



UNIVERSIDAD TECNOLÓGICA DE LA MIXTECA

SOLUCIÓN DEL PROBLEMA INVERSO DE  
ESTIMACIÓN DE DISTRIBUCIONES DE TAMAÑOS  
DE PARTÍCULA

TESIS:

PARA OBTENER EL TÍTULO DE  
LICENCIADO EN MATEMÁTICAS APLICADAS

PRESENTA:

MILTON ALEJANDRO ESCOBAR VÁZQUEZ

DIRECTOR DE TESIS:

DRA. SILVIA REYES MORA

CO-DIRECTOR DE TESIS:

DR. ÁNGEL SINUE CRUZ FÉLIX

HUAJUAPAN DE LEÓN, OAXACA.

ENERO DE 2023.



*Dedicado a  
mis padres.*

# Agradecimientos

A los Doctores Silvia Reyes Mora y Ángel Sinue Cruz Félix por el tiempo dedicado a la dirección, revisión y corrección de esta tesis, y por toda su ayuda y buena disposición a lo largo de la licenciatura.

A la Universidad Tecnológica de la Mixteca por el apoyo brindado durante estos años de estudio, en los cuales he logrado terminar mi carrera profesional, la cual constituye un aliciente para continuar con mi superación.

A mis profesores, Adolfo Maceda, Alejandro Tello, Virgilio Vázquez, Franco Barrágan y Silvia Reyes por confiar en mi y brindarme su ayuda en los momentos más difíciles de mi carrera.

A mis amigos, Luis Mayoral, Daniel Mayoral, Luis López e Iván Martínez por su grata compañía durante las largas jornadas de estudio y por ayudarme a despejar mi mente en los momentos de cansancio.

A mi padre, por su cariño incondicional y su compañía incomparable.

A mis hermanas, Elizabeth, Gabriela y Heidy por su afecto, cariño y apoyo.

A mi sobrina, América, por existir y por tú amor y alegría que llenan mi vida.

A mi madre, Araceli Vázquez, a quien le debo todo lo que soy y todo lo que tengo. Por todo el esfuerzo puesto en pos de mi educación, por enseñarme el valor del trabajo duro y la perseverancia, por nunca perder la confianza en mi y por inculcarme los valores morales que rigen hoy mi vida, y sobre todo, por brindarme una vida plena. Esto será la mejor de las herencias; lo reconozco y lo agradeceré eternamente.

A todos ellos, muchísimas gracias.

# Presentación

En este trabajo de tesis se estudian los principales conceptos matemáticos y físicos relacionados con el problema inverso de estimación de distribuciones de tamaños de partículas (por sus siglas en inglés PSD). Se analiza la ecuación que describe dicho problema, sus componentes, los métodos de solución y una aplicación sintética que muestra la forma en como abordar el problema inverso usando el método de regularización de Tikhonov.

En el Capítulo 1 se definen conceptos básicos de teoría matemática que nos permiten caracterizar el problema inverso de forma teórica y numérica, como propiedades de matrices, conceptos de normas matriciales, espacios de Hilbert, operadores lineales y continuos, el mal o buen planteamiento de un problema, y los valores singulares y el número de condición de una matriz. También, se estudian conceptos sobre la descomposición en valores singulares de una matriz y sobre la matriz de proyección ortogonal ya que son de importancia para resolver el problema inverso.

En el Capítulo 2 se presenta el planteamiento del problema inverso de estimación de PSD, es decir, se describe de forma física y matemática el problema inverso. La descripción física comienza con la técnica óptica de medición, turbidimetría, y con el funcionamiento de un espectrofotómetro con el cual es posible medir el espectro de turbidez. Se define la ecuación operacional continua que describe la relación entre la turbidez y el tamaño de partículas, así como los conceptos básicos de la teoría de Mie que se relacionan con el problema, como la extinción, absorción, dispersión de luz y los coeficientes de Mie. También se estudian las condiciones necesarias para que el problema inverso se encuentre en la llamada región de Rayleigh y sea posible discretizarlo. En la descripción matemática se estudia la ecuación operacional del problema inverso en su forma discreta y matricial, y la clasificación de problemas inversos de acuerdo a su número de parámetros y datos.

En el Capítulo 3 se describen los métodos matemáticos utilizados durante esta tesis para resolver problemas inversos mal condicionados. Se describen los métodos: i) inversa generalizada funcional y matriz pseudoinversa implementado mediante la inversa de Moore-Penrose; ii) regularización de Tikhonov implementado mediante mínimos cuadrados, método de ecuaciones normales y descomposición en valores singulares.

En el capítulo 4 se resuelve el problema inverso de estimación de PSD para partículas esféricas de látex blandos, inmersas en agua, con diámetros entre 80 y 85  $nm$ , y para longitudes de onda entre 300 y 600  $nm$ . Se calculan los valores de la función de distribución de tamaños de partícula  $f(D_j)$  la cual depende del diámetro de la partícula  $D_j$ ; se calculan los valores de las funciones de índice de refracción de la partícula  $n_p(\lambda_i)$ , y del medio  $n_m(\lambda_i)$ , así como los valores del índice de refracción relativo  $m_i$ . También, se calculan los parámetros de tamaño  $x_{ij}$ , los coeficientes de extinción  $Q_{ij}(x_{ij}, m_i)$ , los valores del espectro de turbidez  $\tau(\lambda_i)$  y los coeficientes de la matriz  $A$  de la ecuación matricial del problema inverso discreto. Se resuelve el problema directo y para resolver el problema inverso a la matriz  $A$  se le aplica la inversa de Moore-Penrose y la regularización de Tikhonov mediante Descomposición en Valores Singulares (por sus siglas en inglés SVD) para obtener una solución aproximada  $f_\alpha(D_j)$  a la solución original  $f(D_j)$ .

Finalmente, se presentan las principales conclusiones de la tesis y se plantean algunas sugerencias para trabajos futuros. En la parte final de esta tesis se presenta un apéndice, en el cual se estudian conceptos físicos básicos como: la onda bajo reflexión, la onda bajo refracción, el índice de refracción, el principio de Huygens y la dispersión de luz. Dichos conceptos son de gran ayuda para la comprensión de esta tesis.

# Índice general

<b>Presentación</b>	<b>III</b>
<b>Introducción</b>	<b>1</b>
<b>1. Conceptos Básicos</b>	<b>6</b>
1.1. Introducción . . . . .	6
1.2. Propiedades de Matrices . . . . .	8
1.3. Normas matriciales y Espacios de Hilbert . . . . .	12
1.4. Valores singulares . . . . .	20
1.5. Operadores lineales y continuidad . . . . .	25
1.6. Problemas bien planteados . . . . .	32
1.7. Condicionamiento de una matriz . . . . .	38
<b>2. Planteamiento del problema inverso de PSD</b>	<b>49</b>
2.1. La turbidimetría como técnica óptica de medición . . . . .	50
2.1.1. Medición del espectro de turbidez . . . . .	53
2.1.2. Extinción, absorción, y dispersión de luz . . . . .	58
2.2. Teoría de Mie . . . . .	61
2.2.1. Coeficientes de dispersión de luz . . . . .	62
2.2.2. Partícula Pequeña: Teoría de Rayleigh . . . . .	68
2.3. Descripción matemática del problema . . . . .	71
<b>3. Solución del problema inverso de PSD.</b>	<b>75</b>
3.1. Inversa Generalizada de Moore-Penrose . . . . .	75
3.1.1. Inversa generalizada funcional y matriz pseudoinversa . . . . .	75
3.1.2. Inversa de Moore-Penrose . . . . .	80
3.1.3. Inversa generalizada aplicada al problema inverso de PSD . . . . .	91
3.2. Regularización de Tikhonov . . . . .	93

<i>ÍNDICE GENERAL</i>	VI
3.2.1. Mínimos cuadrados . . . . .	93
3.2.2. Método de ecuaciones normales . . . . .	95
3.2.3. Problemas de mínimos cuadrados con rango deficiente . . .	98
3.2.4. Métodos de Regularización . . . . .	101
<b>4. Aplicación del problema inverso</b>	<b>116</b>
4.1. Solución del problema directo . . . . .	117
4.2. Solución del problema inverso . . . . .	123
4.2.1. Solución obtenida por la inversa generalizada . . . . .	123
4.2.2. Solución obtenida por la regularización de Tikhonov . . . .	129
<b>Conclusiones</b>	<b>134</b>
<b>A. Conceptos físicos básicos</b>	<b>136</b>
<b>Bibliografía</b>	<b>163</b>

# Índice de figuras

1.1. Esquema de una medición sobre la superficie en la posición $x$ para obtener información sobre una anomalía geológica $x'$ en el interior de la tierra, localizada a una profundidad $h$ . . . . .	7
1.2. Proyección ortogonal de $v \in V$ sobre $W$ . . . . .	29
2.1. Experimento de dispersión de luz. . . . .	51
2.2. Esquema de un espectrofotómetro para medición de la turbidez. .	52
2.3. Efectos de luz al atravesar la muestra $M$ . $S_t$ es la luz difundida por las partículas suspendidas en $M$ . $R_1$ y $R_2$ son los rayos periféricos del haz de luz difundido. . . . .	56
2.4. Extinción por una partícula arbitraria en un medio no absorbente.	59
2.5. Sistema de coordenadas polares esférico centrado en una partícula esférica de radio $a$ . . . . .	63
3.1. Proyección ortogonal de $b$ sobre $Im(A)$ . . . . .	96
A.1. Óptica geométrica. . . . .	138
A.2. Trayectoria en línea recta. . . . .	138
A.3. Difracción de ondas. . . . .	139
A.4. Fuente puntual de ondas . . . . .	139
A.5. Reflexión de un pulso con extremo fijo. . . . .	140
A.6. Reflexión especular. . . . .	141
A.7. Reflexión difusa. . . . .	141
A.8. Reflexión especular. . . . .	141
A.9. Refracción. . . . .	142
A.10. Cuando un haz de luz pasa del aire al vidrio, la luz disminuye su velocidad al entrar a éste y su trayectoria se dobla hacia la normal.	143

A.11. Cuando el haz de luz se mueve del vidrio al aire, la luz aumenta su velocidad al entrar al aire y su trayectoria se dobla alejándose de la normal. . . . . 144

A.12. Luz que pasa de un átomo a otro en un medio. Los puntos son electrones y las flechas verticales representan sus oscilaciones. . . . 144

A.13. Cuando una onda se mueve del medio 1 al medio 2, cambia su longitud de onda, pero su frecuencia permanece constante. . . . . 145

A.14. Construcción de Huygens para una onda plana que se propaga a la derecha. . . . . 147

A.15. Construcción de Huygens para demostrar la ley de reflexión. . . . 148

A.16. Construcción de Huygens para demostrar la ley de la refracción de Snell. . . . . 150

A.17. Formación del arco iris primario. . . . . 151

A.18. Ángulo del arco iris. . . . . 155

A.19. Formación del arco iris secundario. . . . . 157

A.20. Ángulo del arco iris primario y secundario. . . . . 160

A.21. Los rayos se desplazan de un medio de índice de refracción  $n_1$  hacia un medio de índice de refracción  $n_2$ , donde  $n_2 < n_1$ . Cuando el ángulo de incidencia aumenta  $\theta_1$ , el ángulo de refracción  $\theta_2$  aumenta hasta que  $\theta_2$  es de  $90^\circ$  (rayo 4). La línea discontinua indica que no hay energía que se propague en esta dirección. Para ángulos de incidencia incluso mayores, se presenta reflexión interna total (rayo 5). . . . . 161

A.22. El ángulo de incidencia que produce un ángulo de refracción igual a  $90^\circ$  es el ángulo crítico  $\theta_c$ , con este ángulo de incidencia se refleja toda la energía de la luz incidente. . . . . 162

# Introducción

En la mayoría de los problemas que se plantean en distintas áreas del conocimiento, incluso en los problemas de la vida cotidiana, es posible encontrar problemas directos y problemas inversos. Los problemas directos, son aquellos en los que se provee la información necesaria para llevar a cabo un proceso bien definido, que lleva a una única solución ([15]). Los problemas inversos están presentes en muchas situaciones, incluso dan la impresión de estar en todas partes; dichos problemas son más difíciles e interesantes, esto se debe a que tienen soluciones múltiples o bien son insolubles, o bien tienen problemas de ser inestables en sentido de Hadamard ([4]). La resolución de un problema directo, involucra análisis o razonamiento progresivo, de causas a efectos; en cambio, la resolución de un problema inverso involucra síntesis, o razonamiento regresivo, de efectos a causas ([4]). En diversos problemas en ciencia y en ingeniería, tales como la meteorología, astronomía, formación de imágenes biomédicas, exploración geológica y dispersión inversa de ondas, entre otros, se dispone de información indirecta acerca de un objeto o magnitud física de interés. En el área de las matemáticas el estudio de los problemas inversos es muy común. Por ejemplo, en [20] podemos encontrar algunos ejemplos de problemas inversos, que están asociados a problemas directos, como los siguientes:

- Hallar un polinomio  $P$  de grado  $n$  que toma valores dados,  $y_1, \dots, y_n \in \mathbb{R}$ , en los puntos dados,  $x_1, \dots, x_n \in \mathbb{R}$ , es el problema inverso al problema directo de calcular los valores del polinomio dado  $P$  en los puntos  $x_1, \dots, x_n$ ; a este problema inverso se le conoce como interpolación.
- Encontrar la matriz diagonal  $D$ , dada la matriz simétrica  $A$  de tamaño  $n \times n$  y  $n$  números reales,  $\lambda_1, \dots, \lambda_n$ , que sean los valores propios de la matriz  $A + D$ ; este problema es inverso al problema directo de calcular los valores propios de la matriz dada  $A + D$ .

Como se muestra en los ejemplos anteriores, los problemas inversos juegan un papel importante en la caracterización del objeto de interés, a partir de información indirecta disponible. El planteamiento de un problema inverso comienza por modelar matemáticamente el problema directo, esto es, el problema físico que relaciona el objeto de interés con la información adquirida; el modelo obtenido nos permite clasificar el problema inverso en lineal o no lineal y nos permite determinar si la caracterización del objeto es sensible a perturbaciones respecto a la información adquirida; es decir, si el problema inverso es mal condicionado o no. Si el problema es mal condicionado es necesario determinar métodos de regularización; esto es, métodos eficientes que permitan obtener una estimación del objeto de interés.

El problema inverso de estimación de distribuciones de tamaños de partículas, ha sido estudiado durante los últimos diez años por varias ramas de la ciencia, tales como la química, física, óptica, matemáticas y computación. Según [18], las propiedades dispersantes de una partícula dependen de su tamaño, de su forma, de la parte real y de la parte imaginaria de su índice de refracción, y de la distribución de tamaños de partícula (PSD). Por ejemplo, en la atmósfera, donde se encuentran las nubes, el polvo y los aerosoles, la distribución de tamaños de partículas no es despreciable si se relaciona con una longitud de onda de radiación; es decir, con energía que se libera en forma de ondas electromagnéticas ([18]). La PSD es una característica física de gran importancia en sistemas coloidales particulados como aerosoles, emulsiones<sup>1</sup>, suspensiones<sup>2</sup>, dispersiones<sup>3</sup>, polvos, etc. ([7]). El comportamiento reológico y la estabilidad química de emulsiones y dispersiones, las velocidades de reacción y difusión, y ciertas propiedades magnéticas y ópticas son influenciadas por la PSD ([7]). En aplicaciones industriales, la PSD puede afectar el sabor y textura de ciertos alimentos, las propiedades de las pinturas, tintas y toners, los procesos de fabricación de cerámicas, y la velocidad de consumo de combustibles y explosivos ([7]). En el caso particular de los coloides poliméricos<sup>4</sup>

---

<sup>1</sup>Sistema heterogéneo constituido por dos líquidos inmiscibles. Un líquido es dispersado en otro. Un ejemplo clásico de emulsión es la mezcla entre agua y aceite.

<sup>2</sup>Una suspensión es un tipo de mezcla heterogénea constituida por pequeñas partículas de un sólido dispersas en un medio líquido en el que no pueden disolverse.

<sup>3</sup>Una dispersión es un sistema en el que las partículas distribuidas de un material se dispersan en una fase continua de otro material. Las dos fases pueden estar en el mismo estado o en diferentes estados de la materia.

<sup>4</sup>Es una dispersión acuosa, carente de toxicidad, que tienen las partículas del plástico y del caucho dispersas en agua. Son partículas muy pequeñas que están dispuestas de manera segura y conveniente para fabricar este polímero que luego se transforma en látex para muros o adhesivos.

(látex), la PSD es una propiedad muy importante dado que afecta su formulación, su procesabilidad y las propiedades de uso final del material (propiedades reológicas<sup>5</sup>, físicas y mecánicas), cuando se le utiliza en adhesivos, recubrimientos, pinturas o tintas ([18]). Por ejemplo, la PSD resulta muy importante en la producción de látex de alto contenido en sólidos; donde una PSD bimodal permite una reducción de la viscosidad del látex con respecto a una PSD unimodal de igual tamaño medio de partícula ([18]). La PSD también influye en la formación de partículas poliméricas<sup>6</sup>, su crecimiento y su interacción entre ellas dentro de un proceso de polimerización<sup>7</sup>. Por estas razones, el conocimiento preciso de la PSD es necesario para caracterizar el producto final, para entender e interpretar los procesos fisicoquímicos que tienen lugar en los procesos de polimerización, y para desarrollar políticas de control sobre dichos procesos ([18]).

Una herramienta factible para el monitoreo de tamaños de partícula en reacciones de polimerización<sup>8</sup> es la Turbidimetría ([18]). La estimación de la PSD a partir de mediciones de turbidimetría, requiere la resolución de un problema inverso mal condicionado, el cual consiste en invertir la medición obtenida sobre la base del modelo matemático que describe el fenómeno de turbidez; es decir, la teoría de dispersión de Mie ([18]). En 1908, el físico alemán Gustav Mie (1868-1957) desarrolló la primera teoría sobre la dispersión por una partícula esférica, la cual describe la interacción de una onda plana con un dieléctrico esférico y relaciona las mediciones de turbidez con la distribución de tamaños de partícula ([7]). Según Mie, se produce más dispersión hacia adelante que en ninguna otra dirección y conforme aumenta el tamaño de la partícula aumenta la dispersión hacia adelante. Apesar de la simplicidad experimental del método de turbidimetría pocos investigadores a intentado determinar la PSD usando esta técnica. Esto se debe a las dificultades numéricas involucradas en el tratamiento de datos y a la falta de información precisa sobre las propiedades ópticas requeridas ([24]). En un experimento de turbidimetría se miden a distintas longitudes de onda, la atenuación de un haz de luz que pasa a través de una o varias partículas sumergidas en agua, a este haz de luz

---

<sup>5</sup>La reología es la rama de la física que se estudia el modo en que los materiales se deforman o fluyen en respuesta a fuerzas o tensiones aplicadas. Las propiedades reológicas estudian el comportamiento de los fluidos sometidos a carga mecánica.

<sup>6</sup>Partículas elaboradas a partir de materiales poliméricos de diferente naturaleza.

<sup>7</sup>Es el proceso químico de convertir una mezcla de monómeros o moléculas de bajo peso molecular en una molécula de gran peso, llamado polímero.

<sup>8</sup>La reacción de polimerización es consecuencia de un proceso de polimerización. En los adhesivos es común observar un cambio de fase pasando del estado líquido al estado sólido una vez que ocurre la reacción de polimerización.

se le conoce como espectro de turbidez. Para realizar las mediciones de turbidez se usa un equipo llamado espectrofotómetro, el cual tiene la capacidad de proyectar un haz de luz monocromática a través de una muestra y medir la cantidad de luz que es absorbida por dicha muestra. Un espectrofotómetro está compuesto por los siguientes dispositivos ([10, 18]):

- **Fuente de Luz:** un espectrofotómetro cuenta con dos lámparas, una halógena (espectro visible (Vis) ) y una de deuterio (espectro ultravioleta (UV) ), las cuales emiten la radiación en un espectro de longitudes de onda continuo.
- **Monocromador:** El monocromador aísla las radiaciones de longitud de onda deseada que inciden sobre la muestra a analizar, se usa para obtener luz monocromática.
- **Fotómetro o Detector:** es el encargado de captar la radiación y a su vez dejarla en evidencia para su estudio posterior; es decir, detecta la cantidad de fotones que se absorben en la muestra y envía una señal a un galvanómetro o pantalla digital. Existen dos tipos de fotómetros: los que responden a los fotones y los que responden al calor.
- **Compartimiento de muestra:** es donde se produce la interacción entre la partícula y la longitud de onda. Durante este proceso, se aplica la ley de Beer-Lambert con base en sus leyes de absorción.

Las técnicas de dispersión de luz como la turbidimetría, la dispersión estática y la dispersión dinámica, no solo permiten calcular la intensidad de luz transmitida, también es posible estimar distribuciones de tamaños de partícula y diámetros medios de partículas en sistemas dispersos ([7]). Durante el estudio del problema inverso de estimación de la PSD, se han propuesto diferentes métodos para resolverlo a partir del conocimiento de la turbidez basados en técnicas de regularización ([20]). En 2012, Luis Gugliotta y Luis Clementi estudiaron el método de redes neuronales de regresión generalizada para la solución al problema inverso de estimación de la PSD de látex en sistemas coloidales a partir de mediciones de turbidimetría. Dicho método se analiza sobre muestras de látex de poliestireno en agua con diámetros promedios de partícula de  $50\text{ nm}$ ,  $100\text{ nm}$  y  $200\text{ nm}$ . En 1969, Yakamoto y Tanaka estimaron la PSD de aerosoles basados en mediciones de turbidimetría, suponiendo que el índice de refracción de los aerosoles es 1.50

y discutieron el efecto de adoptar un índice de refracción diferente en la estimación. En 1971, H. Grassl presentó un método de iteración para la determinación de la PSD de aerosoles a partir de mediciones de turbidimetría; dicha iteración considera el factor de extinción de partículas como un conjunto de funciones de ponderación. También discute la influencia de la humedad en el índice de refracción de la partícula y el medio no absorbente.

El presente trabajo de tesis se centra en el estudio de forma teórica y numérica del problema inverso de estimación de PSD para partículas esféricas de látex industriales, a partir de mediciones de turbidimetría, lo cual requiere la resolución del problema inverso mal condicionado dado por:

$$\tau = Af;$$

donde  $\tau$  son las mediciones de turbidimetría o espectro de turbidez,  $A$  es la matriz que se obtiene a partir de la teoría de Mie, y  $f$  es la PSD que se quiere estimar. En este caso, el problema inverso consiste en hallar  $f$  a partir del conocimiento de  $\tau$  y la matriz mal condicionada  $A$ . Este trabajo también pretende ser una guía para abordar el problema inverso de estimación de PSD en su forma matricial, presentando las principales definiciones y conceptos asociados a la ecuación operacional que describe la relación entre la turbidez y el tamaño de partículas en términos de un número finito de datos, además, se realiza una aproximación a la solución original en una aplicación sintética del problema inverso mediante métodos de regularización como la inversa generalizada de Moore-Penrose y la regularización de Tikhonov.

# Capítulo 1

## Conceptos Básicos

Este capítulo comienza con una introducción a los problemas inversos; en las secciones restantes, se analizan definiciones y resultados básicos en los cuales se basa la modelación y la solución al problema inverso de estimación de distribuciones de tamaños de partícula, los cuales serán útiles en los capítulos siguientes.

### 1.1. Introducción

Los problemas inversos representan la dualidad de todo problema que se presenta en las matemáticas, es la contraparte del problema directo y muchas de las veces resulta más complejo de resolver. En general, al resolver un problema inverso se busca caracterizar el objeto de interés, a partir de información indirecta, es decir, a partir de sus consecuencias. Por ejemplo, la prospección geológica abordada en el Ejemplo 1.1.

**Ejemplo 1.1.** (*Prospección geológica*)

La prospección geológica consiste en determinar la ubicación, la forma o algunos parámetros de anomalías geológicas en el interior de la tierra, a partir de mediciones en la superficie ([20]). Es una de las actividades involucradas en el descubrimiento de yacimientos de minerales útiles. Esta actividad se realiza a partir de un estudio de la estructura geológica del terreno junto con un mapa geológico el cual proporciona una distribución probable de minerales útiles en los territorios bajo investigación. En lugares donde las consideraciones teóricas indican la presencia de depósitos enterrados que no llegan a la superficie, la prospección geológica implica la perforación de pozos exploratorios y, en ocasiones, la excavación de pozos y socavones profundos.

Como se ha dicho anteriormente, en la prospección geológica se quiere conocer información acerca de la composición de la tierra que está localizada a una profundidad  $h$ , a partir de mediciones obtenidas desde la superficie. Por ejemplo, cuando se quiere determinar los cambios en la densidad de masa  $\rho(x)$ ,  $0 \leq x \leq 1$ , de una región anómala localizada a una profundidad  $h$ , se realizan mediciones de la componente vertical  $f_v(x)$  del cambio de fuerza en el punto  $x$ . La masa de un elemento de volumen en el punto  $x'$  está dada por:

$$m = \rho(x')\Delta x',$$

y la distancia entre el instrumento con el que se mide en la superficie, ubicado en la posición  $x$  y la región anómala, ubicada en la posición  $x'$  con profundidad  $h$ , se define como:

$$d(x, x') = \sqrt{(x - x')^2 + h^2}.$$

El cambio de gravedad se describe mediante la ley de la gravedad de Newton definida como:

$$f = \gamma m/d^2,$$

con  $\gamma$  la constante gravitacional.

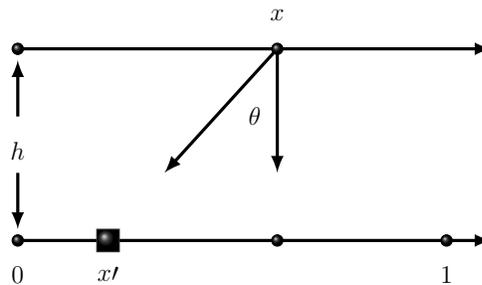


Figura 1.1: Esquema de una medición sobre la superficie en la posición  $x$  para obtener información sobre una anomalía geológica  $x'$  en el interior de la tierra, localizada a una profundidad  $h$ .

Si el lector desea encontrar más información y detalles de este problema inverso puede consultar [20].

## 1.2. Propiedades de Matrices

En esta sección se enuncian algunos resultados importantes sobre matrices que se usarán en este trabajo, dichos resultados fueron tomados de [5, 8, 16, 32].

Sea  $V$  un espacio vectorial sobre un campo  $\mathbb{K}$  de escalares ( $\mathbb{K} = \mathbb{R}$  o  $\mathbb{C}$ ), y denotemos por  $M_{n \times n}(\mathbb{K})$  al conjunto de matrices de tamaño  $n \times n$  con entradas en  $\mathbb{K}$ . Las entradas de una matriz  $A$  se denotan como  $a_{i,j}$  o  $A[i, j]$ .

**Nota 1.1.** Los espacios vectoriales con los que se trabajará son de dimensión finita (salvo se diga lo contrario).

**Definición 1.1.** Dada una matriz  $A \in M_{n \times n}(\mathbb{C})$ , se define y denota por:

- $A^t$ , a la matriz que cumple que  $A^t[i, j] = A[j, i]$  para todo  $i, j$  y que es llamada matriz transpuesta de  $A$ .
- $A^*$ , a la matriz que cumple que  $A^*[i, j] = \overline{A[j, i]}$  para todo  $i, j$  y que es llamada matriz transpuesta conjugada o matriz adjunta de  $A$ .
- Matriz hermitiana, si cumple que  $A = A^*$ .
- Matriz normal si cumple que  $AA^* = A^*A$ .
- Matriz simétrica, si  $A \in M_{n \times n}(\mathbb{R})$ , y  $A = A^t$ .

El siguiente corolario es consecuencia del Teorema 1.18 el cual se enuncia más adelante.

**Corolario 1.1.** Sean  $A, B \in M_{n \times n}(\mathbb{C})$ . Entonces se cumple que:

1.  $(A + B)^* = A^* + B^*$ .
2.  $(\alpha A)^* = \bar{\alpha}A^*$ ,  $\forall \alpha \in \mathbb{K}$ .
3.  $(AB)^* = B^*A^*$ .
4.  $A^{**} = A$ .
5.  $I^* = I$ .

**Teorema 1.1.** Suponga que  $A \in M_{m \times n}(\mathbb{K})$  y  $B \in M_{n \times p}(\mathbb{K})$ . Entonces se cumple que:

1.  $(A^t)^t = A$ .

2.  $(AB)^t = B^t A^t$ .
3. Si  $A, B \in M_{m \times n}(\mathbb{K})$ , entonces  $(A + B)^t = A^t + B^t$ .
4. Si  $A$  es invertible, entonces  $A^t$  es invertible y  $(A^t)^{-1} = (A^{-1})^t$ .

**Definición 1.2.** Sea  $A \in M_{n \times n}(\mathbb{K})$ , se dice que es regular o invertible si existe otra matriz cuadrada de dimensión  $n$ , llamada matriz inversa de  $A$  y denotada por  $A^{-1}$ , tal que  $AA^{-1} = A^{-1}A = I_{n \times n}$ .

**Definición 1.3.** Se dice que  $\lambda \in \mathbb{K}$  es un autovalor o valor propio de la matriz  $A \in M_{n \times n}(\mathbb{K})$  si:

$$\exists v \in V, v \neq 0 : Av = \lambda v.$$

En tal caso, al vector  $v$  se le llama autovector o vector propio asociado a  $\lambda$ .

**Teorema 1.2.** Sea  $A \in M_{n \times n}(\mathbb{K})$ .

1. Un escalar  $\lambda$  es un valor propio o eigenvalor de  $A$  si y sólo si,

$$p(\lambda) = \det(A - \lambda I) = 0.$$

2. Un vector  $v$  es un vector propio o eigenvector de  $A$  correspondiente a un valor propio  $\lambda$  si y sólo si  $v$  es una solución no trivial del sistema:

$$(A - \lambda I)v = 0.$$

**Definición 1.4.** Una matriz  $A \in M_{n \times n}(\mathbb{K})$  es definida positiva, si es simétrica y si  $x^t Ax > 0$  para todo  $x \in \mathbb{K}^n$ , donde  $x \neq 0$ .

**Observación 1.1.** Una consecuencia de la Definición 1.4 es que si la matriz  $A$  es definida positiva, entonces todos los valores propios de la matriz  $A$  son positivos.

Las siguientes son propiedades de las matrices definidas positivas:

1. Toda matriz definida positiva es invertible.
2. Si  $A$  es una matriz definida positiva y  $\alpha$  es un número real positivo, entonces  $\alpha A$  es definida positiva.
3. Si  $A$  y  $B$  son matrices definidas positivas, entonces la suma  $A + B$  también lo es.

4. Si  $AB = BA$ , entonces  $AB$  es también definida positiva.

**Definición 1.5.** Se llama espectro de  $A$  y se denota como  $sp(A)$  al conjunto de valores propios de  $A$ .

**Definición 1.6.**  $N_A$  se denomina el espacio nulo o nulidad de  $A$  y está definido y denotado por:

$$N_A = \{x \in \mathbb{R}^n : Ax = 0\}.$$

**Nota 1.2.** A la dimensión de  $N_A$ , se le denota por  $\nu(A)$  y se denomina nulidad de  $A$ . Si  $N_A$  contiene solo al vector cero, entonces  $\nu(A) = 0$ .

**Definición 1.7.** Sea  $A \in M_{m \times n}(\mathbb{R})$ . Entonces el espacio imagen de  $A$ , denotado por  $Im(A)$ , está dada por:

$$Im(A) = \{y \in \mathbb{R}^m : Ax = y \text{ para alguna } x \in \mathbb{R}^n\}.$$

**Definición 1.8.** Sea  $A \in M_{m \times n}(\mathbb{R})$ . Entonces el rango de  $A$ , denotado por  $ran(A)$ , está dado por:

$$ran(A) = \dim Im(A).$$

**Definición 1.9.** Si  $A \in M_{m \times n}(\mathbb{R})$ , sean  $\{r_1, r_2, \dots, r_m\}$  los renglones de  $A$  y  $\{c_1, c_2, \dots, c_n\}$  las columnas de  $A$ . Entonces se define:

$$R_A = \text{gen}\{r_1, r_2, \dots, r_m\} \text{ y } C_A = \text{gen}\{c_1, c_2, \dots, c_n\}.$$

$R_A$  es llamado espacio de los renglones de  $A$  y  $C_A$  es llamado espacio de las columnas de  $A$ .

**Definición 1.10.** Una matriz  $A \in M_{m \times n}(\mathbb{R})$  se dice de rango completo por renglones si se cumple que:

$$\dim R_A = m;$$

es decir, la cantidad de renglones linealmente independientes de  $A$  es  $m$ .

**Definición 1.11.** Una matriz  $A \in M_{m \times n}(\mathbb{R})$  se dice de rango completo por columnas si se cumple que:

$$\dim C_A = n;$$

es decir, la cantidad de columnas linealmente independientes de la matriz  $A$  es  $n$ .

**Definición 1.12.** Una matriz  $Q \in M_{n \times n}(\mathbb{R})$  se llama ortogonal si:

$$Q^t Q = I_n \quad y \quad Q Q^t = I_n.$$

El conjunto de las matrices ortogonales de orden  $n$  se denota y se define como:

$$O(n, \mathbb{R}) = \{Q \in M_{n \times n}(\mathbb{R}) : Q^t Q = I_n \quad y \quad Q Q^t = I_n\}.$$

**Teorema 1.3.** Sea la matriz  $Q \in M_{n \times n}(\mathbb{R})$ . Las siguientes condiciones son equivalentes:

a)  $Q \in O(n, \mathbb{R})$ .

b)  $Q$  preserva el producto interno de vectores:

$$\forall x, y \in \mathbb{R}^n, \quad \langle Qx, Qy \rangle = \langle x, y \rangle.$$

c)  $Q$  preserva la norma euclidiana de vectores, denotada por  $\|\cdot\|_2$ :

$$\forall x \in \mathbb{R}^n, \quad \|Qx\|_2 = \|x\|_2.$$

**Teorema 1.4.** Para cualquier matriz  $A$ ,  $C_A = \text{Im}(A)$ . Es decir, la imagen de una matriz es igual al espacio de sus columnas.

**Teorema 1.5.** Si  $A \in M_{m \times n}(\mathbb{R})$ , entonces,

$$\dim R_A = \dim C_A = \dim \text{Im}(A) = \text{ran}(A).$$

**Teorema 1.6.** Sea  $A \in M_{m \times n}(\mathbb{R})$ . Entonces,

$$\text{ran}(A) + \nu(A) = n;$$

esto es, el rango de  $A$  más la nulidad de  $A$  es igual al número de columnas de  $A$ .

**Teorema 1.7.** Sea  $A \in M_{n \times n}(\mathbb{R})$ . Entonces  $A$  es invertible si y solo si,

$$\text{ran}(A) = n.$$

**Teorema 1.8.** Sea  $A \in M_{m \times n}(\mathbb{C})$ . Entonces se cumple lo siguiente:

$$\text{Im}(A^*) = [N_A]^\perp \quad \text{y} \quad N_{A^*} = [\text{Im}(A)]^\perp.$$

**Teorema 1.9.** Una matriz cuadrada es simétrica real si y sólo si es ortogonalmente diagonalizable.

**Teorema 1.10.** Para cualquier matriz  $A \in M_{m \times n}(\mathbb{R})$ , y un vector  $u \in \mathbb{R}^n$  y  $v \in \mathbb{R}^m$ , se cumple que:

$$\langle Au, v \rangle = \langle u, A^t v \rangle.$$

**Demostración:**

$$\langle Au, v \rangle = (Au)^t v = (u^t A^t) v = u^t (A^t v) = \langle u, A^t v \rangle$$

□

### 1.3. Normas matriciales y Espacios de Hilbert

A continuación trabajaremos con matrices de tamaño  $m \times n$ , con coeficientes en el campo de los números complejos  $\mathbb{C}$  y presentaremos definiciones y algunos resultados basados en [27, 34, 2, 8, 42].

**Definición 1.13.** Una función  $\|\cdot\| : \mathbb{C}^n \rightarrow \mathbb{R}$  se dirá una norma vectorial sobre  $\mathbb{C}^n$  si satisface las propiedades:

1. a)  $\|x\| \geq 0$ , para todo  $x \in \mathbb{C}^n$ .

- b)  $\|x\| = 0 \Leftrightarrow x = 0$ , para todo  $x \in \mathbb{C}^n$ .
2.  $\|\alpha x\| = |\alpha| \|x\|$ , para todo  $x \in \mathbb{C}^n$  y  $\alpha \in \mathbb{C}$ .
3.  $\|x + y\| \leq \|x\| + \|y\|$ , para todo  $x, y \in \mathbb{C}^n$ .

**Ejemplo 1.2.** A continuación se presentan las normas más comunes en  $\mathbb{C}^n$ . Si  $x = (x_1, \dots, x_n)^t \in \mathbb{C}^n$ ,

1. La norma vectorial suma (o norma 1) está definida por:

$$\|x\|_1 = \sum_{i=1}^n |x_i|.$$

2. La norma vectorial euclidiana (o norma 2) se define como:

$$\|x\|_2 = \left( \sum_{i=1}^n |x_i|^2 \right)^{1/2}.$$

(A partir de este momento la norma euclidiana se denotará como  $\|\cdot\|_2$ )

3. La norma vectorial  $p$ , está definida por:

$$\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}, \quad \text{para } p \in \mathbb{R}, p \geq 1.$$

(Observe que  $\|\cdot\|_1$  y  $\|\cdot\|_2$  son casos particulares de  $\|\cdot\|_p$ )

4. La norma máx o norma  $\infty$ , o norma de Chebyshev, se define mediante:

$$\|x\|_\infty = \max\{|x_1|, \dots, |x_n|\}.$$

(El uso del subíndice  $\infty$  se debe a que  $\forall x \in \mathbb{C}^n$ ,  $\|x\|_\infty = \lim_{p \rightarrow \infty} \|x\|_p$ )

A continuación se enuncian y prueban algunos resultados importantes sobre las normas vectoriales.

**Teorema 1.11.** *Dos normas vectoriales  $\|\cdot\|_\alpha$  y  $\|\cdot\|_\beta$  sobre  $\mathbb{C}^n$  se dicen equivalentes si existen constantes  $c_1 > 0$ ,  $c_2 > 0$  tales que:*

$$\forall x \in \mathbb{C}^n, \quad c_1 \|x\|_\alpha \leq \|x\|_\beta \leq c_2 \|x\|_\alpha.$$

**Lema 1.1.** Para cualquier  $x \in \mathbb{C}^n$ , las normas vectoriales  $\|\cdot\|_1$ ,  $\|\cdot\|_2$ ,  $\|\cdot\|_\infty$  son equivalentes, es decir, se cumple que:

$$\begin{aligned}\|x\|_2 &\leq \|x\|_1 \leq \sqrt{n} \|x\|_2, \\ \|x\|_\infty &\leq \|x\|_2 \leq \sqrt{n} \|x\|_\infty, \\ \|x\|_\infty &\leq \|x\|_1 \leq n \|x\|_\infty.\end{aligned}$$

**Teorema 1.12.** Toda norma vectorial  $\|\cdot\|$  sobre  $\mathbb{C}^n$  es una función continua.

**Demostración:** Sea  $\|\cdot\|$  cualquier norma vectorial sobre  $\mathbb{C}^n$ .

Si  $x = (x_1, \dots, x_n)^t \in \mathbb{C}^n$  y  $e_1, e_2, \dots, e_n$  los vectores de la base canónica de  $\mathbb{C}^n$ , se tiene que:

$$\begin{aligned}\|x\| &= \|x_1 e_1 + \dots + x_n e_n\| \\ &\leq |x_1| \|e_1\| + \dots + |x_n| \|e_n\| \\ &\leq M (|x_1| + \dots + |x_n|) \quad \text{con } M = \max\{\|e_1\|, \dots, \|e_n\|\} \\ &\leq Mn \|x\|_\infty \\ &\leq Mn \|x\|_2.\end{aligned}$$

Entonces,  $\forall x, y \in \mathbb{C}^n$  se cumple que:

$$\left| \|x\| - \|y\| \right| \leq \|x - y\| \leq Mn \|x - y\|_2;$$

así, si la distancia euclidiana entre  $x$  e  $y$  es pequeña, también lo será la distancia entre los números  $\|x\|$  y  $\|y\|$ ; esto muestra la continuidad de  $\|\cdot\|$ .  $\square$

La teoría de normas de vectores es posible extenderla a matrices, a continuación se presentan definiciones y resultados sobre normas matriciales.

**Definición 1.14.** Una función  $\|\cdot\| : M_{m \times n}(\mathbb{C}) \rightarrow \mathbb{R}$  se dirá una norma matricial sobre  $M_{m \times n}(\mathbb{C})$  si cumple con las siguientes propiedades:

1. a)  $\|A\| \geq 0$ , para toda  $A \in M_{m \times n}(\mathbb{C})$ .  
b)  $\|A\| = 0 \Leftrightarrow A = 0$ , para toda  $A \in M_{m \times n}(\mathbb{C})$ .
2.  $\|\lambda A\| = |\lambda| \|A\|$ , para toda  $A \in M_{m \times n}(\mathbb{C})$  y todo  $\lambda \in \mathbb{C}$ .
3.  $\|A + B\| \leq \|A\| + \|B\|$ , para toda  $A, B \in M_{m \times n}(\mathbb{C})$ .

**Ejemplo 1.3.** Sea la matriz  $A \in M_{m \times n}(\mathbb{C})$  y  $a_{i,j} \in \mathbb{C}$  sus respectivos coeficientes. Las normas matriciales más conocidas son:

1. La norma vectorial suma o norma 1, cuando se aplica a matrices, conduce a la función  $\|\cdot\|_1$  definida por:

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|.$$

2. La norma vectorial euclidiana o norma 2, aplicada a matrices, determina la función  $\|\cdot\|_2$  definida por:

$$\|A\|_2 = \left( \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2}.$$

La norma matricial  $\|\cdot\|_2$  también es conocida como norma de Frobenius, norma de Schur o norma de Hilbert-Schmidt.

3. La norma vectorial  $p$ , aplicada a matrices, determina la función  $\|\cdot\|_p$ , definida por:

$$\|A\|_p = \left( \sum_{i=1}^m \sum_{j=1}^n |a_{i,j}|^p \right)^{1/p}.$$

4. La norma  $\infty$  o norma máx, cuando se aplica a matrices, conduce a la función  $\|\cdot\|_\infty$ , definida por:

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{i,j}|.$$

**Proposición 1.1.** *Las siguientes desigualdades, son propiedades de las normas matriciales:*

1.  $\|A^k\| \leq \|A\|^k$ , para toda  $A \in M_{m \times n}(\mathbb{C})$  y  $k \in \mathbb{N}$ .
2.  $\|I_{n \times n}\| \geq 1$ , con  $I_{n \times n}$  la matriz identidad de  $n \times n$ .
3. Si  $\|A\| < 1$ , entonces  $A^k \rightarrow 0$ , cuando  $k \rightarrow \infty$ , para toda  $A \in M_{m \times n}(\mathbb{C})$  y todo  $k \in \mathbb{N}$ .
4.  $\rho(A) \leq \|A\|$ , para cualquier norma matricial y para toda  $A \in M_{m \times n}(\mathbb{C})$ .

## Norma matricial inducida por una norma vectorial

En la siguiente sección estudiamos el procedimiento para construir la norma matricial inducida por una norma vectorial dada sobre  $\mathbb{C}^n$ .

Consideremos los espacios vectoriales  $\mathbb{C}^n$  y  $\mathbb{C}^m$  con normas vectoriales  $\|\cdot\|_n$ ,  $\|\cdot\|_m$ , y bases canónicas  $\beta$  y  $\gamma$ , respectivamente. Sea  $A \in M_{m \times n}(\mathbb{C})$  la representación matricial de la transformación lineal:

$$\begin{aligned} T : \mathbb{C}^n &\rightarrow \mathbb{C}^m, \\ T(x) &= Ax, \quad \forall x \in \mathbb{C}^n, \end{aligned}$$

sobre las bases canónicas  $\beta$  y  $\gamma$ . Sabemos que la magnitud de una matriz  $A$  está dada por su norma matricial  $\|A\|$ , pero si dicha matriz es la representación matricial de una transformación lineal  $T$ , ¿cómo podríamos medir la magnitud de dicho operador? Lo lógico sería pensar que la forma en como medimos la magnitud de una matriz debería estar relacionada con la forma en la que querríamos medir la longitud de una transformación lineal. La idea general para medir una transformación lineal es ver cuánto crece un vector de entrada  $x$  a medida que se convierte en un vector de salida  $T(x)$ . Dado un vector  $x \in \mathbb{C}^n$ , quisieramos saber como se transforma ese vector, y luego querríamos medir que tan grande es su vector de salida, para eso podríamos considerar las normas vectoriales  $\|\cdot\|_n$  sobre  $\mathbb{C}^n$  y  $\|\cdot\|_m$  sobre  $\mathbb{C}^m$  de tal forma que la magnitud del vector de salida  $T(x)$  es:

$$\|T(x)\|_m.$$

Como nos interesa saber cuánto crece el vector de salida  $T(x)$  en relación con el vector de entrada  $x$ , dividimos entre la magnitud del vector  $x$ , siempre y cuando  $x \neq 0$ , y obtenemos:

$$\frac{\|T(x)\|_m}{\|x\|_n}. \quad (1.1)$$

Pero queremos ver todos los vectores posibles  $x \in \mathbb{C}^n$  y elegir el que está más magnificado y medirlo. Para esto, consideramos el elemento supremo, dado por:

$$\alpha = \sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|T(x)\|_m}{\|x\|_n}.$$

El número  $\alpha$  nos daría la máxima cantidad por la que crecería cualquier vector  $x$  bajo el efecto de  $T$ . Dado que  $A$  es la representación matricial de la transformación

lineal  $T$ , entonces el número  $\alpha$  se puede escribir como:

$$\alpha = \sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|_m}{\|x\|_n};$$

y representa la máxima cantidad por la que crecería cualquier vector  $x$  bajo el efecto de la matriz  $A$ ; tal número  $\alpha$  también es una medida del tamaño de  $A$ . Esta observación motiva una clase de normas matriciales conocidas como normas matriciales inducidas.

**Definición 1.15.** Sean  $\|\cdot\|_n : \mathbb{C}^n \rightarrow \mathbb{R}$  y  $\|\cdot\|_m : \mathbb{C}^m \rightarrow \mathbb{R}$  dos normas vectoriales. Se denomina norma matricial  $\|\cdot\|$  sobre  $M_{m \times n}(\mathbb{C})$  inducida por dichas normas vectoriales a la aplicación:

$$\|\cdot\| : M_{m \times n}(\mathbb{C}) \longrightarrow \mathbb{R},$$

definida como:

$$\|A\| = \sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|_m}{\|x\|_n}.$$

**Proposición 1.2.** Para  $x \in \mathbb{C}^n$ , se cumple que:

$$\sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|_m}{\|x\|_n} = \max_{\|x\|_n=1} \|Ax\|_m.$$

**Demostración:** Sea  $x \in \mathbb{C}^n$ , observemos que:

$$\sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|_m}{\|x\|_n} = \sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{1}{\|x\|_n} \|Ax\|_m.$$

Dado que  $\frac{1}{\|x\|_n} \in \mathbb{C}$ , entonces por la propiedad 2 de la Definición 1.14 se tiene que:

$$\begin{aligned} \sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|_m}{\|x\|_n} &= \sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \left\| \frac{Ax}{\|x\|_n} \right\|_m \\ &= \sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \left\| A \frac{x}{\|x\|_n} \right\|_m. \end{aligned}$$

Haciendo  $y = \frac{x}{\|x\|_n}$ , se obtiene que:

$$\sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|_m}{\|x\|_n} = \sup_{\|y\|_n=1} \|Ay\|_m. \quad (1.2)$$

Observemos que todo vector  $x \in \mathbb{C}^n$  se puede volver unitario si lo dividimos por su norma, así es posible cambiar la notación y reescribir la ecuación (1.2) como:

$$\sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|_m}{\|x\|_n} = \sup_{\|x\|_n=1} \|Ax\|_m. \quad (1.3)$$

Por el Teorema 1.12 sabemos que la función  $\|Ax\|_m$  es continua; además, el conjunto:

$$S = \{x \in \mathbb{C}^n : \|x\|_n = 1\},$$

es compacto. Dado que una función real continua alcanza su máximo y su mínimo en un conjunto compacto, entonces se tiene que la función  $\|Ax\|_m$  alcanza su valor máximo y mínimo en el conjunto  $S$  para todo  $x \in S$ ; por lo tanto, se debe cumplir que:

$$\sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|_m}{\|x\|_n} = \max_{\|x\|_n=1} \|Ax\|_m. \quad (1.4)$$

□

Con esto concluimos la revisión de las normas vectoriales, matriciales y sus equivalencias. A continuación estudiamos conceptos básicos sobre el producto escalar estándar sobre  $\mathbb{C}$  y algunos resultados sobre las sucesiones de Cauchy y los espacios de Hilbert.

## Sucesiones de Cauchy y Espacios de Hilbert

**Definición 1.16.** *El producto escalar estándar en  $\mathbb{C}^n$  es una aplicación,*

$$\mathbb{C}^n \times \mathbb{C}^n \longrightarrow \mathbb{C},$$

$$(x, y) \longmapsto \langle x, y \rangle = y^* x = \sum_{i=1}^n x_i \bar{y}_i ;$$

donde  $y^*$  es el vector transpuesto conjugado de  $y$ . El producto escalar estándar satisface lo siguiente para todo  $x, y, z \in \mathbb{C}^n$  y  $\alpha \in \mathbb{C}$ :

1.  $\langle x, x \rangle \geq 0$ , y  $\langle x, x \rangle = 0$  si y solo si  $x = 0$ ;
2.  $\langle x, y \rangle = \overline{\langle y, x \rangle}$ ;
3.  $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ ;
4.  $\alpha \langle x, y \rangle = \langle \alpha x, y \rangle = \langle x, \alpha y \rangle$ .

**Definición 1.17.** Sea el espacio  $\mathbb{C}^n$ . Un vector  $x \in \mathbb{C}^n$  es un vector unitario si se cumple que  $\|x\| = 1$ . Los vectores  $x, y \in \mathbb{C}^n$  son ortogonales si  $\langle x, y \rangle = 0$ . Un subconjunto  $S$  de  $\mathbb{C}^n$  es ortogonal si cualquier par de elementos distintos de  $S$  es ortogonal. Finalmente un subconjunto  $S$  de  $\mathbb{C}^n$  es ortonormal si  $S$  es ortogonal y está formado únicamente de vectores unitarios.

**Definición 1.18.** Una sucesión  $\{x_n\}$  en un espacio normado se dice ser una sucesión de Cauchy, si dado  $\epsilon > 0$ , existe  $N \in \mathbb{N}$  tal que:

$$\|x_n - x_m\| < \epsilon, \quad \forall n, m \geq N.$$

**Teorema 1.13.** Convergencia de una sucesión de Cauchy.

- a) En cualquier espacio métrico  $X$ , toda sucesión convergente es una sucesión de Cauchy.
- b) Si  $X$  es un espacio métrico compacto y si  $\{P_n\}$  es una sucesión de Cauchy en  $X$ , entonces  $\{P_n\}$  converge a algún punto de  $X$ .
- c) En  $\mathbb{R}^k$ , todas las sucesiones de Cauchy convergen, para  $k \in \mathbb{N}$ .

**Definición 1.19.** Un espacio vectorial lineal normado  $X$  es completo si toda sucesión de Cauchy en  $X$  es convergente en  $X$ . Un espacio vectorial lineal normado completo es llamado espacio de Banach.

**Observación 1.2.** Se dice que un espacio vectorial  $X$  es lineal, si se cumple que para cada  $x \in X$ , este se puede escribir como combinación lineal de vectores en  $X$ .

**Definición 1.20.** Un espacio pre-Hilbert es un espacio vectorial lineal  $X$  con un producto interno definido en  $X \times X$ .

**Definición 1.21.** Un espacio completo pre-Hilbert es llamado espacio de Hilbert.

## 1.4. Valores singulares

Tanto los valores singulares como la descomposición en valores singulares de una matriz son conceptos muy importantes para el objetivo de este trabajo, por lo que definimos el concepto de valor singular y se dan algunos resultados relacionados. Si el lector desea más información sobre las demostraciones de los resultados de esta sección puede consultar [32, 28].

Dada la matriz  $A \in M_{m \times n}(\mathbb{R})$ , se tiene que la matriz  $A^t A$  es simétrica, pues se cumple que:

$$(A^t A)^t = A^t (A^t)^t = A^t A,$$

además, también es semidefinida positiva, ya que:

$$x^t A^t A x = (Ax)^t (Ax) = \|Ax\|^2 \geq 0.$$

Luego, por la Definición 1.4 los valores propios de  $A^t A$  son reales y positivos.

**Definición 1.22.** Sea  $A \in M_{m \times n}(\mathbb{R})$ . Sean  $\lambda_1, \dots, \lambda_n$  los valores propios de  $A^t A$  ordenados en forma decreciente, es decir,

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0.$$

Entonces  $\sigma_i = \sqrt{\lambda_i}$  es el  $i$ -ésimo valor singular de  $A$ .

Los valores singulares  $\sigma_1$  y  $\sigma_n$  de una matriz proporcionan la siguiente información.

**Proposición 1.3.** Sea  $A \in M_{m \times n}(\mathbb{R})$ . Entonces, si  $\sigma_1$  y  $\sigma_n$  son, respectivamente, el mayor y el menor valor singular de  $A$ , se tiene que:

$$\max_{\|x\|=1} \|Ax\| = \sigma_1 \quad y \quad \min_{\|x\|=1} \|Ax\| = \sigma_n.$$

El siguiente resultado es clave para la descomposición en valores singulares de una matriz. Para esto, recordemos que por el Teorema 1.9 toda matriz simétrica de  $n \times n$  es ortogonalmente diagonalizable, es decir, existe una base ortonormal de  $\mathbb{R}^n$  compuesta por sus vectores propios.

**Teorema 1.14.** Sea  $A \in M_{m \times n}(\mathbb{R})$ . Supongamos que  $\lambda_1, \lambda_2, \dots, \sigma_n$  son los valores propios de  $A^t A$  y que:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_n = 0,$$

en otras palabras, los autovalores de  $A^t A$  están ordenados en forma decreciente y el número de valores propios no nulos es  $r$ .

Sea  $\{v_1, \dots, v_n\}$  una base ortonormal de  $\mathbb{R}^n$  tal que  $A^t A v_i = \lambda_i v_i$ , entonces:

1.  $\{A v_1, \dots, A v_n\}$  es un conjunto ortogonal y  $\|A v_i\| = \sqrt{\lambda_i} = \sigma_i$  para todo  $i = 1, \dots, n$ ;
2.  $\left\{ \frac{A v_1}{\sigma_1}, \frac{A v_2}{\sigma_2}, \dots, \frac{A v_r}{\sigma_r} \right\}$  es una base ortonormal para  $C_A$ .
3.  $\{v_{r+1}, \dots, v_n\}$  es una base ortonormal para  $N_A$ ;
4.  $\text{ran}(A) = r = \text{número de valores singulares no nulos}$ .

**Demostración:** Sea  $\{v_1, \dots, v_n\}$  una base ortonormal de  $\mathbb{R}^n$  tal que:

$$A^t A v_i = \lambda_i v_i,$$

1. Probemos primero que  $\{A v_1, \dots, A v_n\}$  es un conjunto ortogonal:

$$\begin{aligned} \langle A v_i, A v_j \rangle &= (A v_i)^t (A v_j) \\ &= v_i^t A^t A v_j \\ &= v_i^t \lambda_j v_j \\ &= \lambda_j (v_i^t v_j) \\ &= \begin{cases} \lambda_j, & \text{si } i = j, \\ 0, & \text{si } i \neq j. \end{cases} \end{aligned}$$

Ahora, probemos que  $\|A v_i\| = \sqrt{\lambda_i} = \sigma_i$ :

$$\begin{aligned} \|A v_i\|^2 &= (A v_i)^t A v_i \\ &= v_i^t A^t A v_i \\ &= v_i^t \lambda_i v_i \\ &= \lambda_i \|v_i\|^2 \\ &= \lambda_i \geq 0. \end{aligned}$$

2. Probemos primero que el conjunto  $\left\{ \frac{Av_1}{\sigma_1}, \frac{Av_2}{\sigma_2}, \dots, \frac{Av_r}{\sigma_r} \right\}$  es ortonormal:

$$\begin{aligned} \left\langle \frac{Av_i}{\sigma_i}, \frac{Av_j}{\sigma_j} \right\rangle &= \frac{1}{\sigma_i \sigma_j} \langle Av_i, Av_j \rangle \\ &= \frac{1}{\sigma_i \sigma_j} \langle A^t Av_i, v_j \rangle \quad \dots \text{Teorema 1.10} \\ &= \frac{1}{\sigma_i \sigma_j} \langle \lambda_i v_i, v_j \rangle \\ &= \frac{\lambda_i}{\sigma_i \sigma_j} \langle v_i, v_j \rangle \\ &= \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \end{aligned}$$

Para  $i \neq j$ , los  $v_i$  son ortogonales, de modo que  $\langle v_i, v_j \rangle = 0$ . Para  $i = j$  tenemos que  $\frac{\lambda_i}{\sigma_i^2} = 1$ , por la definición de valores singulares, y dado que  $\langle v_i, v_i \rangle = \|v_i\|^2 = 1$ , por que  $v_i$  son vectores normales.

Por lo tanto, el conjunto  $\left\{ \frac{Av_1}{\sigma_1}, \frac{Av_2}{\sigma_2}, \dots, \frac{Av_r}{\sigma_r} \right\}$  es ortonormal.

Solo resta probar que:

$$\text{gen}\{Av_1, \dots, Av_r\} = C_A.$$

Consideremos la transformación lineal  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , definida como:

$$T(x) = Ax, \quad \forall x \in \mathbb{R}^n.$$

Observemos que  $\text{Im}(T) = C_A$ , además, como  $\{v_1, \dots, v_n\}$  es una base ortonormal de  $\mathbb{R}^n$ , y  $T(v_i) = Av_i$ , entonces se tiene que:

$$C_A = \text{Im}(T) = \text{gen}\{Av_1, \dots, Av_n\};$$

además,  $\|Av_i\| = \sigma_i = 0$ , si  $i \geq r + 1$ , por lo tanto:

$$C_A = \text{gen}\{Av_1, \dots, Av_r\}.$$

3. Sabemos que  $\{v_{r+1}, \dots, v_n\}$  es un conjunto ortonormal, y que  $\|Av_i\| = 0$  para todo  $i \geq r + 1$ , así por propiedades de las normas vectoriales se tiene que  $Av_i = 0$ . Luego,  $v_i \in N_A$ , además, por el Teorema 1.6 se tiene que

$\dim N_A = n - r$ , por lo tanto,  $\{v_{r+1}, \dots, v_n\}$  es una base ortonormal de  $N_A$ .

4. Se deduce del punto 2 que  $\dim(C_A) = r$ . Luego, por el Teorema 1.5 se tiene que:

$$\dim C_A = \text{ran}(A) = r.$$

Además,  $r$  también es el número de valores propios no nulos de  $A^t A$ , los cuales coinciden con el número de valores singulares no nulos de  $A$ . Por lo tanto, se cumple que:

$$r = \text{ran}(A) = \text{num. de valores singulares de } A.$$

□

A continuación se presentan definiciones y algunos resultados de la teoría de descomposición en valores singulares de una matriz  $A \in M_{m \times n}(\mathbb{R})$  que será de ayuda en la regularización de Tikhonov usada para resolver el problema inverso.

## Descomposición en valores singulares (DVS)

A continuación definimos lo que se conoce como descomposición en valores singulares (DVS) de una matriz.

**Definición 1.23.** Sea  $A \in M_{m \times n}(\mathbb{R})$ . Una descomposición en valores singulares de  $A$  es una factorización de la forma:

$$A = U \Sigma V^t;$$

con  $U \in M_{m \times m}(\mathbb{R})$  y  $V \in M_{n \times n}(\mathbb{R})$  ortogonales y  $\Sigma \in M_{m \times n}(\mathbb{R})$  definida como:

$$\Sigma = \begin{bmatrix} D & \vdots & 0_{r \times (n-r)} \\ \dots & \dots & \dots \\ 0_{(m-r) \times r} & \vdots & 0_{(m-r) \times (n-r)} \end{bmatrix}, \quad \text{donde} \quad D = \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix}$$

con

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0, \quad r \leq m, n.$$

La notación  $0_{k \times l}$  representa la matriz de tamaño  $k \times l$  con coeficientes nulos.

El siguiente resultado nos da una forma de escribir a la matriz  $A$  a partir de su DVS.

**Lema 1.2.** Si  $A = U\Sigma V^t$  es una DVS, con  $\Sigma$  igual que la Definición 1.23 y los vectores  $v_i$  y  $u_i$  son las  $i$ -ésimas columnas de  $V$  y  $U$  respectivamente, entonces la matriz  $A$  se puede escribir como:

$$A = \sum_{i=1}^r \sigma_i u_i v_i^t,$$

donde cada una de las matrices  $u_i v_i^t$  es de rango igual a 1.

**Definición 1.24.** Si  $A = U\Sigma V^t$  es una DVS de  $A$ , a los vectores  $v_i$  que aparecen como columnas de la matriz  $V$  se les denomina vectores singulares derechos de  $A$ , mientras que a los que aparecen como columnas de  $U$  se les denomina vectores singulares izquierdos de  $A$ .

El siguiente resultado nos permite establecer que toda matriz admite una DVS.

**Teorema 1.15.** Sea  $A \in M_{m \times n}(\mathbb{R})$ . Entonces existe una descomposición en valores singulares de  $A$ .

**Demostración:** Sean  $\lambda_1, \dots, \lambda_n$  los valores propios de la matriz simétrica  $A^t A$  ordenados en forma decreciente. Supongamos que exactamente  $r$  valores propios son no nulos ( $r$  podría ser  $n$ ). Sean  $\{v_1, \dots, v_n\}$  los vectores propios de la matriz  $A^t A$ , los cuales forman una base ortonormal de  $\mathbb{R}^n$  tal que  $A^t A v_i = \lambda_i v_i$ , para todo  $i = 1, \dots, n$ . Así, definimos la matriz ortogonal,  $V = [v_1 \ \dots \ v_n]$ , y las matrices:

$$\Sigma = \begin{bmatrix} D & \vdots & 0_{r \times (n-r)} \\ \dots & \dots & \dots \\ 0_{(m-r) \times r} & \vdots & 0_{(m-r) \times (n-r)} \end{bmatrix} \quad \text{con} \quad D = \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix},$$

donde,

$$\sigma_i = \sqrt{\lambda_i}.$$

Hasta ahora solo tenemos las matrices  $V$  y  $\Sigma$  de la DVS de  $A$ , por lo que nos falta hallar la matriz ortogonal  $U$ . Con el objetivo de hallar la matriz  $U$ , consideramos los vectores:

$$u_1 = \frac{A v_1}{\sigma_1}, u_2 = \frac{A v_2}{\sigma_2}, \dots, u_r = \frac{A v_r}{\sigma_r}.$$

Sabemos por el Teorema 1.14 que el conjunto  $\{u_1, u_2, \dots, u_r\}$  es ortonormal. Observemos que si  $r = m$ , entonces es posible definir  $U = \begin{bmatrix} u_1 & \dots & u_r \end{bmatrix}$  que es una matriz ortogonal, en el caso de que  $r < m$  buscamos vectores  $u_{r+1}, \dots, u_m$  de tal forma que  $\{u_1, \dots, u_m\}$  sea una base ortonormal de  $\mathbb{R}^m$  (En [2] se describe una forma de extender un conjunto ortonormal  $S$  a una base ortonormal  $B$ .) Así, es posible definir  $U = \begin{bmatrix} u_1 & \dots & u_m \end{bmatrix}$  que es una matriz ortogonal. Ahora veamos si se cumple que  $A = U\Sigma V^t$ . Por un lado tenemos que:

$$AV = \begin{bmatrix} Av_1 & Av_2 & \dots & Av_n \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 & \dots & \sigma_r u_r & Av_{r+1} & \dots & Av_n \end{bmatrix}.$$

Dado que  $\lambda_{r+1} = \dots = \lambda_n = 0$ , entonces se cumple que:

$$Av_{r+1} = \dots = Av_n = 0,$$

luego,

$$AV = \begin{bmatrix} \sigma_1 u_1 & \dots & \sigma_r u_r & 0 & \dots & 0 \end{bmatrix}.$$

Por otro lado,

$$U\Sigma = \begin{bmatrix} \sigma_1 u_1 & \dots & \sigma_r u_r & 0 & \dots & 0 \end{bmatrix}.$$

Por lo tanto, se cumple que:

$$AV = U\Sigma,$$

y como  $V$  es una matriz ortogonal, es decir,  $V^t = V^{-1}$ , así llegamos a la igualdad:

$$A = U\Sigma V^t.$$

□

Con estos resultados enunciados anteriormente la descomposición en valores singulares de una matriz está completa.

## 1.5. Operadores lineales y continuidad

En esta sección enunciamos resultados importantes sobre transformaciones lineales entre espacios vectoriales formados por funciones, las cuales transforman unas funciones en otras, denominadas operadores lineales. En específico nos interesa el operador lineal llamado proyección ortogonal. Si se requiere más información

respecto a los resultados enunciados puede consultar [26, 11, 16, 20].

**Definición 1.25.** Sean  $V$  y  $W$  espacios vectoriales (sobre un campo  $\mathbb{K}$ ). Un operador  $T : V \rightarrow W$  se dice que es lineal si para toda  $x, y \in V$  y  $\mu, \lambda \in \mathbb{K}$  se cumple que:

$$T(\lambda x + \mu y) = \lambda T(x) + \mu T(y)$$

**Definición 1.26.** Sean  $V$  y  $W$  espacios vectoriales y sea  $T : V \rightarrow W$  un operador lineal. Definimos al espacio nulo  $N(T)$  del operador  $T$  como:

$$N(T) = \{x \in V : T(x) = 0\}.$$

Definimos la imagen  $R(T) \subseteq W$  del operador  $T$  como:

$$R(T) = \{T(x) : x \in V\}.$$

**Definición 1.27.** Sean  $X$  e  $Y$  dos espacios vectoriales normados y  $T : X \rightarrow Y$  un operador lineal. Entonces  $T$  es continuo si y solo si verifica la siguiente condición:

$$\exists M > 0 \in \mathbb{R} : \|T(x)\| \leq M \|x\|, \quad \forall x \in X.$$

**Proposición 1.4.** Un operador  $T$  que va de un espacio normado  $X$  a un espacio normado  $Y$  es continuo en el punto  $x_0 \in X$  si y solo si para cada  $x_n \rightarrow x_0$  implica que  $T(x_n) \rightarrow T(x_0)$ .

**Teorema 1.16.** Un operador lineal es acotado si y solo si es continuo.

**Definición 1.28.** Dado un operador lineal  $\phi : X \rightarrow Y$  con  $X$  e  $Y$  espacios de Hilbert. Una sucesión de elementos  $\{x_n\}$  de un conjunto  $X$  es llamada sucesión minimizante si se cumple que:

$$\lim_{n \rightarrow \infty} \phi(x_n) = \inf_{x \in X} \phi(x).$$

**Definición 1.29.** Sea  $T : X \rightarrow Y$  un operador lineal y acotado sobre espacios de Hilbert. Entonces existe uno y solo un operador lineal acotado  $T^* : Y \rightarrow X$  con la propiedad:

$$\langle T(x), y \rangle = \langle x, T^*(y) \rangle, \quad \forall x \in X, y \in Y.$$

Este operador  $T^* : Y \rightarrow X$  es llamado el operador adjunto de  $T$ .

**Nota 1.3.** El operador adjunto se basa en la generalización de la matriz transpuesta conjugada.

**Definición 1.30.** Sean  $V$  y  $W$  espacios vectoriales de dimensión finita con bases ordenadas  $\beta = \{x_1, \dots, x_n\}$  y  $\gamma = \{y_1, \dots, y_m\}$ . Sea  $T : V \rightarrow W$  un operador lineal. Entonces existen escalares únicos  $a_{ij} \in \mathbb{K}$  ( $i = 1, \dots, m; j = 1, \dots, n$ ) tales que:

$$T(x_i) = \sum_{j=1}^m a_{ij} y_j, \quad \text{para } 1 \leq i \leq n.$$

Llamaremos a la matriz  $A \in M_{m \times n}(\mathbb{K})$ , definida mediante  $A[ij] = a_{ij}$ , como la representación matricial de  $T$  en las bases ordenadas  $\beta$  y  $\gamma$ , y la escribiremos como  $A = [T]_{\beta}^{\gamma}$ . Si  $V = W$  y  $\beta = \gamma$ , escribiremos  $A = [T]_{\beta}$ .

**Definición 1.31.** Sean  $X$  e  $Y$  espacios de Hilbert y  $T : X \rightarrow Y$  un operador compacto con su operador adjunto  $T^* : Y \rightarrow X$ . Las raíces cuadradas  $\sigma_j = \sqrt{\lambda_j}$ ,  $j \in J$  de los valores propios  $\lambda_j$  del operador auto-adjunto  $T^*T : X \rightarrow X$  son llamados los valores singulares de  $T$ . Aquí  $J \subset \mathbb{N}$  podría ser finito o  $J = \mathbb{N}$ .

**Nota 1.4.** Note que cada valor propio  $\lambda$  de  $T^*T$  es no negativo, pues:

$$T^*T(x) = \lambda x,$$

esto implica que:

$$\lambda \langle x, x \rangle = \langle T^*T(x), x \rangle = \langle T(x), T(x) \rangle \geq 0, \text{ i.e., } \lambda \geq 0.$$

**Teorema 1.17.** Sea  $T : X \rightarrow Y$  un operador lineal compacto,  $T^* : Y \rightarrow X$  su operador adjunto, y  $\sigma_1 \geq \sigma_2 \geq \sigma_3 \dots > 0$  la secuencia ordenada de los valores singulares positivos de  $T$ , contados en relación con su multiplicidad. Entonces existen sistemas ortonormales  $\{x_j\} \subset X$  e  $\{y_j\} \subset Y$  con las siguientes propiedades:

$$T(x_j) = \sigma_j y_j \quad \text{y} \quad T^*(y_j) = \sigma_j x_j, \quad \forall j \in J.$$

El sistema  $(\sigma_j, x_j, y_j)$  es llamado un sistema singular para  $T$ . Cada valor  $x \in X$  posee la descomposición en valores singulares:

$$x = x_0 + \sum_{j \in J} \langle x, x_j \rangle x_j,$$

para algún  $x_0 \in N(T)$  y

$$T(x) = \sum_{j \in J} \sigma_j \langle x, x_j \rangle y_j.$$

**Teorema 1.18.** *Sea  $V$  un espacio con producto interior dimensionalmente finito y sean  $T$  y  $U$  operadores lineales en  $V$ . Entonces, se cumple que:*

1.  $T^{**} = T$ .
2.  $(T + U)^* = T^* + U^*$ .
3.  $(\alpha T)^* = \bar{\alpha} T^*$ ,  $\forall \alpha \in \mathbb{K}$ .
4.  $(T \circ U)^* = U^* \circ T^*$ .
5.  $I^* = I$ .
6. Si  $T^* = T$ , se dice que el operador lineal es autoadjunto.

En específico nos interesa el operador lineal llamado operador de proyección ortogonal, para esto es necesario estudiar el concepto de complemento ortogonal y algunos otros resultados que ayudarán en la comprensión de dicho operador.

**Definición 1.32.** *Sea  $V$  un espacio vectorial con producto interior y sea  $W$  un subespacio de  $V$ . Definimos al subespacio  $W^\perp$  como el conjunto de todos aquellos vectores de  $V$  que son ortogonales a todos los vectores de  $W$ ; esto quiere decir que  $W^\perp = \{x \in V : \langle x, y \rangle = 0, \forall y \in W\}$ . Al subespacio  $W^\perp$  se le llama complemento ortogonal de  $W$ .*

**Definición 1.33.** *Sean  $V$  un espacio vectorial y  $W_1$  un subespacio de  $V$ . Un operador  $T : V \rightarrow V$  se llama proyección sobre  $W_1$  si:*

1. Existe un subespacio  $W_2$ , tal que  $V = W_1 \oplus W_2$ .
2. Para todo  $x \in V$ , existen  $x_1 \in W_1$  y  $x_2 \in W_2$ , tales que  $x = x_1 + x_2$ .

**Definición 1.34.** *Sea  $W$  un subespacio dimensionalmente finito de un espacio con producto interior  $V$ . Entonces,*

$$V = W \oplus W^\perp.$$

**Teorema 1.19.** *Un operador lineal  $T$  es inyectivo si y solo si  $N(T) = \{0\}$ .*

**Teorema 1.20.** *Dado el operador lineal  $T : \mathbb{C}^n \rightarrow \mathbb{C}^m$ , la restricción*

$$T|_{R(T^*)} : R(T^*) \rightarrow R(T),$$

*es un isomorfismo de espacios vectoriales.*

**Lema 1.3.** *Sea  $T : \mathbb{C}^n \rightarrow \mathbb{C}^m$  un operador lineal. Entonces  $N(T|_{R(T^*)}) = \{0\}$ .*

**Definición 1.35.** *Sea  $V$  un espacio con producto interior, y sea  $T : V \rightarrow V$  una proyección. Decimos que  $T$  es una proyección ortogonal si  $[R(T)]^\perp = N(T)$  y  $[N(T)]^\perp = R(T)$ .*

Para comprender la diferencia entre una proyección arbitraria y una proyección ortogonal sobre  $W$ , consideremos el ejemplo en [11]. Sea el espacio vectorial  $V = \mathbb{R}^2$  y el subespacio  $W = \{(x, y) : y = x\} \subseteq V$ . Definimos y denotamos los operadores  $U$  y  $T$  como en la Figura 1.2 donde  $U(v)$  y  $T(v)$  son la representación gráfica de una proyección arbitraria y la proyección ortogonal sobre  $W$ , respectivamente. Observemos que el vector  $v - T(v) \in W^\perp$  y  $v - U(v) \notin W^\perp$ .

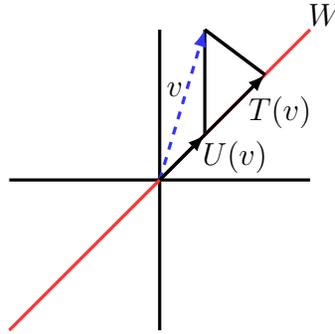


Figura 1.2: Proyección ortogonal de  $v \in V$  sobre  $W$ .

En la Figura 1.2 se puede observar que  $T(v)$  es la mejor aproximación en  $W$  para  $v$ ; es decir, si  $w \in W$ ,  $\|w - v\| \geq \|T(v) - v\|$ . Esta es una propiedad importante que caracteriza al operador lineal proyectador ortogonal.

**Teorema 1.21.** *Sea  $W$  un subespacio dimensionalmente finito de un espacio con producto interior  $V$ , y sea  $T$  la proyección ortogonal sobre  $W$ . Entonces, para cualquier  $v \in V$ : el vector  $T(v)$  es el único elemento de  $W$  que se acerca más a  $v$ ; esto es,  $\|v - T(v)\| \leq \|v - w\|$  para toda  $w \in W$ .*

Otra manera de definir al operador de proyección ortogonal es como el único operador lineal de  $\mathbb{C}^n$  sobre  $W$  tal que:

$$P_W(x) = \begin{cases} x, & x \in W, \\ 0, & x \in W^\perp. \end{cases}$$

Dado que la proyección ortogonal de  $v \in \mathbb{C}^n$  sobre  $W$  es un endomorfismo en  $\mathbb{C}^n$ , entonces tiene una matriz asociada  $P_r \in M_{n \times n}(\mathbb{C})$  tal que:

$$P_r v = \hat{v};$$

donde  $\hat{v}$  es la proyección ortogonal de  $v$ . Para hallar la matriz de proyección ortogonal  $P_r$ , consideremos el conjunto  $B = \{b_1, b_2, \dots, b_d\}$  una base para  $W$ . Definiendo la matriz  $P_B \in M_{n \times d}(\mathbb{C})$  como:

$$P_B = \begin{bmatrix} b_1 & b_2 & \dots & b_d \end{bmatrix};$$

tendremos que:

$$P_B c = \hat{v};$$

siendo  $c$  el vector de coordenadas respecto de la base  $B$  del vector proyectado. Dado que  $v$  y  $\hat{v}$  son vectores ortogonales, se cumple que:  $\langle \hat{v}, v - \hat{v} \rangle = 0$ , con  $\hat{v} \in W$  y  $v - \hat{v} \in W^\perp$ . Reescribiendo esta ecuación en función de  $P_B$ ,  $v$  y  $c$  obtenemos lo siguiente:

$$\langle P_B c, v - P_B c \rangle = 0. \quad (1.5)$$

Dado que el producto escalar es igual a cero, podemos asegurar la conmutatividad, luego, por la Definición 1.16 es posible reescribir la ecuación (1.5) como:

$$\begin{aligned} \langle P_B c, v - P_B c \rangle &= (P_B c)^* (v - P_B c) \\ &= c^* P_B^* (v - P_B c) \\ &= c^* (P_B^* v - P_B^* P_B c) \\ &= 0. \end{aligned}$$

Observemos que el único vector ortogonal a todos los vectores  $c \in \mathbb{C}^d$  es el vector

$0 \in \mathbb{C}^d$ , esto significa que se debe cumplir lo siguiente:

$$P_B^* v - P_B^* P_B c = 0;$$

por lo tanto,

$$P_B^* v = P_B^* P_B c. \quad (1.6)$$

Dado que las columnas de la matriz  $P_B$  son linealmente independientes, entonces se cumple que la matriz  $P_B^* P_B$  es invertible. Luego, al multiplicar ambos miembros de la ecuación (1.6) por  $(P_B^* P_B)^{-1}$  obtenemos lo siguiente:

$$(P_B^* P_B)^{-1} P_B^* v = c.$$

Premultiplicando por  $P_B$ , se obtiene:

$$P_B (P_B^* P_B)^{-1} P_B^* v = P_B c = \hat{v}.$$

Por lo tanto la matriz de proyección ortogonal está dada por:

$$P_r = P_B (P_B^* P_B)^{-1} P_B^*. \quad (1.7)$$

Las siguientes son propiedades de la matriz de proyección ortogonal  $P_r$  :

1. La matriz  $P_r$  es simétrica.

**Demostración:**

$$\begin{aligned} P_r^* &= (P_B (P_B^* P_B)^{-1} P_B^*)^* \\ &= (P_B^*)^* (P_B (P_B^* P_B)^{-1})^* \\ &= P_B ((P_B^* P_B)^{-1})^* P_B^* \\ &= P_B ((P_B^* P_B)^*)^{-1} P_B^* \\ &= P_B (P_B^* P_B)^{-1} P_B^* \\ &= P_r. \end{aligned}$$

□

2. La matriz  $P_r$  es idempotente.

**Demostración:**

$$\begin{aligned}
 P_r^2 &= (P_B (P_B^* P_B)^{-1} P_B^*)^2 \\
 &= P_B (P_B^* P_B)^{-1} P_B^* P_B (P_B^* P_B)^{-1} P_B^* \\
 &= P_B (P_B^* P_B)^{-1} P_B^* \\
 &= P_r.
 \end{aligned}$$

□

El siguiente resultado nos da una condición para que una matriz sea un operador de proyección ortogonal sobre su imagen.

**Teorema 1.22.** *Sea  $A \in M_{n \times n}(\mathbb{C})$ . Entonces,  $A$  es un operador de proyección ortogonal de  $\mathbb{C}^n$  sobre  $R(A)$ , es decir,  $A = P_{R(A)}$ , si y solo si,  $A^2 = A = A^*$ .*

## 1.6. Problemas bien planteados

En esta sección se presentan definiciones básicas para la comprensión del teorema del mapeo abierto y su aplicación en los problemas bien planteados.

**Definición 1.36.** *Sean  $X$  y  $Y$  dos espacios normados y  $T : X \rightarrow Y$  una aplicación lineal o no lineal. La ecuación  $T(x) = y$  se denomina bien planteada si se cumple lo siguiente:*

- *Existencia:* Para todo  $y \in Y$  existe al menos un  $x \in X$  tal que  $T(x) = y$ .
- *Unicidad:* Para todo  $y \in Y$  existe a lo más un  $x \in X$  con  $T(x) = y$ .
- *Estabilidad:* La solución  $x$  depende continuamente de los datos  $y$ ; esto es para cada sucesión  $\{x_n\} \subset X$  tal que  $T(x_n) \rightarrow T(x)$  cuando  $n \rightarrow \infty$ , entonces se tiene que  $x_n \rightarrow x$  cuando  $n \rightarrow \infty$ .

Para ejemplificar la importancia del buen planteamiento de un problema consideremos el problema de Cauchy para la ecuación de Laplace y probemos que el problema inverso resulta en un problema mal planteado.

**Ejemplo 1.4.** Sea el problema de Cauchy para la ecuación de Laplace, definido por el siguiente problema de contorno:

$$\Delta u(x, y) = \frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} = 0, \quad (x, y) \in \mathbb{R} \times [0, \infty); \quad (1.8)$$

con condiciones iniciales:

$$\begin{cases} u(x, 0) = f(x), \\ \frac{\partial u(x, 0)}{\partial y} = g(x), \end{cases} \quad \forall x \in \mathbb{R}.$$

- El **problema directo** consiste en encontrar el valor de las funciones  $f$  y  $g$  a partir del conocimiento de la función solución  $u(x, y)$ .
- El **problema inverso** consiste en encontrar la solución  $u(x, y)$  a partir del conocimiento de las funciones  $f$  y  $g$ .

Nos interesa resolver el problema inverso; es decir, encontrar una solución a la ecuación (1.8) a partir del conocimiento de las funciones  $f$  y  $g$ , a los que llamaremos datos de Cauchy. Para esto, consideremos el caso particular en que  $f$  y  $g$  son las funciones definidas por:

$$\begin{aligned} f(x) &= 0, \\ g(x) &= \frac{1}{n} \operatorname{sen}(nx). \end{aligned}$$

Al resolver el problema inverso, encontramos que la única solución al problema de Cauchy que satisface las condiciones iniciales, está dada por:

$$u(x, y) = \frac{1}{n^2} \operatorname{sen}(nx) \operatorname{senh}(ny), \quad x \in \mathbb{R}, \quad y \geq 0;$$

la cual satisface:

$$\begin{aligned} u(x, 0) &= \frac{1}{n^2} \operatorname{sen}(nx) \operatorname{senh}(0) \\ &= 0, \end{aligned}$$

y también:

$$\begin{aligned} \frac{\partial u(x, 0)}{\partial y} &= \frac{1}{n^2} \operatorname{sen}(nx) (n \operatorname{cosh}(0)) \\ &= \frac{1}{n^2} \operatorname{sen}(nx) n \\ &= \frac{1}{n} \operatorname{sen}(nx). \end{aligned}$$

Ahora vamos a analizar el error de aproximación entre los datos de Cauchy y el error entre las soluciones, con respecto de la norma del supremo. Observemos que el error en los datos está dado por:

$$\begin{aligned} \sup_{x \in \mathbb{R}} \{|f(x)| + |g(x)|\} &= \sup_{x \in \mathbb{R}} \left\{ 0 + \left| \frac{1}{n} \operatorname{sen}(nx) \right| \right\} \\ &= \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \operatorname{sen}(nx) \right| \\ &= \frac{1}{n} \sup_{x \in \mathbb{R}} |\operatorname{sen}(nx)|. \end{aligned}$$

Sabemos que  $\sup_{x \in \mathbb{R}} |\operatorname{sen}(nx)| = 1$ . Por lo tanto,

$$\sup_{x \in \mathbb{R}} \{|f(x)| + |g(x)|\} = \frac{1}{n};$$

luego, se cumple que  $\frac{1}{n} \rightarrow 0$ , cuando  $n \rightarrow \infty$ .

Mientras que el error en la solución está dado por:

$$\begin{aligned} \sup_{x \in \mathbb{R}} |u(x, y)| &= \sup_{x \in \mathbb{R}} \left| \frac{1}{n^2} \operatorname{sen}(nx) \operatorname{senh}(ny) \right| \\ &= \frac{1}{n^2} \operatorname{senh}(ny) \sup_{x \in \mathbb{R}} |\operatorname{sen}(nx)| \\ &= \frac{1}{n^2} \operatorname{senh}(ny); \end{aligned}$$

entonces, se cumple que  $\frac{1}{n^2} \operatorname{senh}(ny) \rightarrow \infty$  cuando  $n \rightarrow \infty$ , para todo  $y > 0$ ; es decir, el error en los datos tienden a cero, pero el error en la solución tiende a infinito. Por lo tanto, la solución al problema de Cauchy  $u(x, y)$ , no depende continuamente de las condiciones iniciales  $f(x)$  y  $g(x)$ , así el problema es un problema mal planteado, pues no se cumple la estabilidad de la Definición 1.36.

**Teorema 1.23.** *Teorema del operador abierto*

Sean  $X$  e  $Y$  dos espacios de Banach y  $T : X \rightarrow Y$  un operador lineal acotado.

- a) Si  $T$  es sobreyectivo, entonces  $T$  es abierto.
- b) Si  $T$  es biyectivo, entonces  $T^{-1}$  es continuo y por tanto acotado.

El teorema del operador abierto da condiciones para que un operador lineal acotado sea abierto y también da condiciones para que un operador lineal acotado

tenga inverso continuo; esto permite demostrar que los problemas cuyo operador inverso sea continuo, son problemas bien planteados.

A continuación, consideremos el Ejemplo 1.5 que ilustra un problema bien planteado y es una aplicación directa del Teorema 1.23.

**Ejemplo 1.5.** Sea  $X = C^2[a, b]$  un espacio de Banach. Definimos la norma de  $f \in X$ , como:

$$\|f\| = \|f\|_\infty + \|f'\|_\infty + \|f''\|_\infty,$$

es decir, tendremos que:

$$\|f\| = \max_{t \in [a, b]} |f(t)| + \max_{t \in [a, b]} |f'(t)| + \max_{t \in [a, b]} |f''(t)| \quad (1.9)$$

Además, también consideremos el espacio de Banach  $Y = C^2[a, b] \times \mathbb{R}^2$ , en el cual la norma de  $g \in Y$  se define como:

$$\|(g, \alpha, \beta)\| = \max_{t \in [a, b]} |g(t)| + |\alpha| + |\beta| \quad (1.10)$$

donde  $\alpha, \beta \in \mathbb{R}$ . Definimos el operador  $T : X \rightarrow Y$  como:

$$T(f) = (y_0 f'' + y_1 f' + y_2 f, f(a), f(b)),$$

donde,  $y_0, y_1, y_2$  son constantes positivas. Demostremos que el operador  $T$  es lineal, para esto consideremos la combinación lineal:

$$l = \alpha f + \beta g, \quad \text{con } f, g \in X, \quad y \quad \alpha, \beta \in \mathbb{R};$$

entonces, se debe cumplir que:

$$T(l) = (y_0 l'' + y_1 l' + y_2 l, l(a), l(b)).$$

Calculamos cada una de las derivadas y cada una de las evaluaciones:

1.  $l'' = \alpha f'' + \beta g''$
2.  $l' = \alpha f' + \beta g'$
3.  $l(a) = \alpha f(a) + \beta g(a)$
4.  $l(b) = \alpha f(b) + \beta g(b)$

Sustituyendo los valores de  $l''$  y  $l'$  en  $T(l)$ , obtenemos lo siguiente:

$$\begin{aligned} y_0 l'' + y_1 l' + y_2 l &= y_0(\alpha f' + \beta g'') + y_1(\alpha f' + \beta g') + y_2(\alpha f + \beta g) \\ &= \alpha(y_0 f'' + y_1 f' + y_2 f) + \beta(y_0 g'' + y_1 g' + y_2 g). \end{aligned}$$

Luego, al sustituir  $l(a)$ ,  $l(b)$  y el valor encontrado anteriormente, en el valor del operador  $T$ , tendremos que este se puede reescribir como:

$$\begin{aligned} T(l) &= (\alpha(y_0 f'' + y_1 f' + y_2 f) + \beta(y_0 g'' + y_1 g' + y_2 g), \alpha f(a) + \beta g(a), \alpha f(b) + \beta g(b)) \\ &= \alpha((y_0 f'' + y_1 f' + y_2 f), f(a), f(b)) + \beta((y_0 g'' + y_1 g' + y_2 g), g(a), g(b)) \\ &= \alpha T(f) + \beta T(g) \end{aligned}$$

Por lo tanto, se cumple que:

$$T(\alpha f + \beta g) = \alpha T(f) + \beta T(g).$$

Así, concluimos que el operador  $T : C^2[a, b] \rightarrow C^2[a, b] \times \mathbb{R}^2$ , es un operador lineal. Ahora probemos que el operador lineal  $T$ , es acotado.

Del Teorema 1.16 sabemos que es necesario probar que el operador  $T$  es continuo, es decir, debemos probar la existencia de una constante  $M > 0$ , tal que para toda  $f \in C^2[a, b]$  se cumpla que:

$$\|T(f)\| \leq M \|f\|.$$

En este caso, por la desigualdad del triángulo y la cota de  $f$  para toda  $x \in [a, b]$ , se tiene lo siguiente:

$$\begin{aligned} \|T(f)\| &= \|(y_0 f'' + y_1 f' + y_2 f, f(a), f(b))\| \\ &= \max_{x \in [a, b]} |y_0 f''(x) + y_1 f'(x) + y_2 f(x)| + |f(a)| + |f(b)| \\ &\leq \max_{x \in [a, b]} |y_0 f''(x) + y_1 f'(x) + y_2 f(x)| + 2 \max_{x \in [a, b]} |f(x)| \\ &\leq y_0 \max_{x \in [a, b]} |f''(x)| + y_1 \max_{x \in [a, b]} |f'(x)| + (y_2 + 2) \max_{x \in [a, b]} |f(x)|; \end{aligned}$$

luego, al factorizar las constantes positivas  $y_0$ ,  $y_1$  y  $(y_2 + 2)$ , se obtiene lo siguiente:

$$\begin{aligned} \|T(f)\| &\leq y_0 y_1 (y_2 + 2) \left( \frac{\max_{x \in [a, b]} |f''(x)|}{y_1 (y_2 + 2)} + \frac{\max_{x \in [a, b]} |f'(x)|}{y_0 (y_2 + 2)} + \frac{\max_{x \in [a, b]} |f(x)|}{y_0 y_1} \right) \\ &\leq y_0 y_1 (y_2 + 2) \left( \max_{x \in [a, b]} |f''(x)| + \max_{x \in [a, b]} |f'(x)| + \max_{x \in [a, b]} |f(x)| \right) \\ &= y_0 y_1 (y_2 + 2) \|f\|; \end{aligned}$$

donde  $M = y_0 y_1 (y_2 + 2) > 0$ , esto pues,  $y_0, y_1, y_2 > 0$ . Por lo tanto, se tiene que el operador lineal  $T$  es acotado. Ahora probemos que el operador lineal acotado  $T$ , es un operador inyectivo. Sean  $f, g \in C^2[a, b]$  supongamos que:

$$T(f) = T(g);$$

entonces, probemos que se cumple:

$$f = g.$$

Recordemos que dos funciones continuas  $f$  y  $g$  son iguales, si tienen el mismo dominio, codominio, y si además, se cumple que:

$$f(x) = g(x), \quad \forall x \in [a, b].$$

En este caso, ambas funciones tienen el mismo dominio y el mismo codominio. Dado que  $T(f) = T(g)$ , entonces se debe cumplir:

$$(y_0 f'' + y_1 f' + y_2 f, f(a), f(b)) = (y_0 g'' + y_1 g' + y_2 g, g(a), g(b));$$

esto significa que:

$$y_0 f'' + y_1 f' + y_2 f = y_0 g'' + y_1 g' + y_2 g \quad (1.11)$$

$$f(a) = g(a) \quad (1.12)$$

$$f(b) = g(b) \quad (1.13)$$

Si denotamos como  $w = f - g$ , entonces obtenemos la ecuación diferencial:

$$y_0 w'' + y_1 w' + y_2 w = 0 \quad (1.14)$$

Observemos que la ecuación diferencial (1.14) de segundo orden es lineal con coeficientes constantes, y como de la ecuación (1.12) se tiene que  $w(a) = 0$ , entonces se satisfacen todas las condiciones del Teorema de Existencia y Unicidad para ecuaciones diferenciales de grado  $n$ , en consecuencia  $w = 0$ , es la única solución en cualquier intervalo que contenga el valor de  $a$ . Por lo tanto, tendremos que:

$$f - g = 0,$$

entonces,

$$f = g.$$

Con esto queda demostrado que el operador lineal acotado  $T$  es inyectivo. Notemos que el operador  $T$  también es sobreyectivo, pues para cada  $f \in C^2[a, b]$  existe una única  $f''$ ,  $f'$ , y las evaluaciones  $f(a)$  y  $f(b)$ , tales que:

$$T(f) = (y_0 f'' + y_1 f' + y_2 f, f(a), f(b)).$$

De todo lo anterior, tenemos que el operador  $T$  es lineal, acotado, inyectivo y sobreyectivo, esto es,  $T$  es biyectivo. Por lo tanto, el operador  $T$  cumple con todas las hipótesis del Teorema 1.23. Así, se puede concluir que el operador  $T$  es abierto y el operador inverso  $T^{-1}$  es continuo, es decir, acotado.

Que el operador  $T^{-1}$  sea continuo, garantiza que el Ejemplo 1.5 sea un problema bien planteado, lo cual nos indica que la solución depende continuamente de los datos. Los ejemplos 1.4 y 1.5 nos dan una idea de la importancia de los cambios que puedan surgir en la solución, respecto a los cambios en los datos de un problema. En el problema de la tesis, se estudia un problema inverso de la forma  $Ax = b$ , en el cual resultan muy importantes los cambios en los datos  $b$  y en la solución  $x$ , es decir, es necesario encontrar condiciones para garantizar el buen planteamiento de nuestro problema. Para esto, estudiemos el condicionamiento de la matriz  $A$  en la siguiente sección.

## 1.7. Condicionamiento de una matriz

En esta sección, nos interesa el estudio de los cambios que puede sufrir la solución de un problema cuando los datos de entrada son modificados ligeramente. En el estudio de dichos cambios, el número de condición de una matriz  $A$ , juega

un papel muy importante, pues es el principal indicador de la estabilidad de la solución del sistema  $Ax = b$ , es decir, es una medida de la sensibilidad del sistema con respecto a los errores en los coeficientes de la matriz  $A$  y las componentes de  $b$ . En general, los problemas inversos resultan mal condicionados o mal planteados, este mal condicionamiento se debe al incumplimiento de alguna de las propiedades del Teorema 1.36 que son existencia, unicidad y estabilidad ([33]). Normalmente los problemas que están mal condicionados resultan en un número de condición muy grande, lo cual no sucede cuando el número de condición resulta muy cercano a uno. Si se requiere más información de los resultados enunciados en esta sección puede consultar [33, 39, 8].

### **Perturbación del término independiente $b$ .**

Según [39], un sistema de ecuaciones lineales  $Ax = b$  se dice bien condicionado si pequeños cambios en la matriz  $A$  o en el vector  $b$  generan pequeños cambios en la solución; y se dice mal condicionado cuando pequeñas perturbaciones en la matriz  $A$  o el vector  $b$  producen cambios relativamente grandes en la solución exacta. La Definición 1.5 nos permite establecer si el sistema que modela el problema inverso está mal o bien condicionado. En la práctica el incumplimiento de la condición de existencia se debe a que no existe un modelo matemático que ajuste perfectamente un conjunto de datos dados, principalmente si el modelo matemático de la física del problema es aproximado o los datos tienen errores ([33]). La existencia de una solución no es problema insuperable, pues siempre se puede forzar esta condición extendiendo el espacio de soluciones, ajustando los datos en algún sentido aproximado ([33]). Resulta un problema mayor la existencia de múltiples soluciones, pues resulta difícil elegir cual de todas las soluciones es la más adecuada. El incumplimiento de la condición de unicidad se debe en algunos casos a que en la práctica sólo una cantidad limitada de datos esta disponible para reconstruir el modelo. Por lo general, el incumplimiento de la condición de estabilidad se da en muchos problemas inversos, esto se debe a que el problema inverso depende continuamente de los datos. Una manifestación de inestabilidad en problemas inversos discretos puede observarse a través del llamado número de condición de la matriz del sistema ([33]).

Consideremos el siguiente sistema de ecuaciones lineales:

$$Ax = b, \text{ con } A \in M_{m \times n}(\mathbb{C}), \quad x \in \mathbb{C}^n, \quad b \in \mathbb{C}^m .$$

En este caso, estudiaremos los cambios producidos en la solución exacta  $x_0$ , cuando se realizan pequeños cambios en el vector  $b$ . Supongamos que existe solución única  $x_0 \neq 0$ , la cual cumple que:

$$Ax_0 = b. \quad (1.15)$$

Dado que, se quieren estudiar las variaciones en el vector solución  $x_0$  ante variaciones del vector  $b$ , se considera el sistema  $Ax = b + \Delta b$ , el cual se asume tendrá solución única  $x_0 + \Delta x_0$ .

De aquí, se obtiene la siguiente ecuación:

$$A(x_0 + \Delta x_0) = b + \Delta b. \quad (1.16)$$

Restando las ecuaciones (1.15) y (1.16) se obtiene lo siguiente:

$$\begin{aligned} Ax_0 - A(x_0 + \Delta x_0) &= b - (b + \Delta b) \\ Ax_0 - Ax_0 - A\Delta x_0 &= b - b - \Delta b \\ A\Delta x_0 &= \Delta b \end{aligned} \quad (1.17)$$

Por lo tanto, para las ecuaciones (1.16) y (1.17) se cumple que:

$$\frac{A\Delta x_0}{\Delta x_0} = \frac{\Delta b}{\Delta x_0} \quad y \quad \frac{Ax_0}{x_0} = \frac{b}{x_0}.$$

Luego, al tomar la norma vectorial  $\|\cdot\|_m$ , y  $\|\cdot\|_n$  en cada una de las ecuaciones, obtenemos lo siguiente:

$$\frac{\|A\Delta x_0\|_m}{\|\Delta x_0\|_n} = \frac{\|\Delta b\|_m}{\|\Delta x_0\|_n} \quad y \quad \frac{\|Ax_0\|_m}{\|x_0\|_n} = \frac{\|b\|_m}{\|x_0\|_n}. \quad (1.18)$$

Por otro lado, por la ecuación (1.4), podemos definir lo siguiente:

$$\begin{aligned} M &= \sup_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|_m}{\|x\|_n} = \max_{\|x\|_n=1} \|Ax\|_m, \\ m &= \inf_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|_m}{\|x\|_n} = \min_{\|x\|_n=1} \|Ax\|_m. \end{aligned}$$

De aquí, se tiene que para todo  $x \neq 0$ , en particular para  $x_0$ , se cumple que:

$$M \geq \frac{\|Ax_0\|_m}{\|x_0\|_n} \quad y \quad m \leq \frac{\|A\Delta x_0\|_m}{\|\Delta x_0\|_n}.$$

Luego, por las igualdades 1.18 se obtiene:

$$M \geq \frac{\|b\|_m}{\|x_0\|_n} \quad y \quad m \leq \frac{\|\Delta b\|_m}{\|\Delta x_0\|_n}.$$

De todo lo anterior,

$$\frac{1}{\|x_0\|_n} \leq \frac{M}{\|b\|_m} \quad y \quad \|\Delta x_0\|_n \leq \frac{\|\Delta b\|_m}{m}.$$

Multiplicando ambas desigualdades, obtenemos:

$$\frac{\|\Delta x_0\|_n}{\|x_0\|_n} \leq \frac{M}{m} \frac{\|\Delta b\|_m}{\|b\|_m}. \quad (1.19)$$

En consecuencia, el valor de  $\frac{M}{m}$  proporciona una cota superior para el crecimiento del error relativo  $\frac{\|\Delta x_0\|_n}{\|x_0\|_n}$ . Esto se verifica con la siguiente definición.

**Definición 1.37.** El valor  $\frac{M}{m}$ , es llamado número de condición de la matriz  $A$  y se define como:

$$c(A) = \frac{\max_{\|x\|_n=1} \|Ax\|_m}{\min_{\|x\|_n=1} \|Ax\|_m}. \quad (1.20)$$

El número de condición  $c(A)$ , proporciona un límite para el crecimiento del error relativo, es decir, nos da un control de los cambios que se producen en la solución exacta  $x_0$ , cuando realizamos cambios en el término independiente  $b$ .

**Observación 1.3.** Por la Proposición 1.3 se tiene que si  $\sigma_1$  y  $\sigma_n$  son, respectivamente, el mayor y el menor valor singular de  $A$ , entonces el número de condición  $c(A)$  se puede reescribir como:

$$c(A) = \frac{\max_k |\sigma_k|}{\min_k |\sigma_k|} = \frac{\sigma_1}{\sigma_n}.$$

En el caso de que la matriz  $A$  sea regular, entonces su número de condición se puede definir como sigue.

**Definición 1.38.** Sea  $\|\cdot\|$  una norma matricial inducida y  $A$  una matriz regular. Entonces el número de condición de la matriz  $A$  se define y se denota como:

$$c(A) = \|A\| \|A^{-1}\| \quad (1.21)$$

Si  $A$  no es regular, se define  $c(A) = +\infty$ .

**Observación 1.4.** En general, el número de condición de una matriz  $A$  cumple lo siguiente:

1.  $c(A) \geq 1$ .
2. Si  $c(A) \approx 1$ , entonces, se dice que  $A$  es una matriz bien condicionada; y si  $c(A) \gg 1$ , entonces, se dice que  $A$  es una matriz mal condicionada.

Considere el siguiente ejemplo donde la matriz  $A \in M_{n \times n}(\mathbb{R})$  es regular y la norma vectorial que induce la norma matricial es la norma vectorial euclidiana  $\|x\|_2$ .

**Ejemplo 1.6.** *Perturbación de la matriz  $A \in M_{n \times n}(\mathbb{R})$*

Consideremos el siguiente sistema de ecuaciones lineales  $Ax = b$  con:

$$A = \begin{pmatrix} 9 & -1 & 15 \\ 3 & 0 & 5 \\ 2 & 5 & 4 \end{pmatrix} \in M_{3 \times 3}(\mathbb{R}) \quad y \quad b = \begin{pmatrix} 11 \\ 17 \\ -3 \end{pmatrix} \in \mathbb{R}^3.$$

Resolvemos el sistema por medio de la Regla de Cramer, para esto calculamos los siguientes determinantes:

$$\det(A) = \begin{vmatrix} 9 & -1 & 15 \\ 3 & 0 & 5 \\ 2 & 5 & 4 \end{vmatrix} = 2$$

$$\det(x_1) = \begin{vmatrix} 11 & -1 & 15 \\ 17 & 0 & 5 \\ -3 & 5 & 4 \end{vmatrix} = 1083$$

$$\det(x_2) = \begin{vmatrix} 9 & 11 & 15 \\ 3 & 17 & 5 \\ 2 & -3 & 4 \end{vmatrix} = 80$$

$$\det(x_3) = \begin{vmatrix} 9 & -1 & 11 \\ 3 & 0 & 17 \\ 2 & 5 & -3 \end{vmatrix} = -643$$

Luego, tendremos lo siguiente:

$$x_1 = \frac{\det(x_1)}{\det(A)} = \frac{1083}{2} = 541.5$$

$$x_2 = \frac{\det(x_2)}{\det(A)} = \frac{80}{2} = 40$$

$$x_3 = \frac{\det(x_3)}{\det(A)} = \frac{-643}{2} = -321.5$$

Por lo tanto, la solución exacta al sistema de  $3 \times 3$  es:

$$x_0 = \begin{pmatrix} 541.5 \\ 40 \\ -321.5 \end{pmatrix}.$$

Ahora, modificamos el coeficiente  $a_{11}$  de la matriz  $A$  por el valor 9.1.

Así, tendremos el sistema  $(A + \Delta A)x = b$ , donde:

$$A = \begin{pmatrix} 9.1 & -1 & 15 \\ 3 & 0 & 5 \\ 2 & 5 & 4 \end{pmatrix} \quad y \quad b = \begin{pmatrix} 11 \\ 17 \\ -3 \end{pmatrix}.$$

Resolvemos el sistema modificado por medio de la Regla de Cramer. Para esto

calculamos los siguientes determinantes:

$$\det(A + \Delta A) = \begin{vmatrix} 9.1 & -1 & 15 \\ 3 & 0 & 5 \\ 2 & 5 & 4 \end{vmatrix} = -0.5$$

$$\det(x_1) = \begin{vmatrix} 11 & -1 & 15 \\ 17 & 0 & 5 \\ -3 & 5 & 4 \end{vmatrix} = 1083$$

$$\det(x_2) = \begin{vmatrix} 9.1 & 11 & 15 \\ 3 & 17 & 5 \\ 2 & -3 & 4 \end{vmatrix} = 88.3$$

$$\det(x_3) = \begin{vmatrix} 9.1 & -1 & 11 \\ 3 & 0 & 17 \\ 2 & 5 & -3 \end{vmatrix} = -651.5$$

De aquí, obtenemos lo siguiente:

$$x_1 = \frac{\det(x_1)}{\det(A + \Delta A)} = \frac{1083}{-0.5} = -2166$$

$$x_2 = \frac{\det(x_2)}{\det(A + \Delta A)} = \frac{88.3}{-0.5} = -176.6$$

$$x_3 = \frac{\det(x_3)}{\det(A + \Delta A)} = \frac{-651.5}{-0.5} = 1303$$

De donde se obtiene que la solución al sistema alterado  $(A + \Delta A)x = b$ , es:

$$x_0 + \Delta x_0 = \begin{pmatrix} -2166 \\ -176.6 \\ 1303 \end{pmatrix}.$$

Observemos que un cambio menor en un coeficiente de la matriz  $A$  produce un

cambio muy significativo en la solución al sistema. Esto es un indicio de que el sistema  $Ax = b$ , es mal condicionado. Por lo que los resultados obtenidos con este tipo de sistemas no son confiables para la toma de decisiones.

Una forma de saber si nuestro sistema de ecuaciones es mal condicionado es realizando el cálculo del número de condición de la matriz  $A$ , el cual se definió en la ecuación 1.21. En este caso, se tiene que  $A$  es una matriz regular, por lo que se cumple que el  $\det(A) \neq 0$ , y existe matriz inversa  $A^{-1}$  dada por:

$$A^{-1} = \begin{pmatrix} -12.5 & 39.5 & -2.5 \\ -1 & 3 & 0 \\ 7.5 & -23.5 & 1.5 \end{pmatrix}.$$

Considerando la norma vectorial euclidiana aplicada a matrices tendremos que el número de condición de la matriz  $A$  está dada por:

$$\begin{aligned} c(A) &= \|A\|_2 \|A^{-1}\|_2 \\ &= (\sqrt{386}) (\sqrt{2343.5}) \\ &= (19.647)(48.410) \\ &= 951.111. \end{aligned}$$

Se puede observar que  $c(A) \gg 1$ , por lo que la matriz  $A$  es mal condicionada.

A continuación estudiamos un ejemplo en donde la perturbación del sistema  $Ax = b$  se da en el vector  $b$ .

**Ejemplo 1.7.** *Perturbación del vector  $b$ .*

Consideremos el sistema  $Ax = b$  con:

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix} \in M_{4 \times 4}(\mathbb{R}) \quad y \quad b = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix} \in \mathbb{R}^4.$$

Realizamos el cálculo de los siguientes determinantes, así se obtiene que:

$$\Delta_s = \begin{vmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{vmatrix} = 1$$

$$\Delta_{x_1} = \begin{vmatrix} 23 & 7 & 8 & 7 \\ 32 & 5 & 6 & 5 \\ 33 & 6 & 10 & 9 \\ 31 & 5 & 9 & 10 \end{vmatrix} = 1$$

$$\Delta_{x_2} = \begin{vmatrix} 10 & 23 & 8 & 7 \\ 7 & 32 & 6 & 5 \\ 8 & 33 & 10 & 9 \\ 7 & 31 & 9 & 10 \end{vmatrix} = 1$$

$$\Delta_{x_3} = \begin{vmatrix} 10 & 7 & 23 & 7 \\ 7 & 5 & 32 & 5 \\ 8 & 6 & 33 & 9 \\ 7 & 5 & 31 & 10 \end{vmatrix} = 1$$

$$\Delta_{x_4} = \begin{vmatrix} 10 & 7 & 8 & 23 \\ 7 & 5 & 6 & 32 \\ 8 & 6 & 10 & 33 \\ 7 & 5 & 9 & 31 \end{vmatrix} = 1$$

Luego, usando la regla de Cramer obtenemos que la solución exacta del sistema es:

$$x = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Supongamos que existe una variación en el vector  $b$ , dada por:

$$\Delta b = \begin{pmatrix} 0.1 \\ -0.1 \\ 0.1 \\ -0.1 \end{pmatrix}.$$

Luego, al aplicar nuevamente la regla de Cramer y resolver el sistema, obtenemos que la solución del sistema alterado  $Ax = b + \Delta b$  es:

$$x + \Delta x = \begin{pmatrix} 9.2 \\ -12.6 \\ 4.5 \\ -1.1 \end{pmatrix}.$$

Calculamos los errores relativos, respecto a la norma vectorial euclidiana:

$$\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}.$$

Así, tendremos lo siguiente:

$$\frac{\|\Delta x\|}{\|x\|} = \frac{\sqrt{268.86}}{\sqrt{4}} = 8.1985 \approx 8.2,$$

$$\frac{\|\Delta b\|}{\|b\|} = \frac{\sqrt{0.04}}{\sqrt{3603}} = 0.0033 \approx 0.003.$$

Notemos que el error relativo respecto a la solución  $x$  del sistema es demasiado grande, lo cual nos indica que existe una diferencia importante entre la solución exacta y la solución del sistema alterado. De aquí, que un cambio pequeño  $\Delta b$  en el vector  $b$ , resulta en cambios muy significativos en la solución  $x$ , por lo que se dice que el sistema está mal condicionado. Por lo tanto, los resultados obtenidos con este sistema resultan no ser confiables para la toma de decisiones.

Calculemos ahora el número de condición de la matriz  $A$ , para esto consideremos

la matriz inversa de  $A$  dada por:

$$A^{-1} = \begin{pmatrix} 25 & -41 & 10 & -6 \\ -41 & 68 & -17 & 10 \\ 10 & -17 & 5 & -3 \\ -6 & 10 & -3 & 2 \end{pmatrix}.$$

Dado que la matriz  $A$  es regular, entonces es posible usar la ecuación (1.21) para hallar el número de condición de  $A$ , luego se tiene que:

$$\begin{aligned} c(A) &= \|A\|_2 \|A^{-1}\|_2 \\ &= (\sqrt{933}) (\sqrt{9708}) \\ &= (30.545)(98.529) \\ &= 3009.568. \end{aligned}$$

Como  $c(A) \gg 1$ , entonces la matriz  $A$  es mal condicionada, y esto nos indica que pequeños cambios en el vector  $b$  del sistema resultan en cambios grandes en la solución.

Para ahorrarnos cálculos al decidir si el sistema está mal condicionado, el número de condición de la matriz  $A$  juega un papel muy importante, pues apartir de él podemos saber si tratamos o no con un sistema de ecuaciones mal condicionado. En esta tesis se trabajará con sistemas mal condicionados, en los que se perturba a los datos, pero no se perturba a la matriz.

# Capítulo 2

## Planteamiento del problema inverso de PSD

En muchos productos y procesos de fabricación se utilizan partículas. En la industria de los alimentos muchos de estos están en forma de polvos secos o molidos. En la industria farmacéutica, en específico, las soluciones farmacéuticas líquidas son entregadas en forma de emulsiones, como cremas para el cuerpo, ungüentos o pomadas, jarabes, o alimentos como la mayonesa, la leche y el yogurt, donde el tamaño de la gotas es importante. Incluso partículas biológicas como exosomas<sup>1</sup> y virus que fluyen en forma de partículas hacia y a través de nuestras venas ([13]). Es decir, los sistemas particulados forman parte de nuestra vida cotidiana, de ahí la importancia de la estimación de tamaños de partículas en materiales como polvos, emulsiones, suspensiones y aerosoles, esto pues, el tamaño es un indicador valioso de calidad y rendimiento del material ([13]). Por ejemplo, el polvo o las gotas en el rango de 2-5  $nm$  ( $1\ nm = 10^{-9}m$ ) se aerosolizan mejor y penetran los pulmones más profundamente que los tamaños más grandes ([13]). El tamaño y la forma de los polvos influyen en las propiedades de fluidez y compactación<sup>2</sup>, esto pues, las partículas más grandes y esféricas fluirán típicamente más fácilmente que las partículas más pequeñas ([13]). Los análisis de distribución de tamaños de partículas se utilizan en mezclas de especias y aromas, como por ejemplo el clavo, la pimienta y el orégano, para conocer los tamaños de las partículas que las componen y sus proporciones, así como para determinar su homogeneidad ([14]).

---

<sup>1</sup>Estructura diminuta en forma de bolsa que se forma dentro de una célula y contiene algunas de las proteínas celulares, así como porciones del ADN y ARN de la célula.

<sup>2</sup>Procedimiento de aplicar energía al suelo suelto para eliminar espacios vacíos, aumentando así su densidad y en consecuencia, su capacidad de soporte y estabilidad entre otras propiedades.

Por estas y por otras razones es importante medir y controlar el tamaño y la forma de las partículas, ya que se utilizan como control de calidad.

Existen distintas técnicas para la estimación de distribuciones de tamaño de partículas, muchas de estas técnicas deben considerar la naturaleza dispersa de las poblaciones de partículas. Los métodos más conocidos son el tamizado, las mallas, la microscopía, la sedimentación, difracción láser y técnicas de dispersión de luz. En este trabajo se usan técnicas de dispersión de luz, las cuales son herramientas importantes para el estudio de las características de sistemas heterogéneos, como los sistemas de partículas inmersas en un medio no absorbente. Estas técnicas miden simultáneamente gran número de partículas basados en la teoría de la dispersión de luz. En específico usaremos la técnica óptica de medición llamada turbidimetría, con la cual es posible medir el espectro de turbidez, y a partir de la teoría de Mie relacionar el espectro de turbidez con la distribución de tamaño de partículas (PSD).

## 2.1. La turbidimetría como técnica óptica de medición

La dispersión de luz puede considerarse como la redirección de la luz que tiene lugar cuando una onda electromagnética (EM) o rayo de luz incidente, encuentra un obstáculo (o una falta de homogeneidad). Un experimento típico de dispersión de luz se muestra en la Figura 2.1. Según [33], cuando la luz incidente reacciona con la muestra, la componente del campo eléctrico de la radiación induce una polarización oscilante de los electrones en las partículas. Entonces, las partículas actúan como fuentes de luz secundaria y en consecuencia son fuentes de radiación dispersada. La luz dispersada a distintos ángulos es registrada por un detector, el cual nos brinda información sobre el movimiento y la estructura de las partículas en el material ([33]). El detector puede registrar tres vectores  $I_s(\theta)$ ,  $I_s(t)$  e  $I_s(\lambda_j)$ , de diferentes técnicas, los cuales registran la intensidad de luz después de atravesar en forma de onda la muestra de partículas.

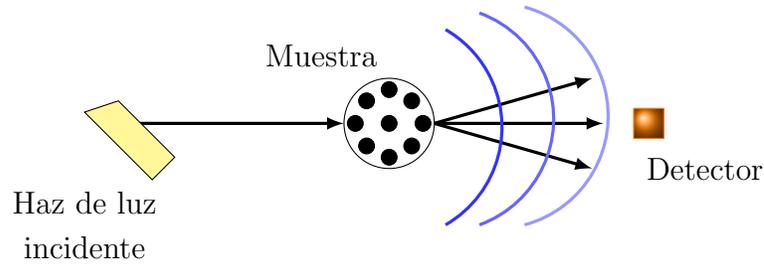


Figura 2.1: Experimento de dispersión de luz.

Las técnicas más empleadas para extraer información de un experimento de dispersión de luz son: *Dispersión de Luz Dinámica (DLD)*, *Dispersión de Luz Estática (DLE)* y *Turbidimetría (T)*. La técnica de Dispersión de Luz Dinámica, consiste en monitorear las variaciones en la intensidad de luz dispersada, denotada por  $I_s$ . Dichas variaciones se deben a que las pequeñas moléculas en la solución se encuentran bajo el movimiento browniano<sup>3</sup>, esto es, la distancia entre los dispersores de luz está cambiando constantemente con el tiempo ([33]). Así, el vector de intensidad de luz dependerá de la variable  $t$ , con lo que obtenemos el vector  $I_s(t)$ . La técnica de *Dispersión de Luz Estática* observa patrones de interferencia de luz entre las partículas midiendo la intensidad de luz como función del ángulo  $\theta$ , por lo que el vector de intensidad de luz está dado por  $I_s(\theta)$  ([33]). La Turbidimetría es una técnica analítica óptica que se basa en el fenómeno físico que ocurre cuando aparecen partículas sólidas en una disolución homogénea ([9]). La pérdida de homogeneidad hace que la luz que atraviesa la disolución no sea de igual intensidad que antes de aparecer la turbidez ([9]). Existen sistemas para la medición de turbidez que operan sobre la base de fenómenos ópticos que se producen al incidir un haz de luz a través de un medio ([1]). La luz se dispersa debido a la existencia de partículas suspendidas en el interior del medio en análisis, que también intervienen en la reducción de la intensidad del haz de luz resultante ([1]). La manera en que la muestra interfiere con la luz está relacionada con el tamaño, la forma, la composición de las partículas en la suspensión y la longitud de onda de la luz incidente ([1]). La estimación de la PSD a partir de mediciones de turbidimetría, requiere la resolución de un problema inverso mal condicionado basado en el modelo matemático que describe la interacción entre partículas y el espectro de turbidez ([33]). La teoría de dispersión de Mie o teoría de dispersión

<sup>3</sup>Movimiento aleatorio que se observa en las partículas que se hallan en un medio fluido (líquido o gas), como resultado de choques contra moléculas de dicho fluido.

esférica, describe la interacción de una onda plana con un conjunto de partículas esféricas y relaciona la distribución de tamaños de partícula con la medición de turbidimetría y el espectro de turbidez ([7]). Para el estudio de la relación entre las partículas y el espectro de turbidez, es necesario medir este espectro haciendo un experimento de turbidimetría por medio de un espectrofotómetro.

Un experimento de turbidimetría como el que se muestra en [7] y [18], consiste en medir a distintas longitudes de onda  $\lambda_i$  ( $i = 1, \dots, m$ ), la atenuación de un haz de luz (fotones) al atravesar un conjunto de partículas esféricas inmersas en un medio no absorbente. En principio, un haz de luz pasa a través de un monocromador el cual divide en un espectro visible el haz incidente, es decir, en sus colores componentes: rojo, anaranjado, amarillo, verde, azul ([44]). Esto se debe a un fenómeno llamado dispersión. Cuando el haz incidente va del aire hacia el interior del cristal y del interior del cristal hacia el aire, el haz de luz es separado en sus colores componentes debido a que cada color presenta un índice de refracción ligeramente diferente. Cada color representa una longitud de onda, se selecciona una determinada, la cual recorre un camino óptico  $\mathcal{L}$  (grosor del tubo de ensayo) e incide sobre la muestra a analizar (ver Figura 2.2). Finalmente, el fotómetro o detector mide la intensidad de luz transmitida  $I(\lambda_i)$ , a través de la muestra ([18]).

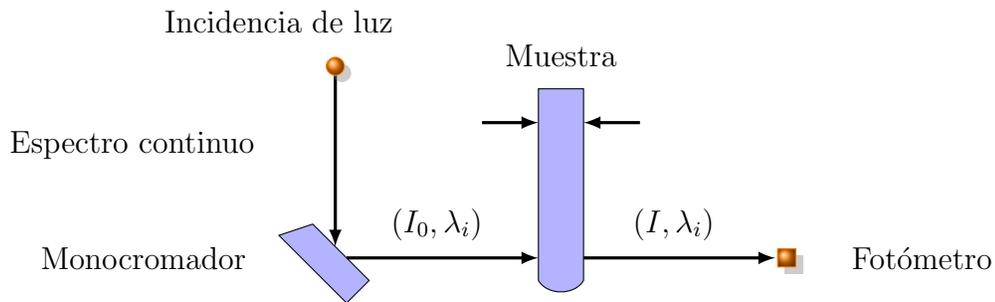


Figura 2.2: Esquema de un espectrofotómetro para medición de la turbidez.

Al medir la intensidad de luz transmitida respecto a una longitud de onda, se puede obtener un valor proporcional a la concentración<sup>4</sup> de la sustancia o sustancias presentes en la disolución responsable de la turbidez ([9]). Mediante el conocimiento de la intensidad de luz transmitida  $I(\lambda_i)$ , es posible calcular el espectro de turbidez que proviene de la Ley de Beer-Lambert, la cual relaciona la

<sup>4</sup>Concentración de una disolución: Según [6], la concentración de una disolución es la cantidad de soluto (sustancia disuelta) que hay en una cantidad de disolvente o bien de disolución.

absorción de la radiación electromagnética de una o varias especies químicas, con su concentración y la distancia que recorre la luz en las interacciones partículas-fotón.

La turbidez se define por la Organización Internacional de Normalización (ISO), como la reducción de la transparencia de un líquido causada por la presencia de partículas no disueltas de material distinto al propio líquido ([1]). La medición de la turbidez tiene diferentes aplicaciones, una de las más importantes es el análisis de la calidad del agua. Un experimento de turbidimetría realizado en un espectrofotómetro como el de la Figura 2.2, consiste básicamente en un haz de luz que pasa a través de un monocromador (prisma), el cual separa los colores de el haz incidente, refractándolos en diferentes direcciones. El esparcimiento angular de todos los colores producidos al enviar luz blanca a través de un prisma, se llama dispersión y la banda de colores así producida se denomina espectro ([44]); por lo que el espectro, después de atravesar la muestra, es llamado espectro de turbidez. Cuando el espectro de luz atraviesa la muestra ocurren diferentes fenómenos físicos como la absorción, fluorescencia o la dispersión que junto con las propiedades ondulatorias de la luz, como su amplitud, fase, polarización y longitud de onda ofrecen información importante de la muestra a estudiar y del medio en la que está inmersa ([1]).

En la siguiente sección se analiza la medición del espectro de turbidez y se dan a conocer los conceptos físicos involucrados en dicha medición.

### 2.1.1. Medición del espectro de turbidez

En un experimento de turbidimetría realizado en un espectrofotómetro se obtiene la medición del llamado espectro de turbidez. Dado que se trata de un indicador de apariencia óptica, ocasionado por la dispersión y absorción de la energía lumínica cuando pasa a través del líquido, entonces la turbidez solo puede ser medida con técnicas ópticas ([1]). Una técnica para el estudio de la interacción entre la radiación electromagnética y la materia, es la espectroscopía. Según [36], la espectroscopía es el estudio de la interacción de la radiación electromagnética con la materia. Cuando la materia es excitada aplicando energía térmica, eléctrica, nuclear o radiante, la materia se relaja y vuelve a su estado fundamental, es decir, a su estado de energía más bajo o natural, emitiendo radiación electromagnética, dando lugar a los espectros de emisión ([36]). También es posible estudiar la interacción de la radiación electromagnética con la materia, observando cuales son

las frecuencias absorbidas, dichas frecuencias aparecerán en el espectro, por lo que estos espectros se les conoce como espectros de absorción. En general, un espectro puede definirse como una representación gráfica de la distribución de la intensidad de la radiación electromagnética, emitida o absorbida por una sustancia, en función de la longitud de onda o la frecuencia de dicha radiación ([36]). Existen tres tipos de espectros: continuo, lineal, y banda; en el caso de un experimento de turbidimetría se trata de un espectro continuo, el sol y el arcoiris son otros ejemplos de espectros continuos. La Espectroscopía, dependiendo de la materia bajo estudio, se puede dividir en dos grandes ramas: atómica y molecular, en el caso de las moléculas, es a través de su momento dipolar o momento magnético como interaccionan con la radiación electromagnética ([36]); al momento de la interacción de la radiación con un sistema molecular este absorbe energía de la radiación, experimentando una transición desde un nivel de energía inferior a uno de energía superior, lo que origina un espectro de absorción ([36]). Existen otro tipo de espectros, los espectros de dispersión, los cuales se originan cuando el sistema molecular dispersa la radiación, de forma que una parte muy pequeña de fotones, al interaccionar inelásticamente con la materia no conserva su energía; es decir, durante el proceso de dispersión parte de la energía del fotón se transfiere a la molécula y viceversa, originándose el denominado espectro de dispersión ([36]). La medición del espectro de turbidez por medio de la turbidimetría se fundamenta en la espectroscopía, principalmente en la relación entre la intensidad de luz incidente y la intensidad de luz dispersada por el medio, las cuales se relacionan mediante la ley de Lambert-Beer, en la que la turbidez es proporcional a la concentración de partículas ([1]). Esta ley establece que una muestra absorberá mayor radiación cuando las dimensiones del material en el que se encuentran las partículas y el medio no absorbente sean mayores, específicamente, el grosor del material  $\mathcal{L}$ , que recorre la luz al entrar y salir de la muestra ([18]). Un haz de radiación monocromática paralelo con intensidad de radiación  $I_0$  llega al medio absorbente perpendicular a la superficie, luego pasa a través de la longitud del medio  $\mathcal{L}$ , que contiene partículas absorbentes que bloquearán la transmisión de la luz (átomos, iones o moléculas), la intensidad del haz disminuye a  $I$  como resultado de la absorción y la dispersión ([1]). Dado que un experimento de turbidimetría está basado en mediciones diferentes de longitudes de onda en el vacío, entonces la intensidad incidente y transmitida dependerán de la longitud de onda  $\lambda$  elegida.

La absorción de radiación cuando un haz de luz atraviesa un soluto <sup>5</sup> absorbente en una disolución no absorbente se obtiene de la ley Beer-Lambert ([36]). Según esta ley, el decrecimiento producido en la intensidad de la radiación, al penetrar una distancia  $d\mathcal{L}$  en una disolución, es directamente proporcional a la intensidad de la radiación  $I(\lambda)$ , a la concentración de la disolución y a la longitud del camino recorrido, es decir, se cumple que ([36]):

$$\frac{dI(\lambda)}{d\mathcal{L}} = -\alpha(\lambda)I(\lambda); \quad (2.1)$$

donde la constante de proporcionalidad,  $\alpha(\lambda)$ , recibe el nombre de coeficiente de absorción y depende de la concentración de la muestra y es función de la frecuencia de radiación usada, en este caso la longitud de onda  $\lambda$ . La ecuación (2.1) es una ecuación diferencial ordinaria de primer orden, por lo que su solución es de la forma:

$$I(\lambda) = I_0(\lambda)e^{-\alpha(\lambda)\mathcal{L}}. \quad (2.2)$$

Si se considera que el material se compone de partículas con diferentes coeficientes de absorción y de dispersión, el coeficiente total de absorción y de dispersión es igual a la suma de los coeficientes de absorción y de dispersión de todas las partículas ([1]). Así, podemos reescribir la ecuación (2.2) como:

$$I(\lambda) = I_0(\lambda)e^{-(\alpha A + \alpha B)C\mathcal{L}};$$

donde  $\alpha A$  es el coeficiente de absorción,  $\alpha B$  es el coeficiente de dispersión, y  $C$  es la concentración del medio no absorbente. Cuando el haz de luz atraviesa el fluido de la muestra (Fig. 2.3), las partículas suspendidas en un medio no absorbente dispersan la luz en todas direcciones, por lo que la reducción de la intensidad del haz de luz se debe a la difusión del haz por dichas partículas ([1]). En este trabajo se considera una cantidad finita de partículas suspendidas en el medio, por lo que, según [1], es posible usar la medida de luz dispersada; esto pues, supondremos que el espectrofotómetro que será usado, cuenta con un fotodetector que detecta pequeños cambios en la intensidad de luz en contraste con un fondo oscuro. Además, se considera el fenómeno de absorción de luz al atravesar la muestra y el medio en análisis. La relación directa entre los datos de turbidez y la concentración de

---

<sup>5</sup>En una disolución, el soluto es la sustancia que se disuelve, es decir, se combina con otra sustancia, que es el disolvente.

partículas suspendidas en el medio depende de muchos factores, como el tamaño de las partículas, la forma, la distribución, el índice de refracción de las partículas de dispersión y de la longitud de onda utilizada ([1]). En este trabajo consideraremos un espectrofotómetro que mide la tasa o relación de intensidades luminosas, tanto la transmisión de la intensidad de luz como su dispersión y absorción.

Por lo general la unidad de medida estándar para la turbidez es la Unidad Nefelométrica de Turbidez (UNT), aunque en algunos casos se utilizan otras unidades de medida establecidas por regulaciones y estándares, como Total de Sólidos Suspendidos (SST) y Concentración de Sedimento (SSC) ([1]). La turbidez se mide utilizando las técnicas de turbidimetría y nefelometría. En este trabajo se considera un experimento de turbidimetría, la cual es una técnica simple, rápida, económica, y puede ser aplicada a una gran variedad de tamaños de partículas ([18]).

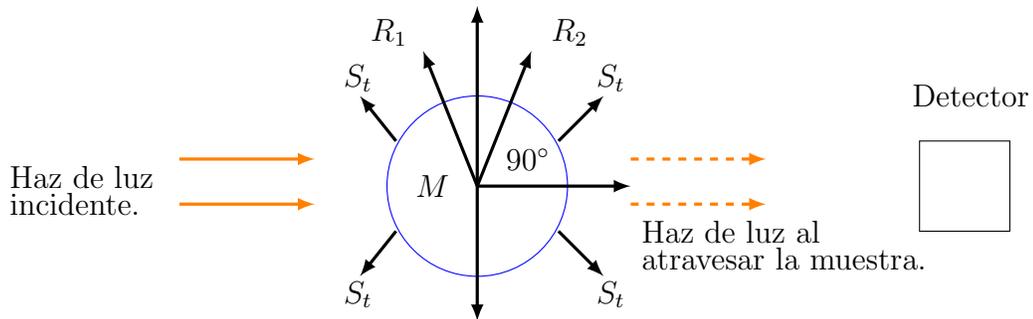


Figura 2.3: Efectos de luz al atravesar la muestra  $M$ .  $S_t$  es la luz difundida por las partículas suspendidas en  $M$ .  $R_1$  y  $R_2$  son los rayos periféricos del haz de luz difundido.

Las mediciones de turbidimetría se realiza en un equipo de medición de turbidez llamado espectrofotómetro como el de la Figura 2.2 el cual se basa en el fenómeno físico que muestra la Figura 2.3. Según [24], dichas mediciones requieren la determinación del espectro de intensidad de luz incidente  $I_0(\lambda_i)$ ; y del espectro de intensidad de luz emergente  $I(\lambda_i)$  (medida en la misma dirección que la intensidad de luz incidente  $I_0$ ).

Llamemos  $\lambda_i$  ( $i = 1, \dots, m$ ) a las longitudes de onda del espectro UV-Vis en el vacío y muestradas a intervalos regulares  $\Delta\lambda$ , entonces, se tiene que el espectro

de turbidez o extinción de luz  $\tau(\lambda_i)$  se puede definir como ([24]):

$$\tau(\lambda_i) = \frac{1}{\mathcal{L}} \log_{10} \left( \frac{I_0(\lambda_i)}{I(\lambda_i)} \right). \quad (2.3)$$

En general, el espectro de turbidez o extinción de luz da una medida de la atenuación de un haz de luz que atraviesa una suspensión de partículas ([21]). De acuerdo a la teoría de Mie, para una suspensión<sup>6</sup> monodispersa<sup>7</sup> de partículas isotrópicas<sup>8</sup> esféricas no absorbentes, en ausencia de dispersión múltiple, es posible definir la relación entre el espectro de turbidez  $\tau(\lambda_i)$  y la distribución de tamaños de partículas a través de la siguiente ecuación ([24, 21]):

$$\tau(\lambda_i) = \frac{\log_{10}(e)\pi\mathcal{L}}{4} \int_0^\infty f(D)Q_{ext}(x_i, m_i)D^2 dD; \quad (2.4)$$

donde  $\log_{10}(e) = 0.434\dots$ ,  $\mathcal{L}$  es el camino óptico recorrido por el haz de luz incidente,  $f(D)$  es la distribución de tamaños de partículas normalizada continua (en número de partículas por  $cm^3$  vs el diámetro de la partícula  $D$ );  $Q_{ext}$  es la eficiencia de la extinción de luz al atravesar la muestra;  $n_m(\lambda_i)$  es la función de índice de refracción del medio;  $n_p(\lambda_i)$  es la función de índice de refracción de la partícula (generalmente compleja);  $m_i$  es el índice refracción relativo; y  $x_i$  es el parámetro de tamaño, estos dos últimos se definen como ([24, 3]):

$$x_i = ka = \pi D \left( \frac{n_m(\lambda_i)}{\lambda_i} \right), \quad (2.5)$$

$$m_i = \frac{n_p(\lambda_i)}{n_m(\lambda_i)}; \quad (2.6)$$

donde  $a$  es el radio de la partícula esférica y  $k$  es el llamado vector de onda, el cual se obtiene al manipular la ecuación (2.5):

$$k = 2\pi \left( \frac{n_m(\lambda_i)}{\lambda_i} \right). \quad (2.7)$$

Para ser mas específicos la atenuación de haz de luz incidente se debe a un

---

<sup>6</sup>Mezcla heterogénea formada por un sólido en polvo o por pequeñas partículas no solubles que se dispersan en un medio líquido.

<sup>7</sup>Caracterizado por partículas de tamaño uniforme en una fase dispersa.

<sup>8</sup>Los materiales isotrópicos tienen propiedades materiales idénticas en todas las direcciones en cada punto dado.

fenómeno físico llamado extinción de luz, el cual se debe a la absorción y dispersión de luz. En la siguiente sección analizaremos dicho fenómeno físico y su relación con el problema inverso de estimación de la PSD; también se define y analiza el coeficiente de extinción  $Q_{ext}$  de acuerdo a la teoría de Mie.

### 2.1.2. Extinción, absorción, y dispersión de luz

En un experimento de turbidimetría se colocan una o más partículas en un haz de radiación electromagnética, como se muestra en la Figura 2.3. La velocidad a la que la energía electromagnética es recibida por un detector  $D$ , pasando a través de un conjunto de partículas se denota por  $U$  ([3]). Si las partículas son eliminadas, el poder recibido por el detector es  $U_0$ , por lo que  $U_0 > U$ . Esto quiere decir que la presencia de las partículas ha provocado la extinción del rayo incidente. Si consideramos que el medio en el que se encuentran las partículas es no absorbente, la diferencia  $U_0 - U$  se puede explicar por la absorción en las partículas (es decir, transformación de energía electromagnética en otras formas) y la dispersión por las partículas ([3]). En general, la extinción es la atenuación de una onda electromagnética por dispersión y absorción a medida que atraviesa un medio particulado que depende de la composición química de las partículas, su tamaño, forma, orientación, el medio circundante, el número de partículas, el estado de polarización y la frecuencia del haz incidente ([3]).

Consideremos la extinción por una sola partícula arbitraria inmersa en un medio no absorbente e iluminada por una onda plana como se muestra en la Figura 2.4. Es posible construir una esfera imaginaria de radio  $r$  alrededor de la partícula; entonces la tasa neta a la que la energía electromagnética atraviesa la superficie  $A$  de esta esfera es ([3]):

$$W_a = - \int_A \vec{S} \cdot \hat{e}_r dA; \quad (2.8)$$

donde  $\vec{S}$  es el vector de Poyting en cualquier punto y  $\hat{e}_r$  es el vector de dirección de la luz dispersada.

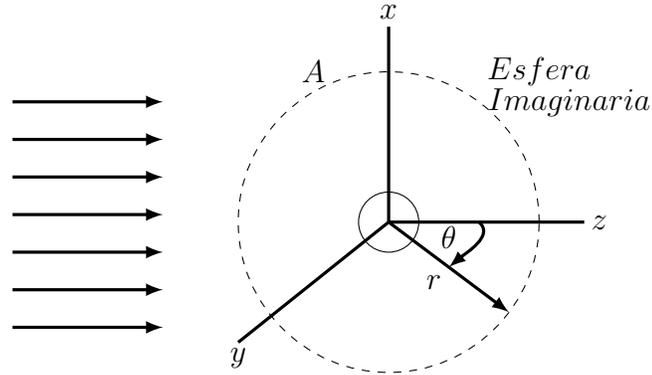


Figura 2.4: Extinción por una partícula arbitraria en un medio no absorbente.

Si  $W_a > 0$ , la energía se absorbe dentro de la esfera, y puesto que el medio es no absorbente, entonces  $W_a$  es la tasa a la que la energía es absorbida por la partícula ([3]). Por otro lado, según [3], el vector de Poynting  $\vec{S}$  promediado en el tiempo en cualquier punto del medio que rodea a la partícula se puede escribir como la suma de tres términos:

$$\vec{S} = \vec{S}_i + \vec{S}_s + \vec{S}_{ext}; \quad (2.9)$$

donde  $\vec{S}_i$ , es el vector de Poynting asociado con la onda incidente, este vector es independiente de la posición si el medio es no absorbente;  $\vec{S}_s$ , es el vector de Poynting del campo dispersado; y se puede interpretar a  $\vec{S}_{ext}$  como el término que surge a partir de la interacción entre las ondas incidente y dispersada. Luego, debido a la ecuación (2.9) es posible escribir  $W_a$  como:

$$W_a = W_i - W_s + W_{ext}; \quad (2.10)$$

donde:

$$W_i = - \int_A \vec{S}_i \cdot \hat{e}_r dA, \quad W_s = \int_A \vec{S}_s \cdot \hat{e}_r dA, \quad W_{ext} = - \int_A \vec{S}_{ext} \cdot \hat{e}_r dA. \quad (2.11)$$

Debido a que la partícula se encuentra en un medio no absorbente la tasa de energía  $W_i$  al interior se puede despreciar y así tendremos que  $W_{ext}$  es la suma de la tasa de absorción de energía  $W_a$  y la tasa de dispersión de energía  $W_s$ :

$$W_{ext} = W_a + W_s. \quad (2.12)$$

Observemos de la Figura 2.4 que la irradiancia incidente  $I_i$  atraviesa la esfera imaginaria alrededor de la partícula como una sección transversal, es decir, como un corte de un plano a una esfera. Así, la relación que existe entre  $W_{ext}$  e  $I_i$  es una cantidad con dimensiones de área, la cual se define como ([24]):

$$C_{ext} = \frac{W_{ext}}{I_i}. \quad (2.13)$$

De la ecuación (2.12) se deduce que la sección transversal de extinción  $C_{ext}$  puede escribirse como la suma de la sección transversal de absorción  $C_{abs}$  y la sección transversal de dispersión  $C_{sca}$  :

$$C_{ext} = C_{abs} + C_{sca}, \quad (2.14)$$

donde  $C_{abs} = W_{abs}/I_i$  y  $C_{sca} = W_s/I_i$ . También es posible definir la eficiencia (o factores de eficiencia de luz) por extinción, dispersión y absorción como ([3]):

$$Q_{ext} = \frac{C_{ext}}{G}, \quad Q_{sca} = \frac{C_{sca}}{G}, \quad Q_{abs} = \frac{C_{abs}}{G}, \quad (2.15)$$

donde  $G$  es el área de la sección transversal de la partícula proyectada en un plano perpendicular a la dirección del haz incidente ([3]). En este trabajo nos interesa conocer específicamente el valor del factor de extinción de luz  $Q_{ext}(x_i, m_i)$  el cual depende del parámetro de tamaño  $x_i$  y el índice de refracción relativo  $m_i$ .

Según [3], la sección transversal de extinción se define como:

$$C_{ext} = \frac{W_{ext}}{I_i} = \frac{2\pi}{k^2} \sum_{n=1}^{\infty} (2n+1) \text{Re}[a_n(x_i, m_i) + b_n(x_i, m_i)], \quad (2.16)$$

donde  $a_n(x_i, m_i)$  y  $b_n(x_i, m_i)$  son los llamados coeficientes de Mie. Así, es posible escribir el coeficiente de extinción  $Q_{ext}$  como:

$$Q_{ext}(x_i, m_i) = \frac{C_{ext}}{G} = \frac{2\pi}{Gk^2} \sum_{n=1}^{\infty} (2n+1) \text{Re}[a_n(x_i, m_i) + b_n(x_i, m_i)]. \quad (2.17)$$

Dado que se consideran partículas esféricas, entonces la sección transversal de la partícula que es proyectada sobre un plano perpendicular a la dirección del haz incidente, es una circunferencia, por lo que  $G = \pi a^2$ . Al sustituir el valor  $G$  y el

valor del vector de onda  $k$  en la ecuación (2.17) tendremos que:

$$\begin{aligned} Q_{ext}(x_i, m_i) &= \frac{2\pi}{\pi a^2 \left( \frac{2^2 \pi^2 n_m^2(\lambda_i)}{\lambda_i^2} \right)} \sum_{n=1}^{\infty} (2n+1) \operatorname{Re} [a_n(x_i, m_i) + b_n(x_i, m_i)] \\ &= \frac{2}{\pi^2 D^2 \left[ \frac{n_m(\lambda_i)}{\lambda_i} \right]^2} \sum_{n=1}^{\infty} (2n+1) \operatorname{Re} [a_n(x_i, m_i) + b_n(x_i, m_i)]. \end{aligned}$$

Por lo tanto, la función de extinción de luz por partículas esféricas se define como:

$$Q_{ext}(x_i, m_i) = \frac{2}{x_i^2} \sum_{n=1}^{\infty} (2n+1) \operatorname{Re} [a_n(x_i, m_i) + b_n(x_i, m_i)]. \quad (2.18)$$

La extinción de luz por partículas es un fenómeno que se observa con bastante frecuencia, por ejemplo: la luz solar a través de una tormenta de polvo o una capa de aire contaminada; faros de automóviles en la niebla; la luz de un buzo en agua turbia. Para entender por completo la función de extinción de luz  $Q_{ext}$  es necesario definir los coeficientes de Mie, y las funciones por las que están compuestos. En la siguiente sección estudiaremos la teoría de Mie para los fenómenos de dispersión y absorción por una esfera, lo cual nos permitirá definir los coeficientes de Mie que son de mucha importancia para el problema inverso de estimación de PSD.

## 2.2. Teoría de Mie

Para el estudio de la dispersión y la absorción de una esfera es importante entender la naturaleza física de estos fenómenos, es decir, es necesario comprender como cambian las cantidades observables con respecto a su tamaño, naturaleza del medio y propiedades ópticas de la esfera ([3]). Uno de los problemas más importante en la teoría de la absorción y la dispersión por partículas pequeñas es el de una esfera de radio e índice de refracción arbitrarios. En 1908, Gustav Mie desarrolló la teoría en un esfuerzo por comprender los variados colores en absorción y dispersión exhibidos por pequeñas partículas coloidales de oro suspendidas en agua ([3]). Casi al mismo tiempo, Peter Debye consideró el problema de la presión de radiación ejercida sobre pequeñas partículas en el espacio. Sin embargo, ni Mie ni Debye fueron los primeros en construir una solución para problema de la esfera, a pesar de eso a la teoría desarrollada para construir la solución es llamada teoría de Mie ([3]).

### 2.2.1. Coeficientes de dispersión de luz

Gustav Mie, llevó a cabo un estudio riguroso del fenómeno de dispersión de luz basado en las ecuaciones de electromagnetismo de Maxwell, obteniendo los coeficientes de extinción de luz  $Q_{ext}$ , y los coeficientes de dispersión  $a_n$  y  $b_n$ , para partículas homogéneas de morfología esférica, de cualquier composición y tamaño, inmersas en un medio homogéneo no absorbente ([7]). A partir de la teoría de Mie, es posible modelar de forma teórica, los fenómenos físicos de dispersión de luz dinámica, dispersión de luz estática, y la turbidimetría, sobre la base de los coeficientes  $Q_{ext}$ ,  $a_n$ , y  $b_n$  ([7]). Una observación importante sobre la teoría de Mie, es que solo es válida bajo condiciones de dispersión simple (o ausencia de dispersión múltiple), es decir, bajo la hipótesis de que la luz dispersada por una partícula no interacciona con ninguna otra partícula del sistema ([7]). En un sistema coloidal<sup>9</sup>, por ejemplo un látex, la hipótesis de dispersión simple es fácilmente alcanzable utilizando bajas concentraciones de la muestra a analizar ([7]).

La idea principal de la teoría de Mie es construir las funciones generadoras  $\psi_{emn}$  y  $\psi_{omn}$  que satisfagan la ecuación de onda en coordenadas polares esféricas:

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial^2 \psi}{\partial \phi^2} + k^2 \psi = 0 \quad (2.19)$$

donde  $\theta$  es el ángulo de dispersión de luz y  $\phi$  es el ángulo azimutal,  $r$  es la variable radial en las coordenadas esféricas de la Figura 2.5.

---

<sup>9</sup>En física y química un coloide, sistema coloidal, suspensión coloidal o dispersión coloidal es un sistema conformado por dos o más fases, normalmente una fluida (líquido o gas) y otra dispersa en forma de partículas generalmente sólidas muy finas, de diámetro comprendido entre  $10^{-9}$  y  $10^{-5}$  m.

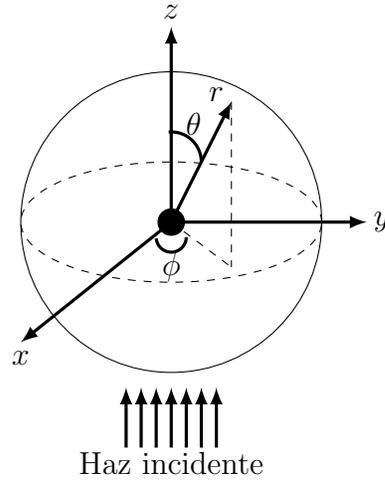


Figura 2.5: Sistema de coordenadas polares esférico centrado en una partícula esférica de radio  $a$ .

Según [3], las funciones generadoras  $\Psi_{emn}$  y  $\Psi_{omn}$  que satisfacen la ecuación de onda en coordenadas polares esféricas, se definen como:

$$\Psi_{emn} = \cos(m\phi) P_n^m(\cos \theta) z_n(\rho), \quad (2.20)$$

$$\Psi_{omn} = \sen(m\phi) P_n^m(\cos \theta) z_n(\rho); \quad (2.21)$$

donde  $P_n^m(\cos \theta)$  son las denominadas funciones asociadas de Legendre de primera clase de grado  $n$  y orden  $m$ ,  $\rho = kr$ , y  $z_n(\rho)$  es alguna de las cuatro funciones esféricas de Bessel  $j_n$ ,  $y_n$ ,  $h_n^{(1)}$  y  $h_n^{(2)}$ .

### Funciones esféricas de Bessel

Las funciones de Bessel forman una clase de función de las denominadas funciones especiales que se encuentran en la solución de determinados problemas físicos. Las funciones esféricas de Bessel  $j_n(\rho)$  y  $y_n(\rho)$ , son soluciones linealmente independientes de la ecuación de Helmholtz en coordenadas esféricas. Las funciones esféricas de Bessel de primer y segundo tipo  $j_n(\rho)$  y  $y_n(\rho)$  se definen como:

$$j_n(\rho) = (-\rho)^n \left( \frac{1}{\rho} \frac{d}{d\rho} \right)^n \frac{\sen \rho}{\rho}, \quad (2.22)$$

$$y_n(\rho) = -(-\rho)^n \left( \frac{1}{\rho} \frac{d}{d\rho} \right)^n \frac{\cos \rho}{\rho}. \quad (2.23)$$

Para  $n = 0, 1$  y  $2$  tenemos:

$$\begin{aligned} j_0(\rho) &= \frac{\text{sen } \rho}{\rho}, \\ j_1(\rho) &= \frac{\text{sen } \rho}{\rho^2} - \frac{\cos \rho}{\rho}, \\ j_2(\rho) &= \frac{3 \text{ sen } \rho}{\rho^3} - \frac{3 \cos \rho}{\rho^2} - \frac{\text{sen } \rho}{\rho}. \end{aligned}$$

Las cuales se obtienen de la siguiente manera:

- Para  $n = 0$  :

$$j_0(\rho) = (-\rho)^0 \left( \frac{1}{\rho} \frac{d}{d\rho} \right)^0 \frac{\text{sen } \rho}{\rho} = \frac{\text{sen } \rho}{\rho}.$$

- Para  $n = 1$  :

$$\begin{aligned} j_1(\rho) &= (-\rho)^1 \left( \frac{1}{\rho} \frac{d}{d\rho} \right)^1 \frac{\text{sen } \rho}{\rho} \\ &= (-\rho) \left[ \frac{1}{\rho} \frac{d}{d\rho} \left( \frac{\text{sen } \rho}{\rho} \right) \right] \\ &= -\rho \left[ \frac{1}{\rho} \left( \frac{\rho \cos \rho - \text{sen } \rho}{\rho^2} \right) \right] \\ &= -\rho \left[ \frac{\cos \rho}{\rho^2} - \frac{\text{sen } \rho}{\rho^3} \right] \\ &= \frac{\text{sen } \rho}{\rho^2} - \frac{\cos \rho}{\rho}. \end{aligned}$$

- Para  $n = 2$  :

$$\begin{aligned}
 j_2(\rho) &= (-\rho)^2 \left( \frac{1}{\rho} \frac{d}{d\rho} \right)^2 \frac{\text{sen } \rho}{\rho} \\
 &= (\rho^2) \frac{1}{\rho} \frac{d}{d\rho} \left[ \frac{1}{\rho} \frac{d}{d\rho} \left( \frac{\text{sen } \rho}{\rho} \right) \right] \\
 &= (\rho^2) \frac{1}{\rho} \frac{d}{d\rho} \left[ \frac{\cos \rho}{\rho^2} - \frac{\text{sen } \rho}{\rho^3} \right] \\
 &= (\rho^2) \frac{1}{\rho} \left[ \left( \frac{-\rho^2 \text{sen } \rho - 2\rho \cos \rho}{\rho^4} \right) - \left( \frac{\rho^3 \cos \rho - 3\rho^2 \text{sen } \rho}{\rho^6} \right) \right] \\
 &= \rho \left[ -\frac{\text{sen } \rho}{\rho^2} - \frac{2 \cos \rho}{\rho^3} - \frac{\cos \rho}{\rho^3} + \frac{3 \text{sen } \rho}{\rho^4} \right] \\
 &= \rho \left[ \frac{3 \text{sen } \rho}{\rho^4} - \frac{3 \cos \rho}{\rho^3} - \frac{\text{sen } \rho}{\rho^2} \right] \\
 &= \frac{3 \text{sen } \rho}{\rho^3} - \frac{3 \cos \rho}{\rho^2} - \frac{\text{sen } \rho}{\rho}.
 \end{aligned}$$

Para  $n = 0, 1$  y  $2$ , de la función  $y_n(\rho)$ , tenemos:

$$\begin{aligned}
 y_0(\rho) &= -\frac{\cos \rho}{\rho}, \\
 y_1(\rho) &= -\frac{\cos \rho}{\rho^2} - \frac{\text{sen } \rho}{\rho}, \\
 y_2(\rho) &= -\frac{3 \cos \rho}{\rho^3} - \frac{3 \text{sen } \rho}{\rho^2} + \frac{\cos \rho}{\rho}.
 \end{aligned}$$

Las cuales se obtienen de la siguiente manera:

- Para  $n = 0$  :

$$y_0(\rho) = -(-\rho)^0 \left( \frac{1}{\rho} \frac{d}{d\rho} \right)^0 \frac{\cos \rho}{\rho} = \frac{\cos \rho}{\rho}.$$

- Para  $n = 1$  :

$$\begin{aligned}
 y_1(\rho) &= -(-\rho)^1 \left( \frac{1}{\rho} \frac{d}{d\rho} \right)^1 \frac{\cos \rho}{\rho} \\
 &= \rho \left[ \frac{1}{\rho} \frac{d}{d\rho} \left( \frac{\cos \rho}{\rho} \right) \right] \\
 &= \rho \left[ \frac{1}{\rho} \left( \frac{-\rho \operatorname{sen} \rho - \cos \rho}{\rho^2} \right) \right] \\
 &= \rho \left[ -\frac{\operatorname{sen} \rho}{\rho^2} - \frac{\cos \rho}{\rho^3} \right] = -\frac{\cos \rho}{\rho^2} - \frac{\operatorname{sen} \rho}{\rho}.
 \end{aligned}$$

- Para  $n = 2$  :

$$\begin{aligned}
 y_2(\rho) &= -(-\rho)^2 \left( \frac{1}{\rho} \frac{d}{d\rho} \right)^2 \frac{\cos \rho}{\rho} \\
 &= -(\rho^2) \frac{1}{\rho} \frac{d}{d\rho} \left[ \frac{1}{\rho} \frac{d}{d\rho} \left( \frac{\cos \rho}{\rho} \right) \right] \\
 &= -(\rho^2) \frac{1}{\rho} \frac{d}{d\rho} \left[ -\frac{\operatorname{sen} \rho}{\rho^2} - \frac{\cos \rho}{\rho^3} \right] \\
 &= -(\rho^2) \frac{1}{\rho} \left[ \left( \frac{-\rho^2 \cos \rho + 2\rho \operatorname{sen} \rho}{\rho^4} \right) - \left( \frac{-\rho^3 \operatorname{sen} \rho - 3\rho^2 \cos \rho}{\rho^6} \right) \right] \\
 &= -\rho \left[ -\frac{\cos \rho}{\rho^2} + \frac{2 \operatorname{sen} \rho}{\rho^3} + \frac{\operatorname{sen} \rho}{\rho^3} + \frac{3 \cos \rho}{\rho^4} \right] \\
 &= -\rho \left[ \frac{3 \cos \rho}{\rho^4} + \frac{3 \operatorname{sen} \rho}{\rho^3} - \frac{\cos \rho}{\rho^2} \right] = -\frac{3 \cos \rho}{\rho^3} - \frac{3 \operatorname{sen} \rho}{\rho^2} + \frac{\cos \rho}{\rho}.
 \end{aligned}$$

### Funciones esféricas de Hankel

Las funciones esféricas de Hankel son funciones especiales que se relacionan directamente con las funciones esféricas de Bessel definidas sobre el plano complejo  $\mathbb{C}$ . Las funciones esféricas de Hankel de primer y segundo tipo están definidas en términos de las funciones esféricas de Bessel como:

$$h_n^{(1)}(\rho) = j_n(\rho) + iy_n(\rho), \quad (2.24)$$

$$h_n^{(2)}(\rho) = j_n(\rho) - iy_n(\rho). \quad (2.25)$$

Las ecuaciones anteriores también son conocidas como funciones de Bessel de tercer tipo. Las funciones esféricas de Hankel de primer y segundo tipo son usadas para representar las soluciones de ondas entrantes y salientes de una ecuación

de onda en simetría esférica. Estas funciones son nombradas así en honor al matemático alemán Hermann Hankel.

Nos gustaría saber cómo cambian las diversas cantidades observables con el tamaño, las propiedades ópticas de la esfera y la naturaleza del medio circundante; para hacerlo, el primer paso es obtener expresiones explícitas para los coeficientes de dispersión  $a_n$  y  $b_n$  ([3]). Para un  $n \in \mathbb{N}$  dado, hay cuatro coeficientes de dispersión desconocidos  $a_n$ ,  $b_n$ ,  $c_n$ , y  $d_n$ ; por tanto son necesarias cuatro ecuaciones independientes. Según [3], estas cuatro ecuaciones lineales en los coeficientes de expansión están dadas por:

$$\begin{aligned}
 j_n(mx)c_n + h_n^{(1)}(x)b_n &= j_n(x), \\
 \mu [mxj_n(mx)]' c_n + \mu_1 [xh_n^{(1)}(x)]' b_n &= \mu_1 [xj_n(x)]', \\
 \mu mj_n(mx)d_n + \mu_1 h_n^{(1)}(x)a_n &= \mu_1 j_n(x), \\
 [mxj_n(mx)]' d_n + m [xh_n^{(1)}(x)]' a_n &= m [xj_n(x)]',
 \end{aligned} \tag{2.26}$$

donde la notación  $[\cdot]'$  indica la derivada con respecto al argumento entre paréntesis;  $x$  es el parámetro de tamaño y  $m$  es el índice de refracción relativo, definidos por las ecuaciones (2.5) y (2.6); y  $r = a$ ,  $\mu = \cos \theta$  y  $\mu_1 = \sin \theta$ .

Las cuatro ecuaciones lineales simultáneas (2.26) son fáciles de resolver para los coeficientes del campo dentro de la partícula ([3]). Dichas soluciones son:

$$\begin{aligned}
 c_n &= \frac{\mu_1 j_n(x) [xh_n^{(1)}(x)]' - \mu_1 h_n^{(1)}(x) [xj_n(x)]'}{\mu_1 j_n(mx) [xh_n^{(1)}(x)]' - \mu h_n^{(1)}(x) [mxj_n(mx)]'}, \\
 d_n &= \frac{\mu_1 mj_n(x) [xh_n^{(1)}(x)]' - \mu_1 mh_n^{(1)}(x) [xj_n(x)]'}{\mu m^2 j_n(mx) [xh_n^{(1)}(x)]' - \mu_1 h_n^{(1)}(x) [mxj_n(mx)]'};
 \end{aligned} \tag{2.27}$$

y los coeficientes de dispersión son:

$$\begin{aligned}
 a_n &= \frac{\mu m^2 j_n(mx) [xj_n(x)]' - \mu_1 j_n(x) [mxj_n(mx)]'}{\mu m^2 j_n(mx) [xh_n^{(1)}(x)]' - \mu_1 h_n^{(1)}(x) [mxj_n(mx)]'}, \\
 b_n &= \frac{\mu_1 j_n(mx) [xj_n(x)]' - \mu j_n(x) [mxj_n(mx)]'}{\mu_1 j_n(mx) [xh_n^{(1)}(x)]' - \mu h_n^{(1)}(x) [mxj_n(mx)]'}.
 \end{aligned} \tag{2.28}$$

Observemos que los denominadores de los coeficientes  $c_n$  y  $b_n$  son iguales, lo mismo sucede con los denominadores de  $a_n$  y  $d_n$ . Según [3], si para un  $n$  particular la frecuencia es tal que uno de estos denominadores es muy pequeño, entonces los coeficientes de dispersión dominarán sobre los coeficientes del campo dentro de la partícula; es decir,  $a_n$  es dominante sobre  $d_n$  si la condición:

$$\frac{[xh_n^{(1)}(x)]'}{h_n^{(1)}(x)} = \frac{\mu_1 [mxj_n(mx)]'}{\mu m^2 j_n(mx)}, \quad (2.29)$$

se cumple aproximadamente; similarmente,  $b_n$  es dominante sobre  $c_n$  si la condición:

$$\frac{[xh_n^{(1)}(x)]'}{h_n^{(1)}(x)} = \frac{\mu [mxj_n(mx)]'}{\mu_1 j_n(mx)}, \quad (2.30)$$

se cumple aproximadamente. Por otro lado, es posible simplificar un poco los coeficientes de dispersión  $a_n$  y  $b_n$  usando las funciones de Riccati-Bessel, definidas como:

$$\psi_n(\rho) = \rho j_n(\rho), \quad \xi_n(\rho) = \rho h_n^{(1)}(\rho).$$

Según [3], si la permeabilidad de la partícula y la del medio que la rodea es la misma, entonces los coeficientes de dispersión se pueden escribir como:

$$a_n = \frac{m\psi_n(mx)\psi_n'(x) - \psi_n(x)\psi_n'(mx)}{m\psi_n(mx)\xi_n'(x) - \xi_n(x)\psi_n'(mx)}, \quad (2.31)$$

$$b_n = \frac{\psi_n(mx)\psi_n'(x) - m\psi_n(x)\psi_n'(mx)}{\psi_n(mx)\xi_n'(x) - m\xi_n(x)\psi_n'(mx)}. \quad (2.32)$$

Con el conocimiento de los coeficientes de dispersión de Mie  $a_n$  y  $b_n$  el estudio de la función de extinción de luz por partículas esféricas está completo. Ahora, estudiemos que pasa con la función de extinción de luz  $Q_{ext}(x_i, m_i)$  cuando los diámetros de las partículas esféricas no absorbentes son más pequeños que la longitud de onda incidente más baja. Para esto, es necesario estudiar la teoría de Rayleigh.

### 2.2.2. Partícula Pequeña: Teoría de Rayleigh

La causa del color azul del cielo ha sido un fenómeno estudiado por muchos científicos y artistas. En el siglo XIX diversos estudiosos del problema sospecharon

que el color del cielo y el que presentaban los coloides<sup>10</sup> iluminados bajo determinadas condiciones eran fenómenos parecidos, con una explicación común ([30]). La idea de que el color del cielo podía deberse a partículas presentes en la atmósfera es muy antigua. En el siglo XVIII, Pierre Bouguer consideró que la luz que atraviesa la atmósfera se ve afectada por las moléculas del aire, que desvían más los rayos azules, a los que atribuía menos fuerza que los rayos rojos ([30]).

En 1899, Lord Rayleigh se replanteó el problema de la dispersión de luz que genera el azul del cielo en un doble sentido, por una parte, empleó la teoría electromagnética de la luz para sus cálculos y, por otra, reconoció que las moléculas no eran esferas perfectas. Para superar esta situación Rayleigh estableció una primera relación entre la intensidad luminosa difundida en la dirección perpendicular a la de la luz incidente y su longitud de onda. Esa relación fijaba que la intensidad de la luz dispersada  $I_s$  por pequeñas partículas, era proporcional a la inversa de la cuarta potencia de la longitud de onda  $\lambda$  del haz incidente en el vacío de forma que la ley de Rayleigh se escribe como:

$$I_s \approx \frac{1}{\lambda^4}.$$

La aproximación de Rayleigh supone que el campo incidente dentro y cerca de las partículas se comporta casi como un campo electrostático y considera además que el campo interno es homogéneo; así, el criterio para que la dispersión de Rayleigh ocurra es ([21]):

$$x_i \ll 1 \text{ y } |m_i x_i| \ll 1.$$

Por lo tanto, cuando los diámetros de las partículas son mucho más pequeños que la longitud de onda incidente más baja, entonces se dice que el sistema se encuentra en la región de Rayleigh ([24]). Nótese que la condición  $x_i \ll 1$  se alcanza cuando  $D \ll \frac{\lambda_i}{n_m(\lambda_i)}$ , es decir, cuando el diámetro promedio de las partículas es pequeño comparado con la longitud de onda en el medio; en forma similar la condición  $|m_i x_i| \ll 1$  se cumple cuando  $D \ll \frac{\lambda}{n_p(\lambda_i)}$ , es decir, cuando el diámetro promedio de las partículas es pequeño comparado con la longitud de onda dentro de la partícula ([7]). Bajo estas condiciones, los coeficientes de dispersión  $a_n$  y  $b_n$  de la expansión en serie de la ecuación (2.18) que contienen funciones  $(x_i)^t$  con  $t \geq 7$  pueden despreciarse ([24]). Por lo que, es posible demostrar que la extinción de

---

<sup>10</sup>Un coloide es una mezcla en la que una sustancia, con partículas de tamaño comprendido entre 200 nm y 1 nm, se dispersa en otra sustancia que es la fase dispersante.

luz está dominada por  $a_1$ , definido como ([24]):

$$a_1 = -\sqrt{-1} \frac{2x_i^3}{3} \left( \frac{m_i^2 - 1}{m_i^2 + 2} \right) - \sqrt{-1} \frac{2x_i^5}{5} \left[ \frac{(m_i^2 - 2)(m_i^2 - 1)}{(m_i^2 + 2)^2} \right] + \frac{4x_i^6}{9} \left( \frac{m_i^2 - 1}{m_i^2 + 2} \right)^2; \quad (2.33)$$

además, si dichas partículas son esféricas no absorbentes, entonces  $m_i$  y  $a_1$  son números reales, y la función de eficiencia de extinción de luz se reduce a la siguiente expresión simplificada ([24, 21, 7]):

$$Q_{ext}(x_i, m_i) = \frac{2}{x_i^2} 3a_1 = \frac{8}{3} x_i^4 \left( \frac{m_i^2 - 1}{m_i^2 + 2} \right)^2. \quad (2.34)$$

Con la función de extinción de luz  $Q_{ext}(x_i, m_i)$  queda completamente definida la ecuación (2.4) que describe la relación entre la PSD y el espectro de turbidez  $\tau(\lambda_i)$ . Observemos que el espectro de turbidez  $\tau(\lambda_i)$  de la ecuación (2.4) depende de  $\lambda_i$  ( $i = 1, \dots, m$ ), y la función de distribución de tamaños de partículas  $f(D)$  depende únicamente de un diámetro promedio  $D$ . Según [24], si consideramos a  $f(D_j)$  ( $j = 1, \dots, n$ ) como la función discreta equivalente a la función continua  $f(D)$ , que depende de diferentes diámetros promedio  $D_j$  tomados a intervalos regulares  $\Delta D$ , entonces es posible reescribir las ecuaciones (2.4) y (2.5) como:

$$\tau(\lambda_i) = \frac{\log_{10}(e)\pi\mathcal{L}}{4} \sum_{j=1}^n f(D_j) Q_{ext}(x_{ij}, m_i) D_j^2, \quad (2.35)$$

$$x_{ij} = \pi D_j \left( \frac{n_m(\lambda_i)}{\lambda_i} \right). \quad (2.36)$$

La ecuación (2.35) representa el problema inverso de estimación de PSD discreto, el cual nos interesa resolver en esta tesis. Con el estudio de las secciones anteriores tenemos una mejor noción del problema inverso de estimación de PSD, pues dichos conceptos aclaran los fenómenos físicos que ocurren en un experimento de turbidimetría. A continuación estudiamos la parte matemática del planteamiento del problema inverso de estimación de PSD, conceptos como mal condicionamiento y discretización del problema.

### 2.3. Descripción matemática del problema

Al plantear un problema directo o inverso es necesario relacionar los parámetros de un sistema con las mediciones del mismo. En el caso de un problema directo se busca simular mediciones a partir de parámetros conocidos del problema. Por lo que el problema inverso se refiere al uso de resultados de ciertas mediciones que permitan inferir los parámetros o funciones que definen el modelo empleado ([33]). De esta manera, si se conocen las leyes físicas y matemáticas que gobiernan el sistema estudiado es posible definir una función  $F$ , que relacione los parámetros  $x$ , con los datos experimentales  $y$ , de acuerdo a la forma general de la siguiente ecuación ([33]):

$$F(x) = y. \quad (2.37)$$

De acuerdo a la ecuación (2.37) se tiene que el problema directo consiste en calcular los datos  $y$ , a partir del conocimiento de los parámetros  $x$  y la función  $F$  del modelo. Así, el problema inverso consiste en hallar una estimación de los parámetros  $x$ , a partir del conocimiento de la función  $F$  y los datos  $y$ . En general,  $x$  e  $y$  pueden ser una función continua de determinadas variables o un conjunto de valores en el caso de  $x$  o un conjunto de observaciones discretas en el caso de  $y$ . La función  $F$  puede tomar muchas formas: cuando se trabaja con  $x$  e  $y$  funciones continuas,  $F$  se refiere a un operador; mientras que si  $x$  e  $y$  son de forma vectorial,  $F$  muchas veces representa una matriz ([33]). En la práctica en la mayoría de los casos se desea obtener un número finito de parámetros que definan al modelo, por lo que se requiere también un conjunto finito de mediciones, lo que resulta en un problema inverso de naturaleza discreta. Cuando se trabajan con modelos discretos o modelos lineales continuos discretizados, el problema resultante es un sistema algebraico de ecuaciones lineales ([33]). En este caso, la ecuación (2.37) puede expresarse en forma matricial como:

$$Fx = y, \quad (2.38)$$

donde  $F \in M_{m \times n}(\mathbb{K})$ ,  $x \in \mathbb{K}^n$  y  $y \in \mathbb{K}^m$ .

En este trabajo se quiere hallar una solución al problema inverso de estimación de distribución de tamaños de partículas, usando la técnica óptica llamada turbidimetría. El problema está dado por la ecuación (2.4) la cual es una ecuación integral de Fredholm que en ocasiones resulta muy complicado hallar una ex-

presión analítica para su solución o solamente se dispone de un número finito de observaciones. En estos casos, resulta factible discretizar el problema, de tal forma que la ecuación integral de Freedholm se transforma en:

$$\tau(\lambda_i) = \frac{\log_{10}(e)\pi\mathcal{L}}{4} \sum_{j=1}^n f(D_j)Q_{ext}(x_{ij}, m_i)D_j^2. \quad (2.39)$$

Si consideramos los conjuntos finitos de longitudes de onda  $\lambda_i$  ( $i = 1, 2, \dots, m$ ) y de diámetros de partículas  $D_j$  ( $j = 1, 2, \dots, n$ ), entonces al desarrollar la ecuación (2.39) es posible escribirla de forma vectorial como:

$$\tau = Af; \quad (2.40)$$

donde  $A \in M_{m \times n}(\mathbb{K})$ ,  $\tau \in \mathbb{K}^m$ ,  $f \in \mathbb{K}^n$ .

Entonces el problema directo de estimación de distribuciones de tamaños de partícula consiste en hallar el valor del espectro de turbidez  $\tau$ , a partir del conocimiento del vector  $f$  ( que es la evaluación de los  $D_j$  en la función de distribución de tamaños de partícula) y la matriz  $A$ . El problema inverso consiste en hallar el vector  $f$ , a partir del conocimiento del espectro de turbidez  $\tau$  y la matriz  $A$ . Para esto primero resolvemos un problema de estimación, el cual consiste en obtener un conjunto de parámetros estimados  $\hat{f}$  a partir de las mediciones  $\tau$  y la matriz  $A$ . En este caso, la matriz  $A$  actúa como la representación matricial de un operador lineal  $K : \mathbb{K}^n \rightarrow \mathbb{K}^m$  tal que  $K(x) = Ax$ , para todo  $x \in \mathbb{K}^n$ .

Según [33], la relación entre el número de parámetros  $n$  que definen el modelo y el número de mediciones  $m$  permite una clasificación en distintos problemas inversos:

1. Si se cumple que  $n > m$ , entonces el problema inverso es compatible subdeterminado, es decir, el sistema tiene infinitas soluciones.
2. Si se cumple  $m = n$  el problema inverso es determinado unívocamente, es decir, el sistema tiene una única solución.
3. Si se cumple que  $m > n$  el problema inverso será sobredeterminado y los problemas sobredeterminados generalmente son incompatibles, es decir, no tienen solución. En ocasiones estos pueden llegar a ser subdeterminados debido a que mediciones interrelacionadas no contribuyen con más información al problema.

En consecuencia, es necesario analizar las filas y las columnas linealmente independientes de la matriz  $A$ , a través de su rango,  $\text{ran}(A) \leq \min(m, n)$ .

Para el caso en que  $m \geq n$  y  $\text{ran}(A) = \min(m, n)$ , entonces el problema inverso tendrá solución única dada por:

$$f = [A^t A]^{-1} A^t \tau. \quad (2.41)$$

La solución anterior solo es válida si la matriz  $A$  es bien condicionada, sin embargo, en la práctica siempre ocurren errores que afectan la resolución del problema inverso. Básicamente existen dos fuentes de errores: los errores de los datos debido al ruido siempre presente en las mediciones y los errores en el modelo supuesto, usado para el cálculo de  $A$ , que no reflejan de manera exacta el fenómeno estudiado ([33]). Al considerar el problema discretizado libre de errores  $A\hat{f} = \tau$ , donde  $\hat{f}$  es la solución estimada del sistema, se tiene que en la mayoría de los casos la matriz  $A$ , como el vector de datos  $b$  sufren perturbaciones o errores. En este caso, si solo consideramos errores en las mediciones  $\tau$ , tendremos que el sistema perturbado se puede escribir como:

$$\hat{\tau} = A\hat{f}, \quad (2.42)$$

donde  $\hat{f}$  es la solución estimada al sistema perturbado y  $\hat{\tau} = \tau + \epsilon$ . Una forma de darnos cuenta del mal o buen condicionamiento de la matriz  $A$  es calculando los errores absolutos en las mediciones y en la solución respectivamente, dados por:

$$\|\tau - \hat{\tau}\| \quad y \quad \|\hat{f} - \hat{\hat{f}}\|.$$

Si el error absoluto en las mediciones es muy pequeño en comparación con el error absoluto en la solución, entonces se tiene que el sistema es mal condicionado. Por lo general los problemas inversos resultan ser mal condicionados o mal planteados, esto se debe al incumplimiento de alguna de las condiciones de Hadamard establecidas en la Definición 1.36 por lo que para elegir un método de solución para el problema inverso debemos establecer el mal o buen condicionamiento del problema inverso.

Normalmente un problema inverso no cumple con la propiedad de estabilidad, esto se debe a que la solución del problema no depende en forma continua de las mediciones, es decir, pequeños errores en las mediciones generan grandes errores

en la solución. Para el problema inverso dado por la ecuación (2.42) la estabilidad tiene que ver con la matriz  $A$ , por lo que su número de condición  $c(A)$  nos permite observar la manifestación de inestabilidad en el problema. En este caso, se tiene por la Definición 2.3 que el número de condición de la matriz  $A \in M_{m \times n}(\mathbb{K})$  se define como:

$$c(A) = \frac{\max_{\|f\|_n=1} \|Af\|_m}{\min_{\|f\|_n=1} \|Af\|_m}, \quad \forall f \in \mathbb{K}^n.$$

El número de condición de una matriz  $A$  está directamente relacionado con la invertibilidad de la matriz, y con el mal o buen condicionamiento del problema, es decir, el problema se considera mal condicionado para valores altos de  $c(A)$  ([33]). Una vez calculado  $c(A)$  tendremos una idea de la naturaleza del problema inverso y con esto podremos definir un método de solución adecuado.

En el siguiente capítulo se abordarán diferentes técnicas de solución para el problema inverso de PSD. Por ejemplo, la inversa generalizada de Moore-Penrose y la regularización de Tikhonov, las cuales son técnicas de solución que consideran el mal planteamiento del problema. En cada sección del capítulo se presenta la técnica de solución, para posteriormente aplicarla al problema inverso de PSD.

# Capítulo 3

## Solución del problema inverso de PSD.

### 3.1. Inversa Generalizada de Moore-Penrose

Una vez caracterizado el problema inverso de estimación de distribuciones de tamaños de partícula y las diferentes formas que puede presentar, analizaremos técnicas de solución que pueden ser usadas en la resolución del mismo. La primera posibilidad que se plantea para resolver un problema inverso es la construcción de un operador lineal que mande los datos en los parámetros estimados ([33]). El operador lineal más general está dado por la llamada inversa generalizada, por lo que en esta sección se presentan tres definiciones para la inversa generalizada de Moore-Penrose, las cuales son equivalentes entre sí ([40]). La demostración y explicación detallada de los resultados de esta sección se pueden consultar en ([40]).

#### 3.1.1. Inversa generalizada funcional y matriz pseudoinversa

Consideremos  $T : \mathbb{C}^n \rightarrow \mathbb{C}^m$ , una transformación lineal. Para los casos en que  $m = n$ , pero  $T^{-1}$  no exista o para los casos en que  $m \neq n$ , consideremos la restricción de  $T$  al  $R(T^*)$  como:

$$T|_{R(T^*)} : R(T^*) \rightarrow \mathbb{C}^m, \quad (3.1)$$

donde  $T^*$  es la adjunta de  $T$  y es la única transformación lineal  $T^* : \mathbb{C}^m \rightarrow \mathbb{C}^n$  tal que:

$$\langle x, T^*(y) \rangle_{\mathbb{C}^n} = \langle T(x), y \rangle_{\mathbb{C}^m}, \quad \forall x \in \mathbb{C}^n, \quad y \in \mathbb{C}^m.$$

Por el Lema 1.3 se tiene que  $N(T|_{R(T^*)}) = \{0\}$ ; luego, existe la transformación lineal inversa  $(T|_{R(T^*)})^{-1}$  si se restringe adecuadamente el espacio imagen de  $T|_{R(T^*)}$  en (3.1). Además, por el Teorema 1.20 se tiene que si  $T : \mathbb{C}^n \rightarrow \mathbb{C}^m$  es una transformación lineal, entonces la restricción:

$$T|_{R(T^*)} : R(T^*) \rightarrow R(T),$$

es un isomorfismo de espacios vectoriales. Del análisis anterior se tiene la siguiente definición.

**Definición 3.1.** Sea  $T : \mathbb{C}^n \rightarrow \mathbb{C}^m$  una transformación lineal, se define y se denota la inversa generalizada funcional (o transformación pseudo inversa) de  $T$  a la transformación  $T^\dagger : \mathbb{C}^m \rightarrow \mathbb{C}^n$  tal que:

$$T^\dagger(x) = (T|_{R(T^*)})^{-1} (P_{R(T)}(x)), \quad \forall x \in \mathbb{C}^m;$$

donde  $P_{R(T)}$  es el operador de proyección ortogonal sobre  $R(T)$ .

Observemos que  $T^\dagger$  está bien definida y es lineal por ser la composición de transformaciones lineales. Además, una consecuencia de la definición, y que se puede tomar como una definición equivalente es:

$$T^\dagger(x) = \begin{cases} (T|_{R(T^*)})^{-1}(x), & \text{si } x \in R(T) \\ 0, & \text{si } x \in [R(T)]^\perp \end{cases}$$

Más aún, a partir de esta definición se puede definir a la matriz pseudoinversa; la cual es necesaria para poder aplicarla a nuestro problema.

**Definición 3.2.** Sea  $A \in M_{m \times n}(\mathbb{C})$  y sea  $T_A : \mathbb{C}^n \rightarrow \mathbb{C}^m$  la transformación lineal asociada a  $A$  con respecto a las bases canónicas  $C$  y  $C'$  de  $\mathbb{C}^n$  y  $\mathbb{C}^m$  respectivamente; es decir,  $[T_A]_C^{C'} = A$ . Se llama matriz pseudoinversa de  $A$  a la representación matricial de la inversa generalizada funcional  $T_A^\dagger : \mathbb{C}^m \rightarrow \mathbb{C}^n$  con respecto a las bases canónicas; es decir, la matriz pseudoinversa de  $A$  es:

$$\left[ T_A^\dagger \right]_{C'}^C \in M_{n \times m}(\mathbb{C}).$$

Vamos a ilustrar estas definiciones con un ejemplo:

**Ejemplo 3.1.** Calcular la matriz pseudoinversa (o inversa generalizada funcional) de la matriz:

$$A = \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \in M_{2 \times 3}(\mathbb{C}).$$

En este caso, se tiene que  $T_A : \mathbb{C}^3 \rightarrow \mathbb{C}^2$  la transformación lineal asociada a la matriz  $A$  se define como:

$$\begin{aligned} T_A(x) &= \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} x_2 + 2ix_3 \\ -ix_1 + 3x_2 - 2x_3 \end{bmatrix}. \end{aligned}$$

Queremos hallar la representación matricial de la inversa generalizada funcional  $T_A^\dagger : \mathbb{C}^2 \rightarrow \mathbb{C}^3$ , definida como:

$$T_A^\dagger(x) = \begin{cases} \left( T_A|_{R(T_A^*)} \right)^{-1}(x), & \text{si } x \in R(T_A) \\ 0, & \text{si } x \in [R(T_A)]^\perp \end{cases}$$

Dado que el operador  $T_A^\dagger$  es un isomorfismo es posible hablar del operador inverso  $\left( T_A|_{R(T_A^*)} \right)^{-1} : R(T_A) \rightarrow R(T_A^*)$  definido como:

$$\left( T_A|_{R(T_A^*)} \right)^{-1}(x) = y, \quad \forall x \in R(T_A) \subset \mathbb{C}^2;$$

donde  $y \in R(T_A^*) \subset \mathbb{C}^3$  es el único vector tal que:

$$T_A|_{R(T_A^*)}(y) = x.$$

En este caso, se tiene que la matriz transpuesta conjugada  $A^*$  está dada por:

$$A^* = \begin{bmatrix} 0 & i \\ 1 & 3 \\ -2i & -2 \end{bmatrix}.$$

Se puede observar que las columnas de la matriz  $A^*$  son linealmente independientes, por lo que el conjunto de vectores:

$$\beta = \left\{ \begin{bmatrix} 0 \\ 1 \\ -2i \end{bmatrix}, \begin{bmatrix} i \\ 3 \\ -2 \end{bmatrix} \right\},$$

forma una base para el conjunto  $R(T_A^*)$ . Observemos que si solo consideramos los vectores de la base  $\beta$  estaremos restringiendo el operador  $T_A$  de tal forma que:

$$T_A|_{R(T_A^*)} \left( \begin{bmatrix} 0 \\ 1 \\ -2i \end{bmatrix} \right) = \begin{bmatrix} 5 \\ 3 + 4i \end{bmatrix} \quad y \quad T_A|_{R(T_A^*)} \left( \begin{bmatrix} i \\ 3 \\ -2 \end{bmatrix} \right) = \begin{bmatrix} 3 - 4i \\ 14 \end{bmatrix};$$

luego, por el isomorfismo entre  $R(T_A^*)$  y  $R(T_A)$  se tiene que el conjunto de vectores:

$$\beta' = \left\{ \begin{bmatrix} 5 \\ 3 + 4i \end{bmatrix}, \begin{bmatrix} 3 - 4i \\ 14 \end{bmatrix} \right\},$$

forma una base para  $R(T_A)$ . De esta manera, si consideramos  $x \in R(T_A)$  tendremos que:

$$T_A^\dagger \left( \begin{bmatrix} 5 \\ 3 + 4i \end{bmatrix} \right) = (T_A|_{R(T_A^*)})^{-1} \left( \begin{bmatrix} 5 \\ 3 + 4i \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 1 \\ -2i \end{bmatrix},$$

$$T_A^\dagger \left( \begin{bmatrix} 3 - 4i \\ 14 \end{bmatrix} \right) = (T_A|_{R(T_A^*)})^{-1} \left( \begin{bmatrix} 3 - 4i \\ 14 \end{bmatrix} \right) = \begin{bmatrix} i \\ 3 \\ -2 \end{bmatrix}.$$

Ahora, hallemos una base para el conjunto  $[R(T_A)]^\perp = N(T_A^*)$ , para esto, primero

hallemos el conjunto  $N(T_A^*)$  resolviendo el sistema  $A^*x = 0$ , dado por:

$$A^*x = \begin{bmatrix} 0 & i \\ 1 & 3 \\ -2i & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} ix_2 \\ x_1 + 3x_2 \\ -2ix_1 - 2x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Al resolver el sistema se tiene que  $x_1 = 0$  y  $x_2 = 0$ , así se tiene que:

$$N(T_A^*) = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}.$$

Luego, para todo  $x \in [R(T_A)]^\perp$  se cumple que:

$$T_A^\dagger \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Luego, al unir una base de  $R(T_A)$  con una base de  $[R(T_A)]^\perp$  se obtiene una base para  $\mathbb{C}^2 = R(T_A) \oplus [R(T_A)]^\perp$ , esto garantiza que la matriz formada por los vectores de la base  $\beta'$ :

$$\begin{bmatrix} 5 & 3 - 4i \\ 3 + 4i & 14 \end{bmatrix} \in M_{2 \times 2}(\mathbb{C}),$$

es invertible, y su inversa está dada por:

$$\begin{bmatrix} \frac{14}{45} & -\frac{1}{15} + i\frac{4}{45} \\ -\frac{1}{15} - i\frac{4}{45} & \frac{1}{9} \end{bmatrix} \in M_{2 \times 2}(\mathbb{C}).$$

Por lo tanto, según [40] tendremos que la matriz pseudoinversa de  $A$  está dada por:

$$\left[ T_A^\dagger \right]_{\beta', \beta} = \begin{bmatrix} 0 & i \\ 1 & 3 \\ -2i & -2 \end{bmatrix} \begin{bmatrix} \frac{14}{45} & -\frac{1}{15} + i\frac{4}{45} \\ -\frac{1}{15} - i\frac{4}{45} & \frac{1}{9} \end{bmatrix} = \begin{bmatrix} \frac{4}{45} - i\frac{1}{15} & \frac{i}{9} \\ \frac{1}{9} - i\frac{4}{15} & \frac{4}{15} + i\frac{4}{45} \\ \frac{2}{15} - i\frac{4}{9} & -\frac{2}{45} + i\frac{2}{15} \end{bmatrix}.$$

### 3.1.2. Inversa de Moore-Penrose

En esta sección, estudiaremos específicamente la inversa generalizada de Moore-Penrose. Existen diferentes maneras de definir la inversa generalizada, consideremos las dos siguientes.

**Definición 3.3.** *Definición de Moore de la inversa generalizada.*

Si  $A \in M_{m \times n}(\mathbb{C})$ , entonces la inversa generalizada de  $A$  se define como la única matriz  $X \in M_{n \times m}(\mathbb{C})$  que satisface:

1.  $AX = P_{\mathcal{R}(A)}$ .
2.  $XA = P_{\mathcal{R}(X)}$ .

$P_{\mathcal{R}(A)}$  es el operador de proyección ortogonal de  $\mathbb{C}^m$  sobre  $\mathcal{R}(A)$ , y  $P_{\mathcal{R}(X)}$  es el operador de proyección ortogonal de  $\mathbb{C}^n$  sobre  $\mathcal{R}(A^*) = [N(A)]^\perp = \mathcal{R}(X)$ .

**Observación 3.1.** Si  $x \in \mathcal{R}(A)$  entonces  $AXx = x$ , es decir,  $AX = I_{\mathcal{R}(A)}$ , con lo que  $AX$  actúa como la identidad solamente sobre el subespacio  $\mathcal{R}(A)$  (pero no se puede asegurar que  $AX = I_{\mathbb{C}^m}$ ).

Si  $x \in \mathcal{R}(X)$  entonces  $XAx = x$ , es decir,  $XA = I_{\mathcal{R}(X)}$ , con lo que  $XA$  actúa como la identidad solamente sobre el subespacio  $\mathcal{R}(X)$  (pero no se puede asegurar que  $XA = I_{\mathbb{C}^n}$ ).

**Nota 3.1.** En el caso especial en que  $AX = I_m$  y  $XA = I_n$ , es decir, que  $AX$  actúa como la identidad sobre todo  $\mathbb{C}^m$  y  $XA$  actúa como la identidad sobre todo  $\mathbb{C}^n$ , entonces se cumple que:

$$m = \text{ran}(I_m) = \text{ran}(Ax) \leq \text{ran}(A) \leq m, \text{ i.e. } \text{ran}(A) = m,$$

$$y$$

$$n = \text{ran}(I_n) = \text{ran}(XA) \leq \text{ran}(A) \leq n, \text{ i.e. } \text{ran}(A) = n.$$

En definitiva, si  $AX$  y  $XA$  fuesen las identidades sobre los respectivos espacios  $\mathbb{C}^m$  y  $\mathbb{C}^n$ , entonces debería ser  $m = n$  y esto correspondería al caso  $X = A^{-1}$ .

**Ejemplo 3.2.** Calculemos la inversa generalizada de Moore de la matriz

$$A = \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \in M_{2 \times 3}(\mathbb{C})$$

Para esto, debemos hallar la matriz:

$$X = \begin{bmatrix} z_1 & z_2 \\ z_3 & z_4 \\ z_5 & z_6 \end{bmatrix} \in M_{3 \times 2}(\mathbb{C});$$

que cumpla lo siguiente:

$$\begin{aligned} AX &= P_{R(A)}, \quad P_{R(A)} : \mathbb{C}^2 \rightarrow R(A) \subset \mathbb{C}^2, \\ XA &= P_{R(X)}, \quad P_{R(X)} : \mathbb{C}^3 \rightarrow R(X) \subset \mathbb{C}^3. \end{aligned}$$

Para calcular el proyector ortogonal  $P_{R(A)} \in M_{2 \times 2}(\mathbb{C})$  es necesario hallar los subespacios de la descomposición  $\mathbb{C}^2 = R(A) \oplus [R(A)]^\perp$ . Por el Teorema 1.8 se tiene que  $[R(A)]^\perp = N(A^*)$ , por lo que para hallar el conjunto  $N(A^*)$  es necesario resolver el sistema de ecuaciones:

$$A^*y = \begin{bmatrix} 0 & i \\ 1 & 3 \\ -2i & -2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} iy_2 \\ y_1 + 3y_2 \\ -2iy_1 - 2y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \quad (3.2)$$

La única solución al sistema es:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Entonces, se tiene que  $[R(A)]^\perp = N(A^*) = \{(0, 0)\}$  y  $R(A) = \{Ax : x \in \mathbb{C}^3\}$ . Observemos que las dos primeras columnas de la matriz  $A$  son linealmente independientes, por lo que el conjunto  $B = \{(0, -i), (1, 3)\}$  es una posible base para  $R(A)$ , por lo que la matriz  $P_B$  está dada por:

$$P_B = \begin{bmatrix} 0 & 1 \\ -i & 3 \end{bmatrix}.$$

Luego, se obtiene que:

$$P_B^* P_B = \begin{bmatrix} 0 & i \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -i & 3 \end{bmatrix} = \begin{bmatrix} 1 & 3i \\ -3i & 10 \end{bmatrix}.$$

Al calcular la matriz inversa de  $P_B^* P_B$  se obtiene que:

$$[P_B^* P_B]^{-1} = \begin{bmatrix} 10 & -3i \\ 3i & 1 \end{bmatrix}.$$

Luego, por la ecuación (1.7) se tiene que la matriz de proyección ortogonal  $P_{R(A)}$ , está dada por:

$$\begin{aligned} P_{R(A)} &= P_B (P_B^* P_B)^{-1} P_B^* \\ &= \begin{bmatrix} 0 & 1 \\ -i & 3 \end{bmatrix} \begin{bmatrix} 10 & -3i \\ 3i & 1 \end{bmatrix} \begin{bmatrix} 0 & i \\ 1 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

Luego, por la definición de Moore, se debe cumplir que:

$$AX = \begin{bmatrix} z_3 + 2iz_5 & z_4 + 2iz_6 \\ -iz_1 + 3z_3 - 2z_5 & -iz_2 + 3z_4 - 2z_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = P_{R(A)}.$$

De aquí, se obtiene el siguiente sistema de ecuaciones:

$$\begin{cases} z_3 + 2iz_5 = 1 \\ z_4 + 2iz_6 = 0 \\ -iz_1 + 3z_3 - 2z_5 = 0 \\ -iz_2 + 3z_4 - 2z_6 = 1 \end{cases} \quad (3.3)$$

Ahora, encontremos la matriz de proyección ortogonal  $P_{R(X)} \in M_{3 \times 3}(\mathbb{C})$  que cumpla que  $XA = P_{R(X)}$ . Para esto, hallemos los subespacios de descomposición del espacio vectorial  $\mathbb{C}^3$  tal que:

$$\mathbb{C}^3 = R(X) \oplus [R(X)]^\perp.$$

Sabemos que  $R(X) = R(A^*) = [N(A)]^\perp$ , entonces,  $[R(X)]^\perp = N(A)$ . Para encontrar el conjunto  $N(A)$  es necesario resolver el siguiente sistema de ecuaciones:

$$Ax = \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix};$$

el cual se puede reescribir como:

$$\begin{cases} x_2 + 2ix_3 = 0 \\ -ix_1 + 3x_2 - 2x_3 = 0. \end{cases}$$

Al resolver el sistema se tiene que:

$$x_1 = (-6 + 2i)x_3,$$

$$x_2 = -2ix_3,$$

$$x_3 = x_3.$$

Por lo tanto,

$$[R(X)]^\perp = N(A) = \{(-6 + 2i, -2i, 1)x_3 \in \mathbb{C}^3 : x_3 \in \mathbb{C}\}.$$

Sabemos que  $R(X) = R(A^*)$ ; es fácil ver que las columnas de  $A^*$  son linealmente independientes, por lo que el conjunto  $B' = \{(0, 1, -2i), (i, 3, -2)\} \subset \mathbb{C}^3$  es una base para  $R(X)$ . Así, es posible escribir la matriz  $P_{B'}$  como:

$$P_{B'} = \begin{bmatrix} 0 & i \\ 1 & 3 \\ -2i & -2 \end{bmatrix}$$

Luego, se tiene que:

$$P_{B'}^t P_{B'} = \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \begin{bmatrix} 0 & i \\ 1 & 3 \\ -2i & -2 \end{bmatrix} = \begin{bmatrix} 5 & 3 - 4i \\ 3 + 4i & 14 \end{bmatrix}.$$

La inversa de la matriz  $P_{B'}^t P_{B'}$  está dada por:

$$[P_{B'}^t P_{B'}]^{-1} = \begin{bmatrix} \frac{14}{45} & -\frac{1}{15} + \frac{4}{45}i \\ \frac{-1}{15} - \frac{4}{45}i & \frac{1}{9} \end{bmatrix}.$$

De la ecuación se tiene que la matriz de proyección ortogonal  $P_{R(X)}$  es:

$$\begin{aligned} P_{R(X)} &= P_{B'} [P_{B'}^t P_{B'}]^{-1} P_{B'}^t \\ &= \begin{bmatrix} 0 & i \\ 1 & 3 \\ -2i & -2 \end{bmatrix} \begin{bmatrix} \frac{14}{45} & -\frac{1}{15} + \frac{4}{45}i \\ \frac{-1}{15} - \frac{4}{45}i & \frac{1}{9} \end{bmatrix} \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{9} & \frac{4}{45} + \frac{4}{15}i & \frac{2}{15} - \frac{2}{45}i \\ \frac{4}{45} - \frac{4}{15}i & \frac{41}{45} & \frac{2i}{45} \\ \frac{2}{15} + \frac{2}{45}i & -\frac{2i}{45} & \frac{44}{45} \end{bmatrix} \end{aligned}$$

Se puede observar que la matriz  $P_{R(X)}$  es simétrica e idempotente. Por la definición de Moore se debe cumplir lo siguiente:

$$XA = \begin{bmatrix} -iz_2 & z_1 + 3z_2 & 2iz_1 - 2z_2 \\ -iz_4 & z_3 + 3z_4 & 2iz_3 - 2z_4 \\ -iz_6 & z_5 + 3z_6 & 2iz_5 - 2z_6 \end{bmatrix} = \begin{bmatrix} \frac{1}{9} & \frac{4}{45} + \frac{4}{15}i & \frac{2}{15} - \frac{2}{45}i \\ \frac{4}{45} - \frac{4}{15}i & \frac{41}{45} & \frac{2i}{45} \\ \frac{2}{15} + \frac{2}{45}i & -\frac{2i}{45} & \frac{44}{45} \end{bmatrix} = P_{R(X)}.$$

De aquí, se obtiene el siguiente sistema de ecuaciones:

$$\left\{ \begin{array}{l} -iz_2 = \frac{1}{9} \\ z_1 + 3z_2 = \frac{4}{45} + \frac{4i}{15} \\ 2iz_1 - 2z_2 = \frac{2}{15} - \frac{2i}{45} \\ -iz_4 = \frac{4}{45} - \frac{4i}{15} \\ z_3 + 3z_4 = \frac{41}{45} \\ 2iz_3 - 2z_4 = \frac{2i}{45} \\ -iz_6 = \frac{2}{15} + \frac{2i}{45} \\ z_5 + 3z_6 = -\frac{2i}{45} \\ 2iz_5 - 2z_6 = \frac{44}{45} \end{array} \right. \quad (3.4)$$

Al resolver el sistema 3.4 se obtienen los siguientes valores:

$$\begin{aligned} z_2 &= \frac{i}{9}, \\ z_1 &= \frac{4}{45} + \frac{4i}{15} - 3z_2 = \frac{4}{45} + \frac{4i}{15} - \frac{3i}{9} = \frac{4}{45} - \frac{i}{15}, \\ z_4 &= i \left( \frac{4}{45} - \frac{4i}{15} \right) = \frac{4i}{45} + \frac{4}{15}, \\ z_3 &= \frac{41}{45} - 3z_4 = \frac{41}{45} + \left( -\frac{4i}{15} - \frac{4}{5} \right) = \frac{1}{9} - \frac{4i}{15}, \\ z_6 &= i \left( \frac{2}{15} + \frac{2i}{45} \right) = \frac{2i}{15} - \frac{2}{45}, \\ z_5 &= -\frac{2i}{45} - 3z_6 = -\frac{2i}{45} + \left( -\frac{2i}{5} + \frac{2}{15} \right) = \frac{2}{15} - \frac{4i}{9}. \end{aligned}$$

Veamos que los valores obtenidos también son solución al sistema (3.3):

$$\begin{aligned} z_3 + 2iz_5 &= \left(\frac{1}{9} - \frac{4i}{15}\right) + 2i\left(\frac{2}{15} - \frac{4i}{9}\right) \\ &= \left(\frac{1}{9} - \frac{4i}{15}\right) + \left(\frac{4i}{15} + \frac{8}{9}\right) \\ &= 1 \end{aligned}$$

$$\begin{aligned} z_4 + 2iz_6 &= \left(\frac{4i}{45} + \frac{4}{15}\right) + 2i\left(\frac{2i}{15} - \frac{2}{45}\right) \\ &= \left(\frac{4i}{45} + \frac{4}{15}\right) + \left(-\frac{4}{15} - \frac{4i}{45}\right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} -iz_1 + 3z_3 - 2z_5 &= -i\left(\frac{4}{45} - \frac{i}{15}\right) + 3\left(\frac{1}{9} - \frac{4i}{15}\right) - 2\left(\frac{2}{15} - \frac{4i}{9}\right) \\ &= \left(-\frac{4i}{45} - \frac{1}{15}\right) + \left(\frac{1}{3} - \frac{4i}{5}\right) + \left(-\frac{4}{15} + \frac{8i}{9}\right) \\ &= \left(\frac{4}{15} - \frac{8i}{9}\right) + \left(-\frac{4}{15} + \frac{8i}{9}\right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} -iz_2 + 3z_4 - 2z_6 &= -i\left(\frac{i}{9}\right) + 3\left(\frac{4i}{45} + \frac{4}{15}\right) - 2\left(\frac{2i}{15} - \frac{2}{45}\right) \\ &= \frac{1}{9} + \left(\frac{4i}{15} + \frac{4}{5}\right) + \left(-\frac{4i}{15} + \frac{4}{45}\right) = 1 \end{aligned}$$

Por lo tanto, de todo lo anterior se puede concluir que la matriz de Moore esta dada por:

$$X = \begin{bmatrix} \frac{4}{45} - i\frac{1}{15} & \frac{i}{9} \\ \frac{1}{9} - i\frac{4}{15} & \frac{4}{15} + i\frac{4}{45} \\ \frac{2}{15} - i\frac{4}{9} & -\frac{2}{45} + i\frac{2}{15} \end{bmatrix}.$$

Probemos que la matriz  $X$  verifica las condiciones de Moore:

1.  $AX = P_{R(A)}$  :

$$\begin{aligned} AX &= \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \begin{bmatrix} \frac{4}{45} - i\frac{1}{15} & \frac{i}{9} \\ \frac{1}{9} - i\frac{4}{15} & \frac{4}{15} + i\frac{4}{45} \\ \frac{2}{15} - i\frac{4}{9} & -\frac{2}{45} + i\frac{2}{15} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= P_{R(A)}. \end{aligned}$$

2.  $XA = P_{R(X)}$  :

$$\begin{aligned} XA &= \begin{bmatrix} \frac{4}{45} - i\frac{1}{15} & \frac{i}{9} \\ \frac{1}{9} - i\frac{4}{15} & \frac{4}{15} + i\frac{4}{45} \\ \frac{2}{15} - i\frac{4}{9} & -\frac{2}{45} + i\frac{2}{15} \end{bmatrix} \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{9} & \frac{4}{45} + \frac{4}{15}i & \frac{2}{15} - \frac{2}{45}i \\ \frac{4}{45} - \frac{4}{15}i & \frac{41}{45} & \frac{2i}{45} \\ \frac{2}{15} + \frac{2}{45}i & -\frac{2i}{45} & \frac{44}{45} \end{bmatrix} \\ &= P_{R(X)}. \end{aligned}$$

La definición de Moore fue dada en 1935 al particular estilo de Moore. En 1955, Penrose dio la siguiente definición algebraica.

**Definición 3.4.** *Definición de Penrose de la inversa generalizada.*

Si  $A \in M_{m \times n}(\mathbb{C})$ , entonces  $Y \in M_{n \times m}(\mathbb{C})$  es la única matriz tal que:

1.  $AYA = A$ .
2.  $YAY = Y$ .
3.  $(AY)^* = AY$ .
4.  $(YA)^* = YA$ .

Las ecuaciones (1)-(4) son llamadas condiciones de Penrose.

**Observación 3.2.** 1. Si  $m = n$  y  $A$  es no singular, entonces la matriz  $A^{-1}$  satisface las condiciones de Penrose, es decir, la pseudoinversa coincide con la matriz inversa de  $A$ .

2. Si  $m > n$  y las columnas de  $A$  son linealmente independientes, es decir,  $A$  es de rango completo por columnas, su pseudoinversa se obtiene con la siguiente expresión:

$$Y = [A^*A]^{-1} A^*. \quad (3.5)$$

**Demostración:** De la definición de Penrose se tiene que  $AYA = A$ , luego al tomar la adjunta tenemos:  $(AYA)^* = A^*$ . Por propiedades de la adjunta se tiene que:  $(AYA)^* = A^*(AY)^*$ , pero por la condición 4) de la definición de Penrose se tiene que:  $A^*(AY)^* = A^*AY$ , luego:

$$A^*AY = A^*. \quad (3.6)$$

Por otro lado, como  $\text{ran}(A^*A) = \text{ran}(A^*)$ , entonces ocurre que si  $\text{ran}(A^*) = n$ , entonces  $\text{ran}(A^*A) = n$ , por lo tanto,  $N(A^*A) = \{0\}$ ; es decir,  $A^*A$  es invertible y por tanto, al multiplicar por la inversa en ambos lados de la ecuación 3.6, tenemos:

$$Y = [A^*A]^{-1} A^*.$$

□

3. Si  $m < n$  y las filas de  $A$  son linealmente independientes, es decir,  $A$  es de rango completo por filas, su pseudoinversa se obtiene como:

$$Y = A^* [AA^*]^{-1}. \quad (3.7)$$

**Demostración:** De igual manera, al tomar nuevamente la adjunta de la condición 1) tenemos:  $A^* = (AYA)^* = (YA)^* A^*$ . Luego, por la condición 4) tenemos:

$$A^* = YAA^*. \quad (3.8)$$

Además, como  $\text{ran}(AA^*) = \text{ran}(A)$ ; si ocurre que  $\text{ran}(A) = m$ , entonces  $N(AA^*) = \{0\}$ ; por lo tanto,  $AA^*$  es invertible, luego al multiplicar por la inversa en ambos lados de la ecuación 3.8 tenemos:

$$Y = A^* [AA^*]^{-1}.$$

□

Consideremos la definición de Penrose y analicemos el siguiente ejemplo.

**Ejemplo 3.3.** Calcular la inversa de Penrose de la matriz:

$$A = \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \in M_{2 \times 3}(\mathbb{C}).$$

Según la definición de Penrose la matriz inversa  $Y$  es del tipo:

$$Y = \begin{bmatrix} z_1 & z_2 \\ z_3 & z_4 \\ z_5 & z_6 \end{bmatrix} \in M_{3 \times 2}(\mathbb{C}).$$

Es evidente que las filas de  $A$  son linealmente independientes, es decir,  $A$  es de rango completo por filas, además se tiene que  $m = 2$  y  $n = 3$ , por lo que  $m < n$ . Luego, de la Observación 3.2 se tiene que la pseudoinversa de  $A$  se obtiene como:

$$\begin{aligned} Y &= A^* [AA^*]^{-1} \\ &= \begin{bmatrix} 0 & i \\ 1 & 3 \\ -2i & -2 \end{bmatrix} \begin{bmatrix} 5 & 3 - 4i \\ 3 + 4i & 14 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 0 & i \\ 1 & 3 \\ -2i & -2 \end{bmatrix} \begin{bmatrix} \frac{14}{45} & \frac{-1}{15} + \frac{4}{45}i \\ \frac{-1}{15} - \frac{4}{45}i & \frac{1}{9} \end{bmatrix} \end{aligned}$$

Por lo tanto, la matriz pseudoinversa de Penrose de la matriz  $A$  es:

$$Y = \begin{bmatrix} \frac{4}{45} - i\frac{1}{15} & \frac{i}{9} \\ \frac{1}{9} - i\frac{4}{15} & \frac{4}{15} + i\frac{4}{45} \\ \frac{2}{15} - i\frac{4}{9} & -\frac{2}{45} + i\frac{2}{15} \end{bmatrix}.$$

Podemos verificar que se cumplen las condiciones de Penrose:

1.  $AYA = A$  :

$$\begin{aligned} AYA &= \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \begin{bmatrix} \frac{4}{45} - i\frac{1}{15} & \frac{i}{9} \\ \frac{1}{9} - i\frac{4}{15} & \frac{4}{15} + i\frac{4}{45} \\ \frac{2}{15} - i\frac{4}{9} & -\frac{2}{45} + i\frac{2}{15} \end{bmatrix} \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} = A. \end{aligned}$$

2.  $YAY = Y$  :

$$\begin{aligned} YAY &= \begin{bmatrix} \frac{4}{45} - i\frac{1}{15} & \frac{i}{9} \\ \frac{1}{9} - i\frac{4}{15} & \frac{4}{15} + i\frac{4}{45} \\ \frac{2}{15} - i\frac{4}{9} & -\frac{2}{45} + i\frac{2}{15} \end{bmatrix} \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \begin{bmatrix} \frac{4}{45} - i\frac{1}{15} & \frac{i}{9} \\ \frac{1}{9} - i\frac{4}{15} & \frac{4}{15} + i\frac{4}{45} \\ \frac{2}{15} - i\frac{4}{9} & -\frac{2}{45} + i\frac{2}{15} \end{bmatrix} \\ &= \begin{bmatrix} \frac{4}{45} - i\frac{1}{15} & \frac{i}{9} \\ \frac{1}{9} - i\frac{4}{15} & \frac{4}{15} + i\frac{4}{45} \\ \frac{2}{15} - i\frac{4}{9} & -\frac{2}{45} + i\frac{2}{15} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = Y. \end{aligned}$$

3.  $(AY)^* = AY$  :

$$\begin{aligned} AY &= \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \begin{bmatrix} \frac{4}{45} - i\frac{1}{15} & \frac{i}{9} \\ \frac{1}{9} - i\frac{4}{15} & \frac{4}{15} + i\frac{4}{45} \\ \frac{2}{15} - i\frac{4}{9} & -\frac{2}{45} + i\frac{2}{15} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = (AY)^* \end{aligned}$$

$$4. (YA)^* = YA$$

$$\begin{aligned} YA &= \begin{bmatrix} \frac{4}{45} - i\frac{1}{15} & \frac{i}{9} \\ \frac{1}{9} - i\frac{4}{15} & \frac{4}{15} + i\frac{4}{45} \\ \frac{2}{15} - i\frac{4}{9} & -\frac{2}{45} + i\frac{2}{15} \end{bmatrix} \begin{bmatrix} 0 & 1 & 2i \\ -i & 3 & -2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{9} & \frac{4}{45} + \frac{4}{15}i & \frac{2}{15} - \frac{2}{45}i \\ \frac{4}{45} - \frac{4}{15}i & \frac{41}{45} & \frac{2i}{45} \\ \frac{2}{15} + \frac{2}{45}i & -\frac{2i}{45} & \frac{44}{45} \end{bmatrix} = (YA)^* \end{aligned}$$

El hecho mas importante de estas tres definiciones, funcional (pseudoinversa), inversa de Moore e inversa de Penrose es que son equivalentes, es decir, las matrices  $\left[T_A^\dagger\right]_{C',C}$ ,  $X$ ,  $Y$  son iguales. Esto permite definir y denotar como  $A^\dagger$  a la única matriz que satisface dichas definiciones, llamada la inversa de Moore-Penrose de la matriz  $A$  ([40]). El Teorema 3.1 lo garantiza.

**Teorema 3.1.** *Sea  $A \in M_{m \times n}(\mathbb{C})$ . Las definiciones funcional (pseudoinversa), de Moore y de Penrose de la inversa generalizada de  $A$  son equivalentes.*

A continuación estudiamos la forma en como se aplica la inversa generalizada al problema inverso de estimación de PSD.

### 3.1.3. Inversa generalizada aplicada al problema inverso de PSD

Consideremos el problema inverso de estimación de PSD dado por:

$$\tau = Af. \tag{3.9}$$

Sin importar el estimador que se elija, siempre se tiene que la transformación lineal más general que va, en este caso, de los datos  $\tau$  al conjunto de parámetros estimados  $\hat{f}$  puede escribirse en su forma matricial como ([33]):

$$\hat{f} = A^\dagger \tau; \tag{3.10}$$

donde  $A^\dagger$  es la inversa generalizada o inversa de Moore-Penrose de la matriz  $A$ . Dado que la matriz  $A$  generalmente no es cuadrada, entonces su inversa no existe,

por lo que los productos  $AA^\dagger$  y  $A^\dagger A$  no necesariamente son iguales a la matriz identidad. Observemos que es posible definir los datos ajustados  $\hat{\tau}$ , considerando como entrada, los parámetros estimados de  $\hat{f}$ , es decir:

$$\hat{\tau} = A\hat{f}. \quad (3.11)$$

Luego, al reemplazar el valor de  $\hat{f}$  de la ecuación (3.10) en la ecuación (3.11), tendremos lo siguiente:

$$\hat{\tau} = AA^\dagger\tau. \quad (3.12)$$

El producto  $D = AA^\dagger$  se define como la matriz de resolución de datos cuya traza, resulta un indicador del número de incógnitas que pueden ser resueltas ([33]). Una forma de ver la importancia de la matriz de resolución de datos es analizando los errores de ajuste en los datos  $\tau$ , denotados por  $e$ , denominados residuos, los cuales son definidos como ([33]):

$$\begin{aligned} \|e\|^2 &= e^t e \\ &= (\tau - \hat{\tau})^t (\tau - \hat{\tau}) \\ &= (\tau - AA^\dagger\tau)^t (\tau - AA^\dagger\tau) \\ &= [(I - AA^\dagger)\tau]^t [(I - AA^\dagger)\tau] \\ &= [(I - D)\tau]^t [(I - D)\tau] \end{aligned}$$

Observemos que la norma de los residuos se anulan cuando  $D = I$  y se incrementan a medida que  $D$  se desvía de la matriz identidad  $I$ . Por otro lado, la relación entre el conjunto de parámetros verdaderos  $f$  y el estimado  $\hat{f}$  puede obtenerse como:

$$\begin{aligned} \hat{f} &= A^\dagger\tau \\ &= A^\dagger(Af) \\ &= A^\dagger A f \end{aligned}$$

En este caso, se define la matriz  $G = A^\dagger A$ , como la matriz de resolución del modelo ([33]). Cuando los parámetros del modelo son resueltos perfectamente, se cumple que la matriz  $G = I$ . En muchos casos los problemas inversos no se pueden resolver usando únicamente la inversa generalizada, por lo que es necesario regularizar el problema inverso, en la siguiente sección se dan los principales conceptos y resultados sobre el método de regularización de Tikhonov.

## 3.2. Regularización de Tikhonov

En esta sección se describe la técnica de regularización de Tikhonov; para ello, iniciamos motivando con el problema de mínimos cuadrados, el cual es un problema básico cuando se busca resolver problemas de la forma  $Ax = b$ , como lo es el problema de esta tesis. Posteriormente, presentamos el método de ecuaciones normales y concluimos con el método de regularización de Tikhonov usando la Descomposición en Valores Singulares (DVS) de la matriz del sistema.

### 3.2.1. Mínimos cuadrados

En la ciencia y en sus aplicaciones se presentan muchos problemas que se pueden reducir a la solución de un problema de mínimos cuadrados. Actualmente la solución de problemas de mínimos cuadrados son importantes en la teoría y la solución de problemas inversos y problemas mal planteados, es decir, el enfoque de mínimos cuadrados tiene una relación directa con la solución de sistemas de ecuaciones sobredeterminados o subdeterminados, por lo que no busca resolver exactamente las ecuaciones del sistema sino minimizar la suma de los cuadrados de los residuos. Por ejemplo, el problema de ajuste de curvas a cierto conjunto de datos se puede reducir a la solución de un problema de mínimos cuadrados. Consideremos el conjunto de  $m$  datos dados  $(x_1, b_1), (x_2, b_2), \dots, (x_m, b_m)$ , el problema consiste en modelar una función  $f(x)$ , de tal forma que  $b_i \approx f(x_i)$ , por medio de una combinación de  $n$  funciones base  $\phi_1(x), \phi_2(x), \dots, \phi_n(x)$ . Para el caso lineal se tiene que la función que mejor se ajusta a los datos es una combinación lineal de la forma:

$$f(x) = u_1\phi_1(x) + u_2\phi_2(x) + \dots + u_n\phi_n(x). \quad (3.13)$$

Entonces, de manera aproximada los datos deben cumplir que:

$$b_i \approx u_1\phi_1(x_i) + u_2\phi_2(x_i) + \dots + u_n\phi_n(x_i). \quad (3.14)$$

La ecuación (3.14) representa un sistema de  $m$  ecuaciones y  $n$  incógnitas, el cual puede ser escrito en su forma matricial como:

$$\begin{bmatrix} \phi_1(x_1) & \phi_2(x_1) & \dots & \phi_n(x_1) \\ \phi_1(x_2) & \phi_2(x_2) & \dots & \phi_n(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(x_m) & \phi_2(x_m) & \dots & \phi_n(x_m) \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}. \quad (3.15)$$

La matriz  $A = [\phi_j(x_i)]$ , con  $A \in M_{m \times n}(\mathbb{R})$ , es denominada matriz de regresión. Las funciones base  $\phi_j(x)$  pueden elegirse de diferentes maneras, por lo general, suelen usarse funciones racionales,  $\phi_j(x) = x^{j-1}/(\alpha_0 + \alpha_1 x + \dots + \alpha_{n-1} x^{n-1})$  con  $\alpha_0, \dots, \alpha_{n-1}$  parámetros dados; polinomios,  $\phi_j(x) = x^{j-1}$ ; funciones exponenciales,  $\phi_j(x) = \exp^{-\lambda_j x}$ , donde los  $\lambda_j$  son denominados parámetros de decaimiento. El ajuste de curvas normalmente resulta en un problema sobredeterminado, es decir, se cumple que  $m > n$  y, por lo tanto, el problema tendrá solución solo si  $b$  se encuentra en el espacio imagen de la matriz  $A$ . Normalmente se tiene que  $b \notin \text{Im}(A)$ , por lo que no es posible encontrar una solución directa  $u$  del sistema (3.15). Entonces, el ajuste de curvas se convierte en un problema de mínimos cuadrados que consiste en encontrar el vector de coeficientes  $u$  que minimice la diferencia  $Ax - b$ , para todo  $x \in \mathbb{R}^n$ . Si consideramos la norma euclidiana en  $\mathbb{R}^m$ , entonces el problema consiste en resolver:

$$\|Au - b\|_2^2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2. \quad (3.16)$$

Esto equivale a encontrar el vector de coeficientes  $u = (u_1, u_2, \dots, u_n)^t$  que minimiza la suma de cuadrados:

$$\min_{u \in \mathbb{R}^n} \sum_{i=1}^m (u_1 \phi_1(x_i) + u_2 \phi_2(x_i) + \dots + u_n \phi_n(x_i) - b_i)^2. \quad (3.17)$$

Una vez encontrado el vector  $u$ , se sustituye en la ecuación (3.13) para obtener la función  $f(x)$  que mejor ajusta los datos dados. En general, el ajuste de curvas es usado para encontrar una función que responda a una muestra de datos obtenidos por alguna medición. En particular, nos da una idea de la solución de un problema sobredeterminado que se puede resolver desde el enfoque de mínimos cuadrados. Esto nos dice que la estimación de parámetros por medio de una regresión lineal

nos conduce a un problema de mínimos cuadrados.

**Definición 3.5.** Dada la matriz  $A \in M_{m \times n}(\mathbb{R})$  con  $m \geq n$  y el vector  $b \in \mathbb{R}^m$ , el problema lineal de mínimos cuadrados consiste en hallar un vector  $x^\dagger \in \mathbb{R}^n$  tal que:

$$\|Ax^\dagger - b\|_2^2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2. \quad (3.18)$$

La siguiente definición nos dice cuando un vector es solución de mínimos cuadrados.

**Definición 3.6.** Sean  $A \in M_{m \times n}(\mathbb{R})$  y  $b \in \mathbb{R}^m$ . Un vector  $u \in \mathbb{R}^n$ , se dice que es una solución de mínimos cuadrados del sistema de ecuaciones  $Ax = b$ , si:

$$\|Au - b\|_2^2 \leq \|Ax - b\|_2^2, \quad \forall x \in \mathbb{R}^n;$$

es decir,

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 = \|Au - b\|_2^2.$$

Existen diferentes métodos para resolver el problema de minimización planteado por la ecuación (3.18), uno de ellos es el método de ecuaciones normales, el cual está basado en el concepto de proyección ortogonal y que tiene relación con conceptos vistos anteriormente, como la inversa de Moore-Penrose.

### 3.2.2. Método de ecuaciones normales

En esta sección se analizará la parte geométrica y algebraica del problema de mínimos cuadrados y se estudiará el método de las ecuaciones normales para resolverlo. Debido a que el problema inverso de estimación de PSD es un problema mal condicionado podemos usar el método de ecuaciones normales para resolverlo, además, este método es una introducción a la regularización de Tikhonov.

Sean la matriz  $A \in M_{m \times n}(\mathbb{R})$  con  $m \geq n$ , los vectores  $x \in \mathbb{R}^n$  y  $b \in \mathbb{R}^m$ , y consideremos la norma euclidiana en  $\mathbb{R}^m$  denotada por  $\|\cdot\|_2$ , la cual usaremos en toda esta sección. Para analizar el argumento geométrico del problema de mínimos cuadrados, examinamos la ecuación matricial  $Ax = b$ . Para esto, supongamos que el vector  $b \notin \text{Im}(A)$ , entonces buscamos el vector  $s \in \text{Im}(A)$  más cercano a  $b$ , es decir, la distancia entre  $b$  y  $\text{Im}(A)$  sea mínima, que equivale a que  $s$  satisfaga:

$$\|s - b\|_2 = \min_{t \in \text{Im}(A)} \|t - b\|_2.$$

Geoméricamente, se tiene que el vector  $s$  es una proyección de  $b$  sobre  $Im(A)$  tal que el vector diferencia  $s - b$  sea ortogonal al espacio imagen  $Im(A)$ , como se muestra en la Figura 3.1. El vector  $s$  es llamado proyección ortogonal de  $b$  sobre  $Im(A)$ , y debido a que  $s \in Im(A)$ , entonces existe un vector  $u \in \mathbb{R}^n$  tal que:

$$Au = s.$$

Así, el problema lineal de mínimos cuadrados consiste en hallar un vector  $u \in \mathbb{R}^n$  que cumpla lo siguiente:

$$\|Au - b\|_2^2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2. \quad (3.19)$$

Para hallar el vector  $u$ , se resuelven las ecuaciones normales dadas por:

$$A^t Au = A^t b. \quad (3.20)$$

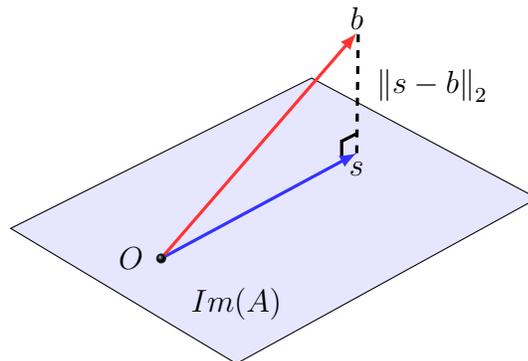


Figura 3.1: Proyección ortogonal de  $b$  sobre  $Im(A)$ .

Consideremos la matriz  $A^t A \in M_{n \times n}(\mathbb{R})$  y supongamos que la matriz  $A$  es de rango completo por columnas. Probemos que la matriz  $A^t A$  es invertible.

Sea  $x \in N_{A^t A}$ , entonces se cumple que:

$$\begin{aligned} A^t Ax &= 0, \\ x^t A^t Ax &= x^t 0. \end{aligned}$$

Recordemos que para cualesquiera dos matrices  $L, K \in M_{m \times n}(\mathbb{R})$  se cumple que  $L^t K^t = (KL)^t$ , y si identificamos el vector  $x \in \mathbb{R}^n$  como una matriz de  $n \times 1$ ,

tendremos que:

$$\begin{aligned}x^t A^t A x &= x^t 0, \\(Ax)^t (Ax) &= 0.\end{aligned}$$

Además, sabemos que para todo  $y \in \mathbb{R}^m$  se cumple que:

$$y^t y = \langle y, y \rangle;$$

así, tendremos que:

$$(Ax)^t (Ax) = \langle Ax, Ax \rangle = \|Ax\|_2^2 = 0.$$

Luego, se debe cumplir que  $Ax = 0$ , es decir,  $x \in N_A$ . Así, por la arbitrariedad de  $x$ , tendremos que  $N_{A^t A} \subseteq N_A$ . Debido a que la matriz  $A$  es de rango completo por columnas se tiene que  $N_A = \{0\}$  y  $\text{ran}(A) = n$ . De aquí, se tiene que  $x \in N_{A^t A} \subseteq N_A = \{0\}$ , es decir, la única solución de la ecuación  $A^t A x = 0$ , es el vector  $x = 0$ . Luego, la matriz  $A^t A$  también es de rango completo por columnas, esto significa que las columnas de  $A^t A$  son linealmente independientes y que  $\text{ran}(A^t A) = n$ . Entonces, la forma escalonada por renglones de la matriz  $A^t A$  tienen  $n$  pivotes, es decir la forma escalonada por renglones de  $A^t A$  es la matriz identidad  $I_n$ . Por lo tanto, la matriz  $A^t A$  es invertible. Así, al multiplicar la matriz  $(A^t A)^{-1}$  por la ecuación (3.20) tendremos que:

$$\begin{aligned}(A^t A)^{-1} A^t A u &= (A^t A)^{-1} A^t b, \\u &= (A^t A)^{-1} A^t b.\end{aligned}$$

Al definir la matriz  $A^\dagger$  como:

$$A^\dagger = (A^t A)^{-1} A^t, \tag{3.21}$$

podemos reescribir la única solución de las ecuaciones normales como:

$$u = A^\dagger b. \tag{3.22}$$

El vector  $u \in \mathbb{R}^n$  es llamado solución de mínimos cuadrados de la ecuación  $Ax = b$ .

De esta forma, según ([31]) es posible definir la función lineal:

$$\begin{aligned} P_{C_A} : C_{A^\dagger} &\longrightarrow C_A \\ u &\longrightarrow Au, \end{aligned} \quad (3.23)$$

que transforma el vector solución  $u$  en el vector proyección  $s$ , mientras que la función inversa transforma el vector proyección  $s$  en el vector  $u$  se define como:

$$\begin{aligned} P_{C_{A^\dagger}} : C_A &\longrightarrow C_{A^\dagger} \\ y &\longrightarrow A^\dagger y. \end{aligned} \quad (3.24)$$

El método de las ecuaciones normales nos da una solución al sistema lineal mal condicionado  $Ax = b$  usando la inversa de la matriz cuadrada  $A^t A$ . Este método es muy parecido al de la inversa generalizada, usado para resolver la ecuación (3.9). El estudio de estos métodos de solución se usan en la mayoría de los casos para problemas mal condicionados o mal planteados.

### 3.2.3. Problemas de mínimos cuadrados con rango deficiente

El método de solución de ecuaciones normales para el problema de mínimos cuadrados supone que la matriz  $A$  es de rango completo, es decir, sus renglones o sus columnas son linealmente independientes. En este caso, supongamos que el rango de la matriz  $A$  de  $m \times n$  es deficiente, entonces la matriz  $A^t A$  no es invertible, por lo que procedemos de manera diferente. Cambiamos la base de  $C_A$  (o  $R_A$ ) por una base que tiene vectores ortonormales. Una forma es usar la Descomposición en Valores Singulares (DVS) de una matriz dada por la Definición 1.23.

Sea el problema de mínimos cuadrados dado por la ecuación:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2. \quad (3.25)$$

Si consideramos la DVS de la matriz  $A$ , entonces el problema de mínimos cuadrados se puede reescribir como:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 = \min_{x \in \mathbb{R}^n} \|U\Sigma V^t x - b\|_2^2.$$

Debido a que la matriz  $U$  es ortogonal, y por el Teorema 1.3 se cumple que:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \|U\Sigma V^t x - b\|_2^2 &= \min_{x \in \mathbb{R}^n} \|U(\Sigma V^t x - U^t b)\|_2^2 \\ &= \min_{x \in \mathbb{R}^n} \|\Sigma V^t x - U^t b\|_2^2. \end{aligned}$$

Definimos,

$$c = U^t b \quad e \quad y = V^t x,$$

con  $c \in \mathbb{R}^m$  y  $y \in \mathbb{R}^n$ . Dado que la matriz  $V^t$  es ortogonal,  $V$  es su matriz inversa, entonces:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \|\Sigma V^t x - U^t b\|_2^2 &= \min_{y \in \mathbb{R}^n} \|\Sigma V^t V y - U^t b\|_2^2 \\ &= \min_{y \in \mathbb{R}^n} \|\Sigma y - c\|_2^2. \end{aligned}$$

Los elementos de la matriz  $\Sigma$  son todos cero fuera de su diagonal principal, y sobre esta diagonal solamente los valores singulares  $\sigma_1, \dots, \sigma_r$  son distintos de cero. Así, se tiene que:

$$\begin{aligned} \min_{y \in \mathbb{R}^n} \|\Sigma y - c\|_2^2 &= \min_{y \in \mathbb{R}^n} \left( \sqrt{\sum_{i=1}^r |\sigma_i y_i - c_i|^2} \right)^2 \\ &= \min_{y \in \mathbb{R}^n} \left( \sum_{i=1}^r |\sigma_i y_i - c_i|^2 \right) \\ &= \min_{y \in \mathbb{R}^n} (|\sigma_1 y_1 - c_1|^2 + \dots + |\sigma_r y_r - c_r|^2 + |0 - c_{r+1}|^2 + \dots + |0 - c_m|^2) \\ &= \min_{y \in \mathbb{R}^n} \left( \sum_{i=1}^r |\sigma_i y_i - c_i|^2 + \sum_{i=r+1}^m |c_i|^2 \right). \end{aligned}$$

De la última igualdad se puede observar que la suma puede ser mínima cuando  $|\sigma_i y_i - c_i|^2 = 0$ , por lo que todo vector  $y$  con componentes:

$$y_i = \begin{cases} \frac{c_i}{\sigma_i}, & i = 1, \dots, r, \\ \text{arbitrario}, & i = r + 1, \dots, n \end{cases} \quad (3.26)$$

minimizan el tamaño del residuo  $\Sigma y - c$ . Por lo tanto, el conjunto de vectores  $Vy = x$ , donde  $y_i$  esta dado por la ecuación (3.26) nos da una familia de soluciones de mínimos cuadrados. De la familia de soluciones, seleccionamos el vector de

menor norma euclidiana y lo definimos como:

$$x^\dagger = Vy.$$

El vector  $x^\dagger$  se llama solución de mínimos cuadrados de norma mínima. Puesto que  $V$  es una matriz ortogonal, y por el Teorema 1.3 se cumple que:

$$\|x^\dagger\|_2 = \|Vy\|_2 = \|y\|_2.$$

Como queremos minimizar  $\|x^\dagger\|_2$  y las componentes del vector  $y$  están dadas por (3.26), entonces basta tomar:

$$y_{i+r} = \dots = y_n = 0.$$

Así, tendremos que:

$$\begin{aligned} x^\dagger &= Vy \\ &= \begin{bmatrix} v_1 \\ \vdots \\ v_r \\ v_{r+1} \\ \vdots \\ v_n \end{bmatrix}_{n \times n} \begin{bmatrix} y_1 \\ \vdots \\ y_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n \times 1} \\ &= \sum_{i=1}^r v_i y_i. \end{aligned}$$

Por lo tanto, la solución de mínimos cuadrados de norma mínima para matrices con rango deficiente es:

$$x^\dagger = \sum_{i=1}^r \frac{c_i}{\sigma_i} v_i. \quad (3.27)$$

De forma general, la solución de mínimos cuadrados de norma mínima  $x^\dagger$  para la ecuación  $Ax = b$  se obtiene calculando una DVS de la matriz  $A = U\Sigma V^t$ , y formando el producto  $c = U^t b$ , para así obtener  $x^\dagger = (c_1/\sigma_1)v_1 + \dots + (c_r/\sigma_r)v_r$ .

En la siguiente sección veremos la definición de un problema mal condicionado y teoremas que nos ayudarán para su solución. Se estudiarán conceptos de error,

como error relativo. También se analizarán los casos cuando la matriz  $A$  y el vector  $b$  tienen errores o perturbaciones y para esto se recordarán conceptos como el número de condición de una matriz.

### 3.2.4. Métodos de Regularización

Como en las secciones anteriores, nos interesa resolver un problema lineal mal planteado que al discretizarlo da lugar a un sistema de ecuaciones lineales de la forma  $Ax = b$ , donde la matriz  $A$  es mal condicionada. La idea de la regularización es reemplazar el problema mal planteado por otro bien planteado de modo que la solución del nuevo problema sea una buena aproximación de la solución buscada ([31]). El matemático ruso Andréi Tikhonov fue el primero en sugerir un nuevo enfoque para resolver problemas mal planteados. Esta formulación modifica el problema mal planteado en otro bien planteado. Así, Tikhonov estableció y justificó el método más popular para resolver problemas mal planteados, llamado método de regularización.

Como lo asegura [20], el método de regularización consiste en elegir una colección de problemas bien planteados, esta colección de problemas puede ser finita o infinita y depende del llamado parámetro de regularización que es elegido a priori, es decir, antes de calcular la solución que regulariza el problema. Para cada valor del parámetro, se tiene un problema distinto, y su solución se llama solución regularizada. En general, dicho parámetro depende de los límites de la solución exacta; los cuales no se conocen, por lo que a veces es necesario usar estrategias para elegir el parámetro de regularización adecuado, como lo son el Principio de Discrepancia de Morozov y el Método de la L-Curva. Una vez elegido el valor del parámetro de regularización, la solución regularizada se toma como aproximación de la solución del problema mal planteado, como se asegura en [31]. En resumen, el método de regularización consiste en calcular las soluciones regularizadoras en función del parámetro de regularización y elegir correctamente el parámetro de regularización.

#### Regularización mediante DVS truncada

El primer método que se usa para regularizar problemas discretos mal planteados es la DVS. Dicho método está basado en los valores singulares de la matriz, definidos en 1.22. Sea la matriz  $A \in M_{m \times n}(\mathbb{K})$  con  $m \geq n$  y  $\text{ran}(A) = r$ , conside-

remos su descomposición en valores singulares, la cual se puede escribir como:

$$A = \underbrace{\begin{bmatrix} u_1 & \dots & u_r & u_{r+1} & \dots & u_m \end{bmatrix}}_U \underbrace{\begin{bmatrix} \sigma_1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & 0 \\ 0 & \dots & \sigma_r & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_\Sigma \underbrace{\begin{bmatrix} v_1 & \dots & v_r & v_{r+1} & \dots & v_n \end{bmatrix}^t}_{V^t}$$

De aquí, se tiene que:

$$\begin{aligned} A &= \begin{bmatrix} u_1 & \dots & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix} \begin{bmatrix} v_1 & \dots & v_r \end{bmatrix}^t + \begin{bmatrix} u_{r+1} & \dots & u_n \end{bmatrix} [0] \begin{bmatrix} v_{r+1} & \dots & v_n \end{bmatrix}^t \\ &= \begin{bmatrix} u_1 & \dots & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_r \end{bmatrix} \begin{bmatrix} v_1 & \dots & v_r \end{bmatrix}^t. \end{aligned}$$

Por la Definición 1.24 sabemos que las columnas de  $U$ , que son  $u_1, u_2, \dots, u_r$  se denominan vectores singulares izquierdos de  $A$ , mientras que las columnas de  $V$ , que son  $v_1, \dots, v_r$  son llamados vectores singulares derechos de  $A$ . Las matrices  $U$  y  $V$  no están determinadas en forma única por  $A$ , en cambio  $\Sigma$  si, porque contiene los valores singulares  $\sigma_1, \dots, \sigma_r$  de  $A$ .

Al realizar el producto de las matrices, se obtiene lo siguiente:

$$A = \begin{bmatrix} \sigma_1 u_1 & \dots & \sigma_r u_r \end{bmatrix} \begin{bmatrix} v_1^t \\ \vdots \\ v_r^t \end{bmatrix} = \sigma_1 u_1 v_1^t + \dots + \sigma_r u_r v_r^t.$$

Por lo tanto, la matriz  $A$  se puede escribir como:

$$A = \sum_{i=1}^r \sigma_i u_i v_i^t. \quad (3.28)$$

El Lema 1.2 garantiza la escritura de la matriz  $A$  como la suma de matrices  $u_i v_i^t$  de rango uno. En algunas ocasiones los valores singulares más pequeños de  $A$  son los responsables de que la matriz resulte mal condicionada, por lo que primero

intentamos removerlos. Al remover los valores singulares más pequeños, truncamos la suma (3.28). Para esto, dado un valor  $k \in \{1, 2, \dots, r\}$ , removemos los  $r - k$  valores singulares más pequeños de la matriz  $A$  y obtenemos una nueva matriz dada por:

$$A_k = \sum_{i=1}^k \sigma_i u_i v_i^t. \quad (3.29)$$

Si sustituimos la matriz  $A$  por la matriz  $A_k$  en el problema de mínimos cuadrados dado por la ecuación (3.25), obtenemos la siguiente colección de problemas:

$$\min_{x \in \mathbb{R}^n} \|A_k x - b\|_2^2, \quad k = 1, 2, \dots, r. \quad (3.30)$$

Para cada problema de la colección se tiene una solución de norma mínima. Dicha solución se obtiene de la misma manera que la solución (3.27). En este caso, se tiene que  $c_i = u_i^t b$ , con  $i = 1, \dots, k$ . Luego, se tiene que la solución para cada problema de la familia (3.30) está dada por:

$$x_k = \sum_{i=1}^k \frac{u_i^t b}{\sigma_i} v_i, \quad k = 1, 2, \dots, r. \quad (3.31)$$

Este método se conoce como DVS truncada, y a la matriz  $A_k$  se le conoce como TDVS de  $A$  con nivel de truncamiento  $k$  ([31]). La idea de la regularización por DVS truncada es quedarse con los coeficientes  $u_i^t b$  asociados a los valores singulares más grandes. El valor de  $k$  es el llamado parámetro de regularización, el cual representa la cantidad de valores singulares más grandes que contribuyen con la matriz  $A_k$ , y la solución regularizada es  $x_k$  ([31]).

En la práctica, muchos problemas discretizados de la forma  $Ax = b$  se ven afectados por errores en las observaciones  $b$ , los cuales repercuten en la solución del problema. Por lo que en ocasiones el vector de observaciones se define como:

$$b = b_{exacto} + \epsilon,$$

donde  $b_{exacto}$  es desconocido, y representa los datos exactos del problema,  $\epsilon$  es un vector aleatorio no observable y representa el error en los datos ([31]). Al sustituir

este nuevo vector en la solución regularizada  $x_k$ , obtenemos lo siguiente:

$$\begin{aligned} x_k &= \sum_{i=1}^k \frac{u_i^t b}{\sigma_i} v_i \\ &= \sum_{i=1}^k \frac{u_i^t (b_{exacto} + \epsilon)}{\sigma_i} v_i \\ &= \sum_{i=1}^k \left( \frac{u_i^t b_{exacto}}{\sigma_i} v_i + \frac{u_i^t \epsilon}{\sigma_i} v_i \right). \end{aligned}$$

Por lo tanto, la solución regularizada de norma mínima se puede expresar como:

$$x_k = \sum_{i=1}^k \frac{u_i^t b_{exacto}}{\sigma_i} v_i + \sum_{i=1}^k \frac{u_i^t \epsilon}{\sigma_i} v_i, \quad k = 1, 2, \dots, r. \quad (3.32)$$

De acuerdo con [31], el primer término representa la solución de norma mínima del problema libre de errores:

$$\min_{x \in \mathbb{R}^n} \|A_k x - b_{exacto}\|_2^2,$$

mientras que el segundo término representa la propagación del error en la solución del problema perturbado:

$$\min_{x \in \mathbb{R}^n} \|A_k x - (b_{exacto} + \epsilon)\|_2^2.$$

El objetivo principal de la DVS truncada, es elegir correctamente el nivel de truncamiento  $k$  que permita atenuar la influencia del error en la solución regularizada.

Una forma de responder la pregunta, ¿cuál es el nivel de truncamiento óptimo que disminuye la influencia del error en la solución regularizada?, es mediante la regularización de Tikhonov que se estudiará a continuación.

### Regularización de Tikhonov

Un método clásico para regularizar sistemas lineales finitos sobredeterminados, es decir, problemas discretos mal planteados de la forma  $Ax = y$ , es penalizar la norma de la solución en el problema de mínimos cuadrados (3.18), este método se conoce como regularización de Tikhonov ([20, 31]). Dado que las soluciones de mínimos cuadrados del problema discreto mal planteado dependen de coeficientes

que tienden al infinito cuando los valores singulares tienden a cero, se incluye una penalización a la norma de la solución y se incorpora al problema de mínimos cuadrados, es decir, se penaliza  $\|x\|_2$ , esto para intentar suprimir el efecto de estos cambios. Así, de acuerdo con [31], la regularización de Tikhonov consiste en hallar un  $x$  que resuelva el problema:

$$\min_{x \in \mathbb{R}^n} \{ \|Ax - y\|_2^2 + \alpha^2 \|x\|_2^2 \}, \quad (3.33)$$

dado un factor  $\alpha > 0$ , el cual es llamado factor de Tikhonov, y actúa como parámetro de regularización. La solución regularizada del problema (3.33) es denotada por  $x_\alpha$ . El parámetro de regularización  $\alpha$  controla el balance entre el tamaño del residuo  $r_\alpha = Ax_\alpha - y$  y el tamaño de la solución regularizada  $x_\alpha$ . Esto es, si el valor de  $\alpha$  es muy grande, entonces el tamaño de la solución regularizada disminuye, y si el valor de  $\alpha$  es muy pequeño, entonces el tamaño de la solución regularizada aumenta, es decir:

$$\begin{aligned} \|x_\alpha\|_2 &\rightarrow 0 \quad \text{cuando} \quad \alpha \rightarrow \infty, \\ \|x_\alpha\|_2 &\rightarrow \infty \quad \text{cuando} \quad \alpha \rightarrow 0. \end{aligned}$$

Por otro lado, mientras más pequeño sea el tamaño de  $\alpha$ , la solución regularizada se acerca más a la solución de norma mínima  $x^\dagger$ , y si el tamaño de  $\alpha$  es muy grande, entonces la solución regularizada se aleja de la solución de norma mínima, esto es:

$$\begin{aligned} \|x_\alpha - x^\dagger\|_2 &\rightarrow 0 \quad \text{cuando} \quad \alpha \rightarrow 0, \\ \|x_\alpha - x^\dagger\|_2 &\rightarrow \infty \quad \text{cuando} \quad \alpha \rightarrow \infty. \end{aligned}$$

En general, un método de regularización consiste en hallar un  $\hat{x}$  que logre minimizar la norma  $\|Kx - y\|_2$ . En otras palabras, dado el operador  $K : X \rightarrow Y$  lineal y acotado se busca un  $\hat{x}$  con  $\|K\hat{x} - y\|_2 \leq \|Kx - y\|_2$  para todo  $x \in X$  y  $y \in Y$ , tal que sea solución de la ecuación normal  $K^*K\hat{x} = K^*y$ , donde  $K^* : Y \rightarrow X$  denota el operador adjunto de  $K$ .

Consideremos lo siguiente:

$$\begin{aligned} \|Kx - y\|_2 &= \|Kx + K\hat{x} - K\hat{x} - y\|_2 \\ &= \|(K\hat{x} - y) + (Kx - K\hat{x})\|_2. \end{aligned} \quad (3.34)$$

Por el teorema del binomio se cumple que para todo  $x \in X$  e  $y \in Y$ :

$$\begin{aligned}\|x - y\|_2^2 &= \|x\|_2^2 + \|y\|_2^2 - 2\operatorname{Re}(\langle x, y \rangle), \\ \|x\|_2^2 &= \langle x, x \rangle,\end{aligned}$$

y usando la propiedad 3 de la Definición 1.16 es posible reescribir la igualdad (3.34) como:

$$\begin{aligned}\|Kx - y\|_2^2 &= \|(K\hat{x} - y) + (Kx - K\hat{x})\|_2^2 \\ &= \langle K\hat{x} - y + Kx - K\hat{x}, K\hat{x} - y + Kx - K\hat{x} \rangle \\ &= \langle K\hat{x} - y, K\hat{x} - y + Kx - K\hat{x} \rangle + \langle Kx - K\hat{x}, K\hat{x} - y + Kx - K\hat{x} \rangle \\ &= \langle K\hat{x} - y, K\hat{x} - y \rangle + \langle K\hat{x} - y, Kx - K\hat{x} \rangle + \langle Kx - K\hat{x}, K\hat{x} - y \rangle \\ &\quad + \langle Kx - K\hat{x}, Kx - K\hat{x} \rangle \\ &= \|K\hat{x} - y\|_2^2 + 2\operatorname{Re}(\langle K\hat{x} - y, K(x - \hat{x}) \rangle) + \|K(x - \hat{x})\|_2^2.\end{aligned}$$

De aquí, se tiene que:

$$\|Kx - y\|_2^2 - \|K\hat{x} - y\|_2^2 = 2\operatorname{Re}(\langle K\hat{x} - y, K(x - \hat{x}) \rangle) + \|K(x - \hat{x})\|_2^2.$$

Luego, por la Definición 1.29 se obtiene lo siguiente:

$$\|Kx - y\|_2^2 - \|K\hat{x} - y\|_2^2 = 2\operatorname{Re}(\langle K^*(K\hat{x} - y), x - \hat{x} \rangle) + \|K(x - \hat{x})\|_2^2. \quad (3.35)$$

Supongamos que  $\hat{x}$  minimiza la ecuaciones normales, entonces se cumple que  $K^*(K\hat{x} - y) = 0$ , por lo que de la ecuación (3.35) tendremos que:

$$\begin{aligned}\|Kx - y\|_2^2 - \|K\hat{x} - y\|_2^2 &= 2\operatorname{Re}(\langle 0, x - \hat{x} \rangle) + \|K(x - \hat{x})\|_2^2 \\ &= \|K(x - \hat{x})\|_2^2 \geq 0.\end{aligned} \quad (3.36)$$

Por lo tanto, se obtiene que:

$$\|Kx - y\|_2^2 \geq \|K\hat{x} - y\|_2^2, \quad \forall x, \hat{x} \in X;$$

esto significa que  $\hat{x}$  minimiza la norma  $\|Kx - y\|_2^2$ . Por otro lado, si suponemos que  $\hat{x}$  minimiza la norma  $\|Kx - y\|_2^2$ , entonces al sustituir cualquier  $x \in X$  se debe cumplir que  $\|Kx - y\|_2^2 \geq \|K\hat{x} - y\|_2^2$ .

En particular, al sustituir  $x = \hat{x} + \alpha z$  en la ecuación (3.35) se obtiene que:

$$2\operatorname{Re}(\langle K^*(K\hat{x} - y), \hat{x} + \alpha z - \hat{x} \rangle) + \|K(\hat{x} + \alpha z - \hat{x})\|_2^2 \geq 0,$$

para cualquier  $\alpha > 0$  y  $z = x - \hat{x} \in X$ . Luego, por propiedades del producto interior en  $X$  se tiene que:

$$2\alpha\operatorname{Re}(\langle K^*(K\hat{x} - y), z \rangle) + \alpha^2 \|Kz\|_2^2 \geq 0. \quad (3.37)$$

Observemos que al dividir la ecuación (3.37) por  $\alpha > 0$  y por  $\alpha \rightarrow 0$  se obtiene la ecuación (3.35), la cual solo se cumple si:  $\operatorname{Re}(\langle K^*(K\hat{x} - y), z \rangle) \geq 0$ ,  $\forall z \in X$ , es decir, si se cumple que  $K^*(K\hat{x} - y) = 0$ . Así,  $\hat{x}$  es solución a la ecuación normal. Los cálculos anteriores justifican el siguiente lema.

**Lema 3.1.** Sea  $X$  e  $Y$  espacios de Hilbert,  $K : X \rightarrow Y$  un operador lineal y acotado e  $y \in Y$ . Entonces existe un  $\hat{x} \in X$  con  $\|K\hat{x} - y\|_2 \leq \|Kx - y\|_2$  para todo  $x \in X$  si y solo si  $\hat{x}$  es solución de la ecuación normal  $K^*K\hat{x} = K^*y$ .

**Nota 3.2.** Todos estos resultados se pueden generalizar para cualquier norma. Para el problema inverso de estimación de tamaños de partícula se considera la norma Euclidiana.

Una consecuencia del Lema (3.1) es que podemos penalizar  $\|x\|_2$  reemplazando la ecuación de primer tipo  $K^*K\hat{x} = K^*y$ , por una ecuación de segundo tipo dada por  $\alpha^2\hat{x} + K^*K\hat{x} = K^*y$ , de acuerdo al lenguaje de la teoría de ecuaciones integrales ([20]). Este punto de vista nos conduce al siguiente problema de minimización. Dado un operador lineal, acotado  $K : X \rightarrow Y$  y el vector de datos  $y \in Y$ , se quiere determinar  $x_\alpha \in X$  que minimiza el funcional de Tikhonov definido como:

$$J_\alpha(x) := \|Ax - y\|_2^2 + \alpha^2 \|x\|_2^2, \quad \forall x \in X, \quad \alpha > 0. \quad (3.38)$$

Entonces, el problema de hallar un  $x$  que resuelva la ecuación (3.33) se convierte en determinar  $x_\alpha \in X$  que minimiza el funcional de Tikhonov dado por la ecuación (3.38). Para esto, es necesario probar primero la existencia de un mínimo de  $J_\alpha$ . Consideremos una sucesión minimizante de vectores  $\{x_n\} \subset X$ , dada por la Definición 1.28. En este caso, dicha sucesión de vectores cumple lo siguiente:

$$\lim_{n \rightarrow \infty} J_\alpha(x_n) = I := \inf_{x \in X} J_\alpha(x). \quad (3.39)$$

Probemos que la sucesión  $\{x_n\}$  es una sucesión de Cauchy.

Sean  $n, m \in \mathbb{Z}$ , se tiene lo siguiente:

$$\begin{aligned} J_\alpha(x_n) + J_\alpha(x_m) &= \|Kx_n - y\|_2^2 + \alpha^2 \|x_n\|_2^2 + \|Kx_m - y\|_2^2 + \alpha^2 \|x_m\|_2^2 \\ &= \|Kx_n - y\|_2^2 + \|Kx_m - y\|_2^2 + \alpha^2 \|x_n\|_2^2 + \alpha^2 \|x_m\|_2^2. \end{aligned} \quad (3.40)$$

Por otro lado, por el teorema del binomio se tiene que:

$$\begin{aligned} \|x + y\|_2^2 &= \|x\|_2^2 + \|y\|_2^2 + 2\operatorname{Re}(\langle x, y \rangle), \\ \|x - y\|_2^2 &= \|x\|_2^2 + \|y\|_2^2 - 2\operatorname{Re}(\langle x, y \rangle). \end{aligned}$$

Luego, una consecuencia del teorema del binomio es:

$$\|x + y\|_2^2 + \|x - y\|_2^2 = 2\|x\|_2^2 + 2\|y\|_2^2. \quad (3.41)$$

Así, aplicando la igualdad (3.41) obtenemos:

$$\begin{aligned} \frac{\alpha^2}{2} (\|x_n - x_m\|_2^2 + \|x_n + x_m\|_2^2) &= \frac{\alpha^2}{2} (2\|x_n\|_2^2 + 2\|x_m\|_2^2) \\ &= \alpha^2 \|x_n\|_2^2 + \alpha^2 \|x_m\|_2^2. \end{aligned} \quad (3.42)$$

Luego, al sustituir la ecuación (3.42) en la ecuación (3.40) tendremos que:

$$\begin{aligned} J_\alpha(x_n) + J_\alpha(x_m) &= \|Kx_n - y\|_2^2 + \|Kx_m - y\|_2^2 + \frac{\alpha^2}{2} (\|x_n - x_m\|_2^2 + \|x_n + x_m\|_2^2) \\ &= \|Kx_n - y\|_2^2 + \|Kx_m - y\|_2^2 + \frac{\alpha^2}{2} \|x_n - x_m\|_2^2 + \frac{\alpha^2}{2} \|x_n + x_m\|_2^2 \\ &= \|Kx_n - y\|_2^2 + \|Kx_m - y\|_2^2 + \frac{\alpha^2}{2} \|x_n - x_m\|_2^2 + \frac{2\alpha^2}{4} \|x_n + x_m\|_2^2. \end{aligned}$$

Luego, se obtiene que:

$$J_\alpha(x_n) + J_\alpha(x_m) = \|Kx_n - y\|_2^2 + \|Kx_m - y\|_2^2 + \frac{\alpha^2}{2} \|x_n - x_m\|_2^2 + 2\alpha^2 \left\| \frac{x_n + x_m}{2} \right\|_2^2. \quad (3.43)$$

Por otra parte, se tiene que:

$$\begin{aligned}
2 \left\| K \left( \frac{x_n + x_m}{2} \right) - y \right\|_2^2 &= 2 \left\| K \left( \frac{x_n + x_m}{2} \right) - \frac{1}{2}y - \frac{1}{2}y \right\|_2^2 \\
&= 2 \left\| \frac{Kx_n - y}{2} + \frac{Kx_m - y}{2} \right\|_2^2 \\
&= \frac{2}{4} \|(Kx_n - y) + (Kx_m - y)\|_2^2 \\
&= \frac{1}{2} \|(Kx_n - y) + (Kx_m - y)\|_2^2.
\end{aligned}$$

De aquí, obtenemos:

$$4 \left\| K \left( \frac{x_n + x_m}{2} \right) - y \right\|_2^2 = \|(Kx_n - y) + (Kx_m - y)\|_2^2. \quad (3.44)$$

Además, por la ecuación (3.41) se obtiene que:

$$\begin{aligned}
\|(Kx_n - y) + (Kx_m - y)\|_2^2 + \|(Kx_n - y) - (Kx_m - y)\|_2^2 \\
= 2 \|Kx_n - y\|_2^2 + 2 \|Kx_m - y\|_2^2.
\end{aligned} \quad (3.45)$$

Luego, al sustituir la ecuación (3.44) en la ecuación (3.45) obtenemos:

$$\begin{aligned}
4 \left\| K \left( \frac{x_n + x_m}{2} \right) - y \right\|_2^2 + \|(Kx_n - y) - (Kx_m - y)\|_2^2 &= 2 \|Kx_n - y\|_2^2 \\
&+ 2 \|Kx_m - y\|_2^2.
\end{aligned}$$

Por lo que,

$$2 \left\| K \left( \frac{x_n + x_m}{2} \right) - y \right\|_2^2 + \frac{1}{2} \|Kx_n - Kx_m\|_2^2 = \|Kx_n - y\|_2^2 + \|Kx_m - y\|_2^2. \quad (3.46)$$

Al sustituir la ecuación (3.46) en la ecuación (3.43) obtenemos lo siguiente:

$$\begin{aligned}
J_\alpha(x_n) + J_\alpha(x_m) &= 2 \left\| K \left( \frac{x_n + x_m}{2} \right) - y \right\|_2^2 + \frac{1}{2} \|K(x_n - x_m)\|_2^2 + \frac{\alpha^2}{2} \|x_n - x_m\|_2^2 \\
&+ 2\alpha^2 \left\| \frac{x_n + x_m}{2} \right\|_2^2
\end{aligned} \quad (3.47)$$

Dado que:

$$2J_\alpha \left( \frac{x_n + x_m}{2} \right) = 2 \left\| K \left( \frac{x_n + x_m}{2} \right) - y \right\|_2^2 + 2\alpha^2 \left\| \frac{x_n + x_m}{2} \right\|_2^2,$$

entonces podemos reescribir la ecuación (3.47) como:

$$J_\alpha(x_n) + J_\alpha(x_m) = 2J_\alpha \left( \frac{x_n + x_m}{2} \right) + \frac{1}{2} \|K(x_n - x_m)\|_2^2 + \frac{\alpha^2}{2} \|x_n - x_m\|_2^2.$$

Puesto que:

$$\lim_{n,m \rightarrow \infty} J_\alpha \left( \frac{x_n + x_m}{2} \right) = I,$$

entonces se cumple:

$$\begin{aligned} J_\alpha(x_n) + J_\alpha(x_m) &= 2J_\alpha \left( \frac{x_n + x_m}{2} \right) + \frac{1}{2} \|K(x_n - x_m)\|_2^2 + \frac{\alpha^2}{2} \|x_n - x_m\|_2^2 \\ &\geq 2I + \frac{\alpha^2}{2} \|x_n - x_m\|_2^2. \end{aligned}$$

Además, se tiene que  $J_\alpha(x_n) + J_\alpha(x_m) \rightarrow 2I$ , cuando  $n, m \rightarrow \infty$ . Por lo tanto,

$$0 \geq \|x_n - x_m\| \geq 0.$$

Esto significa que la sucesión de vectores  $\{x_n\}$  es una sucesión de Cauchy. Luego, por el Teorema 1.13 la sucesión  $\{x_n\}$  es convergente. Sea  $x_\alpha = \lim_{n \rightarrow \infty} x_n$ , con  $x_\alpha \in X$ . Dado que el funcional de Tikhonov es continuo, entonces por la Proposición 1.4 se tiene que  $J_\alpha(x_n) \rightarrow J_\alpha(x_\alpha)$ ; así, por la ecuación (3.39) obtenemos que  $J_\alpha(x_\alpha) = I$ . Esto prueba la existencia de un mínimo de  $J_\alpha$ . Si consideramos a  $x_\alpha$  como un mínimo de  $J_\alpha$ , entonces se debe cumplir que:

$$J_\alpha(x_\alpha) \leq J_\alpha(x), \quad \forall x \in X. \quad (3.48)$$

Además,

$$J_\alpha(x) - J_\alpha(x_\alpha) = (\|Kx - y\|_2^2 - \|Kx_\alpha - y\|_2^2) + \alpha^2 (\|x\|_2^2 - \|x_\alpha\|_2^2),$$

y por la ecuación (3.35) se tiene que:

$$J_\alpha(x) - J_\alpha(x_\alpha) = 2\operatorname{Re}(\langle K^*(Kx_\alpha - y), x - x_\alpha \rangle) + \|K(x - x_\alpha)\|_2^2 + \alpha^2 (\|x\|_2^2 - \|x_\alpha\|_2^2). \quad (3.49)$$

Por el teorema del binomio, se tiene que:

$$\begin{aligned} \|x\|_2^2 &= \|x_\alpha + x - x_\alpha\|_2^2 \\ &= \|x_\alpha\|_2^2 + \|x - x_\alpha\|_2^2 + 2\operatorname{Re}(\langle x_\alpha, x - x_\alpha \rangle); \end{aligned}$$

entonces,

$$\|x\|_2^2 - \|x_\alpha\|_2^2 = \|x - x_\alpha\|_2^2 + 2\operatorname{Re}(\langle x_\alpha, x - x_\alpha \rangle). \quad (3.50)$$

Al sustituir la ecuación (3.50) en la ecuación (3.49) se obtiene:

$$\begin{aligned} J_\alpha(x) - J_\alpha(x_\alpha) &= 2\operatorname{Re}(\langle K^*(Kx_\alpha - y), x - x_\alpha \rangle) + \|K(x - x_\alpha)\|_2^2 + \alpha^2 \|x - x_\alpha\|_2^2 \\ &\quad + 2\alpha^2 \operatorname{Re}(\langle x_\alpha, x - x_\alpha \rangle). \end{aligned}$$

Factorizando términos obtenemos que:

$$\begin{aligned} J_\alpha(x) - J_\alpha(x_\alpha) &= 2\operatorname{Re}(\langle K^*(Kx_\alpha - y), x - x_\alpha \rangle + \alpha^2 \langle x_\alpha, x - x_\alpha \rangle) + \|K(x - x_\alpha)\|_2^2 \\ &\quad + \alpha^2 \|x - x_\alpha\|_2^2. \end{aligned}$$

Por último, aplicando las propiedades 3 y 4 de la Definición 1.16 se tiene que:

$$J_\alpha(x) - J_\alpha(x_\alpha) = 2\operatorname{Re}(\langle K^*(Kx_\alpha - y) + \alpha^2 x_\alpha, x - x_\alpha \rangle) + \|K(x - x_\alpha)\|_2^2 + \alpha^2 \|x - x_\alpha\|_2^2, \quad (3.51)$$

para todo  $x \in X$ . Al sustituir la ecuación (3.51) en la desigualdad (3.48) se obtiene lo siguiente:

$$0 \leq 2\operatorname{Re}(\langle K^*(Kx_\alpha - y) + \alpha^2 x_\alpha, x - x_\alpha \rangle) + \|K(x - x_\alpha)\|_2^2 + \alpha^2 \|x - x_\alpha\|_2^2. \quad (3.52)$$

En particular, la desigualdad (3.52) se cumple para  $x = x_\alpha + \alpha z$ , con  $\alpha > 0$  y  $z = x - x_\alpha \in X$ . Así,

$$\begin{aligned} 0 &\leq 2\operatorname{Re}(\langle K^*(Kx_\alpha - y) + \alpha^2 x_\alpha, \alpha z \rangle) + \|K(\alpha z)\|_2^2 + \alpha^2 \|\alpha z\|_2^2 \\ &= 2\alpha \operatorname{Re}(\langle K^*(Kx_\alpha - y) + \alpha^2 x_\alpha, z \rangle) + \alpha^2 \|Kz\|_2^2 + \alpha^4 \|z\|_2^2. \end{aligned}$$

Al dividir la desigualdad anterior por  $\alpha > 0$  y  $\alpha \rightarrow 0$ , se puede observar que esta

solo se cumple si  $Re(\langle K^*(Kx_\alpha - y) + \alpha^2x_\alpha, z \rangle) \geq 0$  para todo  $z \in X$ ; esto es,

$$K^*(Kx_\alpha - y) + \alpha^2x_\alpha = 0.$$

Por lo tanto,  $x_\alpha$  es solución a la ecuación normal:

$$(\alpha^2I + K^*K)x_\alpha = K^*y.$$

Probemos que dicha solución es única. Para esto, supongamos que  $x_\ell$  también es solución de la ecuación normal, por lo que se debe cumplir que:

$$\alpha^2x_\ell + K^*Kx_\ell = K^*y.$$

Entonces, se tiene que:

$$\alpha^2x_\alpha + K^*Kx_\alpha = \alpha^2x_\ell + K^*Kx_\ell,$$

lo cual solo se cumple si:

$$x_\alpha = x_\ell.$$

Por lo tanto,  $x_\alpha$  es la única solución de la ecuación normal. Por otro lado, probemos que el operador  $A_\alpha = \alpha^2I + K^*K : X \rightarrow X$  es biyectivo. Es fácil ver que  $A_\alpha$  es sobreyectivo, demostremos que  $A_\alpha$  es inyectivo para todo  $\alpha > 0$ .

Sea  $x \in N_{A_\alpha}$ , entonces se tiene que  $A_\alpha x = 0$ . Luego, multiplicando por  $x$  y aplicando la propiedad 3 de la Definición 1.16 se tiene lo siguiente:

$$\begin{aligned} \langle A_\alpha x, x \rangle &= \langle \alpha^2x + K^*Kx, x \rangle \\ &= \langle \alpha^2x, x \rangle + \langle K^*Kx, x \rangle \\ &= \alpha^2\langle x, x \rangle + \langle Kx, Kx \rangle \\ &= \alpha^2\|x\|_2^2 + \|Kx\|_2^2 \\ &= 0 \end{aligned}$$

Así, se debe cumplir que  $x = 0$ . Por lo tanto, por la arbitrariedad de  $x$  se cumple que  $N_{A_\alpha} = \{0\}$ . De aquí, por el Teorema 1.19 se tiene que  $A_\alpha$  es un operador inyectivo. Por lo tanto,  $A_\alpha$  es biyectivo, es decir, es invertible.

Los cálculos anteriores demuestran el siguiente teorema.

**Teorema 3.2.** *Sea  $K : X \rightarrow Y$  un operador lineal y acotado sobre espacios de Hilbert y  $\alpha > 0$ . Entonces el funcional de Tikhonov  $J_\alpha$  tiene un único mínimo  $x_\alpha \in X$ . Este mínimo  $x_\alpha$  es la única solución de la ecuación normal:*

$$\alpha^2 x_\alpha + K^* K x_\alpha = K^* y. \quad (3.53)$$

En particular, si la matriz  $A$  de  $m \times n$  es la representación matricial del operador lineal y acotado  $K : \mathbb{K}^n \rightarrow \mathbb{K}^m$  sobre las bases canónicas de  $\mathbb{K}^n$  y  $\mathbb{K}^m$ , tal que  $K(x) = Ax$ , entonces tendríamos que las ecuaciones normales regularizadas se pueden reescribir como:

$$\alpha^2 x_\alpha + A^t A x_\alpha = A^t y, \quad \text{con } \alpha > 0, \quad (3.54)$$

donde  $A^* = A^t$ , si  $A \in M_{m \times n}(\mathbb{R})$ .

Si consideramos la DVS de la matriz  $A \in M_{m \times n}(\mathbb{R})$  dada por:

$$A = \underbrace{\begin{bmatrix} u_1 & \dots & u_m \end{bmatrix}}_{U_{m \times m}} \underbrace{\begin{bmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n \\ 0 & 0 & 0 \end{bmatrix}}_{\Sigma_{m \times n}} \underbrace{\begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}^t}_{V_{n \times n}^t},$$

tendremos, por la propiedad 2 del Teorema 1.1 que:

$$\begin{aligned} A^t &= (U \Sigma V^t)^t \\ &= (\Sigma V^t)^t U^t \\ &= (V^t)^t \Sigma^t U^t \\ &= V \Sigma^t U^t. \end{aligned} \quad (3.55)$$

Luego, se tiene que:

$$A^t A = (V \Sigma^t U^t) (U \Sigma V^t). \quad (3.56)$$

Dado que las matrices  $U$  y  $V$  son ortogonales se cumple que:

$$U^t U = I_m \quad \text{y} \quad V V^t = I_n;$$

así, tendremos que:

$$A^t A = V \Sigma^t \Sigma V^t.$$

Luego, se tiene lo siguiente:

$$\alpha^2 I_n + A^t A = \alpha^2 V V^t + V \Sigma^t \Sigma V^t = V (\alpha^2 I_n + \Sigma^t \Sigma) V^t. \quad (3.57)$$

Observemos que  $\Sigma^t \Sigma$  es una matriz de  $n \times n$ , dada por:

$$\Sigma^t \Sigma = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n^2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Así, se tiene que:

$$\alpha^2 I_n + \Sigma^t \Sigma = \begin{bmatrix} \alpha^2 + \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \alpha^2 + \sigma_n^2 \\ 0 & 0 & 0 \end{bmatrix} = \text{diag} (\alpha^2 + \sigma_1^2, \dots, \alpha^2 + \sigma_n^2).$$

Al sustituir esta última igualdad en la ecuación (3.57) se obtiene:

$$\alpha^2 I_n + A^t A = V \text{diag} (\alpha^2 + \sigma_1^2, \dots, \alpha^2 + \sigma_n^2) V^t.$$

Observemos que:

$$\begin{aligned} (\alpha^2 I_n + A^t A) V &= V \text{diag} (\alpha^2 + \sigma_1^2, \dots, \alpha^2 + \sigma_n^2) V^t V \\ &= V \text{diag} (\alpha^2 + \sigma_1^2, \dots, \alpha^2 + \sigma_n^2). \end{aligned}$$

Esto nos quiere decir que la matriz diagonal contiene los valores propios de la matriz  $\alpha^2 I_n + A^t A$  y los vectores singulares por la derecha de  $A$  son sus respectivos vectores propios. Como  $\alpha > 0$ , entonces los valores propios de  $\alpha^2 I_n + A^t A$  son positivos, es decir, la matriz  $\alpha^2 I_n + A^t A$  es definida positiva. Dado que toda matriz definida positiva es invertible, entonces se tiene que  $\alpha^2 I_n + A^t A$  es invertible. Dicha

matriz se define como:

$$(\alpha^2 I_n + A^t A)^{-1} = V \operatorname{diag} \left( \frac{1}{\alpha^2 + \sigma_1^2}, \dots, \frac{1}{\alpha^2 + \sigma_n^2} \right) V^t. \quad (3.58)$$

De la ecuación (3.54) se obtiene que:

$$x_\alpha = (\alpha^2 I_n + A^t A)^{-1} A^t y. \quad (3.59)$$

Al sustituir las ecuaciones (3.58) y (3.55) en la ecuación (3.59) se tiene que la solución regularizada  $x_\alpha$ , está dada por:

$$x_\alpha = V \operatorname{diag} \left( \frac{1}{\alpha^2 + \sigma_1^2}, \dots, \frac{1}{\alpha^2 + \sigma_n^2} \right) V^t V \Sigma^t U^t y.$$

Dado que:

$$V^t V = I_n \quad y \quad \Sigma^t U^t y = \begin{bmatrix} \sigma_1 u_1^t y \\ \vdots \\ \sigma_n u_n^t y \end{bmatrix};$$

entonces se cumple:

$$x_\alpha = \begin{pmatrix} v_1 & \dots & v_n \end{pmatrix} \operatorname{diag} \left( \frac{1}{\alpha^2 + \sigma_1^2}, \dots, \frac{1}{\alpha^2 + \sigma_n^2} \right) \begin{bmatrix} \sigma_1 u_1^t y \\ \vdots \\ \sigma_n u_n^t y \end{bmatrix}.$$

Por lo tanto, la solución de la ecuaciones normales regularizadas (3.54) es:

$$x_\alpha = \sum_{i=1}^n \left[ \frac{\sigma_i}{\alpha^2 + \sigma_i^2} u_i^t y \right] v_i; \quad (3.60)$$

donde los vectores  $u_i$  y  $v_i$  son sistemas de vectores ortonormales dados por el Teorema 1.17. La solución regularizada (3.60) minimiza el funcional de Tikhonov, por lo que es posible aplicar dicha solución regularizada para resolver el problema inverso discreto mal condicionado de la PSD. En el siguiente capítulo, se resuelve una aplicación sintética del problema inverso de PSD para partículas de látex NBR, con un conjunto finito de longitudes de onda y de diámetros de partículas.

# Capítulo 4

## Aplicación del problema inverso

La distribución de tamaños de partículas (PSD) de un látex industrial es una característica morfológica importante que determina la procesabilidad y propiedades finales del material cuando estos son usados en adhesivos, revestimientos o pintura ([24]). La microscopía electrónica de barrido (SEM) y la microscopía electrónica de transmisión (TEM) son actualmente las técnicas más confiables para estimar y observar una PSD; sin embargo, ambos métodos son experimentalmente costosos, difíciles y lentos ([24]). Pocos han intentado determinar la PSD por turbidimetría, esto se debe a las dificultades numéricas involucradas con el tratamiento de los datos y a la falta de información precisa sobre las propiedades ópticas requeridas. Sin embargo, en esta aplicación resolveremos el problema inverso de estimación de PSD por turbidimetría UV-Vis cerca de la región de Rayleigh, basados en la teoría según [24], para látex industriales blandos<sup>1</sup> (NBR<sup>2</sup>) con tamaños de partículas en el rango de 80 a 85 *nm*. En este capítulo se presenta un ejemplo sintético que se resuelve mediante la inversa generalizada y el método de regularización de Tikhonov; el ejemplo sintético corresponde al caso de polímeros blandos y en el que se usan los coeficientes de Mie. Para poder simular las mediciones de turbidez es necesario considerar un modelo matemático que relacione el espectro de turbidez (o extinción de luz)  $\tau$  con el tamaño de la partícula;

---

<sup>1</sup>Los materiales poliméricos (los plásticos) se pueden agrupar, de acuerdo a su estructura mecánica, en dos tipos: rígidos y blandos. Los blandos tienen canales internos y huecos que atrapan físicamente las moléculas de agua, es decir, los polímeros blandos son como esponjas donde parte de agua se adhiere a su superficie y parte queda atrapada en los huecos.

<sup>2</sup>Su abreviatura internacional es NBR (Nitrile Butadiene Rubber) y es un caucho que se utiliza para la fabricación de productos de caucho industriales y automotrices, por ejemplo, el látex sintético. Algunas de sus aplicaciones se dan en las juntas de estanqueidad de aceites y gasolinas en tuberías, racores, pletinas, tanques y depósitos.

además, para resolver el problema inverso de estimar la distribución de tamaños de partícula, necesitamos datos reales; sin embargo, debido a que no contamos con ellos vamos a construir un ejemplo sintético.

## 4.1. Solución del problema directo

Como se ha mencionado anteriormente, para resolver el problema inverso, necesitamos resolver el problema directo, debido a que no contamos con datos reales. Es decir, vamos a suponer conocida la matriz  $A$ , los valores de  $f(D_j)$  y calcularemos los valores de  $\tau(\lambda_i)$ . Una vez resuelto el problema directo, vamos a resolver el problema inverso de hallar a  $f(D_j)$ , a partir del conocimiento de la matriz  $A$ , y los datos  $\tau(\lambda_i)$ . La solución del problema inverso debe ser muy parecida a la forma de  $f(D_j)$ .

Para resolver el problema directo, vamos a suponer conocida la función continua  $f(D)$  que, de acuerdo con [24], está dada por:

$$f(D) = \frac{N_p}{D\sigma\sqrt{2\pi}} \exp \left\{ \frac{-[\ln(D) - \ln(D_g)]^2}{2\sigma^2} \right\}; \quad (4.1)$$

con  $N_p = 1.2 \times 10^{12} \text{ part./cm}^3$ ;  $D_g = 80 \text{ nm}$ ; y  $\sigma = 0.1$ . La distribución continua de la ecuación (4.1) se puede cambiar por la función discreta equivalente  $f(D_j)$ , tomando en cuenta los diámetros  $D_j$  y las longitudes de onda  $\lambda_i$ . Para este ejemplo, se consideran cuatro diferentes diámetros promedios de partículas de látex industriales blandos (NBR) inmersas en agua y tres diferentes longitudes de onda, dados por:

$$D_1 = 80.4 \text{ nm},$$

$$D_2 = 82.8 \text{ nm},$$

$$D_3 = 83.6 \text{ nm},$$

$$D_4 = 84.5 \text{ nm},$$

y

$$\lambda_1 = 300 \text{ nm},$$

$$\lambda_2 = 450 \text{ nm},$$

$$\lambda_3 = 600 \text{ nm}.$$

Así, el problema inverso que resolveremos es discreto, y está dado por:

$$\tau(\lambda_i) = \frac{\log_{10}(e)\pi\mathcal{L}}{4} \sum_{j=1}^n f(D_j)Q_{ext}(x_{ij}, m_i)D_j^2, \quad (4.2)$$

donde,

$$x_{ij} = \pi D_j \left( \frac{n_m(\lambda_i)}{\lambda_i} \right), \quad m_i = \frac{n_p(\lambda_i)}{n_m(\lambda_i)}, \quad \mathcal{L} = 1 \text{ cm},$$

y

$$Q_{ext}(x_{ij}, m_i) = \frac{2}{x_{ij}^2} 3a_1 = \frac{8}{3} x_{ij}^4 \left( \frac{m_i^2 - 1}{m_i^2 + 2} \right)^2.$$

Asumiremos que las funciones de índices de refracción del medio  $n_m(\lambda_i)$  y de las partículas  $n_p(\lambda_i)$ , se definen como ([24]):

$$n_m(\lambda_i) = 1.324 + \frac{3046}{\lambda_i^2}, \quad (4.3)$$

$$n_p(\lambda_i) = 1.494 + \frac{6284}{\lambda_i^2}. \quad (4.4)$$

Usando una notación vectorial es posible reescribir la ecuación (4.2) como:

$$\tau = Af, \quad (4.5)$$

donde  $A \in M_{3 \times 4}(\mathbb{R})$ ,  $\tau \in \mathbb{R}^3$ , y  $f \in \mathbb{R}^4$ . Así, el problema directo consiste en hallar el vector de espectro de turbidez  $\tau$  a partir del conocimiento de la matriz  $A$  y el vector  $f$ . Para esto es necesario hallar los coeficientes de la matriz  $A$ , los cuales están dados por ([24]):

$$a_{ij} = \frac{\log_{10}(e)\pi\mathcal{L}}{4} Q_{ext}(x_{ij}, m_i)D_j^2, \quad (4.6)$$

con  $(i = 1, \dots, m)$  y  $(j = 1, \dots, n)$ .

Para ser congruentes con los artículos [24, 21] de donde se obtiene toda la información de este ejemplo, consideramos que las unidades de medida para las longitudes de onda  $\lambda_i$  y los diámetros  $D_j$  están dados en nanómetros; y el camino óptico  $\mathcal{L}$  y la densidad numérica  $N_p$  están dados en centímetros. Esto se hace por conveniencia, pues queremos hallar una aproximación a la inversa generalizada de Penrose; además, todos los cálculos se realizan con ayuda del sistema de cómputo numérico MATLAB.

Comenzamos por calcular los valores de  $f(D_j)$  :

$$f(D_j) = \frac{1.2 \times 10^{12} \text{ #part./cm}^3}{D_j(0.1)\sqrt{2\pi}} \exp \left\{ \frac{-[\ln(D_j) - \ln(80)]^2}{0.02} \right\};$$

$$f(D_1) = 5.9470 \times 10^{10} \frac{\text{#part.}}{\text{cm}^3},$$

$$f(D_2) = 5.4496 \times 10^{10} \frac{\text{#part.}}{\text{cm}^3},$$

$$f(D_3) = 5.1977 \times 10^{10} \frac{\text{#part.}}{\text{cm}^3},$$

$$f(D_4) = 4.8776 \times 10^{10} \frac{\text{#part.}}{\text{cm}^3}.$$

Ahora, hallemos los coeficientes de la matriz  $A$ . Para esto, primero calculamos los índices de refracción de las partículas,  $n_p(\lambda_i)$ , los índices de refracción del medio  $n_m(\lambda_i)$ , y los índices de refracción relativos  $m_i$ , considerando las longitudes de onda  $\lambda_i$  :

$$n_p(\lambda_1) = 1.494 + \frac{6284}{(300)^2} = 1.5638,$$

$$n_p(\lambda_2) = 1.494 + \frac{6284}{(450)^2} = 1.5250,$$

$$n_p(\lambda_3) = 1.494 + \frac{6284}{(600)^2} = 1.5115;$$

$$n_m(\lambda_1) = 1.324 + \frac{3046}{(300)^2} = 1.3578,$$

$$n_m(\lambda_2) = 1.324 + \frac{3046}{(450)^2} = 1.3390,$$

$$n_m(\lambda_3) = 1.324 + \frac{3046}{(600)^2} = 1.3325;$$

$$m_1 = \frac{n_p(\lambda_1)}{n_m(\lambda_1)} = \frac{1.564}{1.358} = 1.1517,$$

$$m_2 = \frac{n_p(\lambda_2)}{n_m(\lambda_2)} = \frac{1.525}{1.339} = 1.1389,$$

$$m_3 = \frac{n_p(\lambda_3)}{n_m(\lambda_3)} = \frac{1.511}{1.332} = 1.1343.$$

Con los datos anteriores es posible calcular los parámetros de tamaño  $x_{ij}$  :

$$x_{11} = \pi D_1 \left( \frac{n_m(\lambda_1)}{\lambda_1} \right) = 1.1432,$$

$$x_{12} = \pi D_2 \left( \frac{n_m(\lambda_1)}{\lambda_1} \right) = 1.1774,$$

$$x_{13} = \pi D_3 \left( \frac{n_m(\lambda_1)}{\lambda_1} \right) = 1.1887,$$

$$x_{14} = \pi D_4 \left( \frac{n_m(\lambda_1)}{\lambda_1} \right) = 1.2015,$$

$$x_{21} = \pi D_1 \left( \frac{n_m(\lambda_2)}{\lambda_2} \right) = 0.7516,$$

$$x_{22} = \pi D_2 \left( \frac{n_m(\lambda_2)}{\lambda_2} \right) = 0.7740,$$

$$x_{23} = \pi D_3 \left( \frac{n_m(\lambda_2)}{\lambda_2} \right) = 0.7815,$$

$$x_{24} = \pi D_4 \left( \frac{n_m(\lambda_2)}{\lambda_2} \right) = 0.7899,$$

$$x_{31} = \pi D_1 \left( \frac{n_m(\lambda_3)}{\lambda_3} \right) = 0.5609,$$

$$x_{32} = \pi D_2 \left( \frac{n_m(\lambda_3)}{\lambda_3} \right) = 0.5777,$$

$$x_{33} = \pi D_3 \left( \frac{n_m(\lambda_3)}{\lambda_3} \right) = 0.5833,$$

$$x_{34} = \pi D_4 \left( \frac{n_m(\lambda_3)}{\lambda_3} \right) = 0.5895.$$

Al sustituir los valores encontrados  $x_{ij}$  y  $m_i$  en la función de extinción  $Q_{ext}(x_{ij}, m_i)$  se obtiene:

$$Q_{ext}(x_{11}, m_1) = \frac{8}{3} x_{11}^4 \left( \frac{m_1^2 - 1}{m_1^2 + 2} \right)^2 = 0.0439,$$

$$Q_{ext}(x_{12}, m_1) = \frac{8}{3} x_{12}^4 \left( \frac{m_1^2 - 1}{m_1^2 + 2} \right)^2 = 0.0493,$$

$$Q_{ext}(x_{13}, m_1) = \frac{8}{3} x_{13}^4 \left( \frac{m_1^2 - 1}{m_1^2 + 2} \right)^2 = 0.0513,$$

$$Q_{ext}(x_{14}, m_1) = \frac{8}{3} x_{14}^4 \left( \frac{m_1^2 - 1}{m_1^2 + 2} \right)^2 = 0.0535,$$

$$Q_{ext}(x_{21}, m_2) = \frac{8}{3} x_{21}^4 \left( \frac{m_2^2 - 1}{m_2^2 + 2} \right)^2 = 0.0069,$$

$$Q_{ext}(x_{22}, m_2) = \frac{8}{3}x_{22}^4 \left( \frac{m_2^2 - 1}{m_2^2 + 2} \right)^2 = 0.0078,$$

$$Q_{ext}(x_{23}, m_2) = \frac{8}{3}x_{23}^4 \left( \frac{m_2^2 - 1}{m_2^2 + 2} \right)^2 = 0.0081,$$

$$Q_{ext}(x_{24}, m_2) = \frac{8}{3}x_{24}^4 \left( \frac{m_2^2 - 1}{m_2^2 + 2} \right)^2 = 0.0084,$$

$$Q_{ext}(x_{31}, m_3) = \frac{8}{3}x_{31}^4 \left( \frac{m_3^2 - 1}{m_3^2 + 2} \right)^2 = 0.0020,$$

$$Q_{ext}(x_{32}, m_3) = \frac{8}{3}x_{32}^4 \left( \frac{m_3^2 - 1}{m_3^2 + 2} \right)^2 = 0.0023,$$

$$Q_{ext}(x_{33}, m_3) = \frac{8}{3}x_{33}^4 \left( \frac{m_3^2 - 1}{m_3^2 + 2} \right)^2 = 0.0023,$$

$$Q_{ext}(x_{34}, m_3) = \frac{8}{3}x_{34}^4 \left( \frac{m_3^2 - 1}{m_3^2 + 2} \right)^2 = 0.0025.$$

Por último, usando la ecuación (4.6) hallamos los coeficientes  $a_{ij}$  de la matriz  $A$  como:

$$a_{11} = \frac{\log_{10}(e)\pi\mathcal{L}}{4}Q_{ext}(x_{11}, m_1)D_1^2 = 96.7045,$$

$$a_{12} = \frac{\log_{10}(e)\pi\mathcal{L}}{4}Q_{ext}(x_{12}, m_1)D_2^2 = 115.3699,$$

$$a_{13} = \frac{\log_{10}(e)\pi\mathcal{L}}{4}Q_{ext}(x_{13}, m_1)D_3^2 = 122.2216,$$

$$a_{14} = \frac{\log_{10}(e)\pi\mathcal{L}}{4}Q_{ext}(x_{14}, m_1)D_4^2 = 130.3319,$$

$$a_{21} = \frac{\log_{10}(e)\pi\mathcal{L}}{4}Q_{ext}(x_{21}, m_2)D_1^2 = 15.2341,$$

$$a_{22} = \frac{\log_{10}(e)\pi\mathcal{L}}{4}Q_{ext}(x_{22}, m_2)D_2^2 = 18.1745,$$

$$a_{23} = \frac{\log_{10}(e)\pi\mathcal{L}}{4}Q_{ext}(x_{23}, m_2)D_3^2 = 19.2538,$$

$$a_{24} = \frac{\log_{10}(e)\pi\mathcal{L}}{4}Q_{ext}(x_{24}, m_2)D_4^2 = 20.5315,$$

$$a_{31} = \frac{\log_{10}(e)\pi\mathcal{L}}{4}Q_{ext}(x_{31}, m_3)D_1^2 = 4.4296,$$

$$a_{32} = \frac{\log_{10}(e)\pi\mathcal{L}}{4}Q_{ext}(x_{32}, m_3)D_2^2 = 5.2846,$$

$$a_{33} = \frac{\log_{10}(e)\pi\mathcal{L}}{4}Q_{ext}(x_{33}, m_3)D_3^2 = 5.5984,$$

$$a_{34} = \frac{\log_{10}(e)\pi\mathcal{L}}{4}Q_{ext}(x_{34}, m_3)D_4^2 = 5.9699.$$

Luego, es posible escribir la matriz  $A$  como:

$$A = \begin{bmatrix} 96.7045 & 115.3699 & 122.2216 & 130.3319 \\ 15.2341 & 18.1745 & 19.2538 & 20.5315 \\ 4.4296 & 5.2846 & 5.5984 & 5.9699 \end{bmatrix} \quad (4.7)$$

Con los datos encontrados hasta ahora se puede resolver el problema directo de estimación de distribución de tamaños de partícula. Por lo tanto, la solución al problema directo de estimación de distribución de tamaños de partículas está dada por:

$$\begin{aligned} \tau &= Af \\ &= \begin{bmatrix} 96.7045 & 115.3699 & 122.2216 & 130.3319 \\ 15.2341 & 18.1745 & 19.2538 & 20.5315 \\ 4.4296 & 5.2846 & 5.5984 & 5.9699 \end{bmatrix} \begin{bmatrix} 5.9470 \times 10^{10}, \\ 5.4496 \times 10^{10}, \\ 5.1977 \times 10^{10}, \\ 4.8776 \times 10^{10} \end{bmatrix} \\ &= \begin{bmatrix} 2.4748 \times 10^{13} \\ 0.3899 \times 10^{13} \\ 0.1134 \times 10^{13} \end{bmatrix} \end{aligned}$$

Se puede observar que la solución al problema directo se realiza de forma muy sencilla y a partir de esta es posible hacerse una idea de la solución del problema inverso. En la siguiente sección se resuelve el problema inverso de estimación de distribución de tamaños de partícula usando dos métodos importantes, la inversa generalizada y el método de regularización de Tikhonov.

## 4.2. Solución del problema inverso

Ahora, vamos a resolver el problema inverso de hallar  $f(D_j)$  a partir del conocimiento del espectro de turbidez  $\tau$  y la matriz  $A$ . Supondremos que  $\tau \in \mathbb{R}^3$  y la matriz  $A \in M_{3 \times 4}(\mathbb{R})$  son datos conocidos, además, como en la práctica siempre existen errores en las mediciones de turbidez, consideramos un error relativo  $\epsilon$ , por lo que el problema (4.2) se puede escribir como:

$$\tau = Af + \epsilon. \quad (4.8)$$

Consideramos dos métodos de solución para hallar el vector  $f \in \mathbb{R}^4$ , el primero la inversa generalizada de Moore-Penrose y el segundo la regularización de Tikhonov.

### 4.2.1. Solución obtenida por la inversa generalizada

Según la Definición 3.4 y la Observación 3.2 es posible hallar la matriz pseudoinversa de Penrose de la matriz  $A$  por medio de la ecuación (3.7). En este caso,  $m = 3$  y  $n = 4$ , es decir,  $m < n$ ; además, es fácil ver que los renglones  $r_1, r_2, r_3 \in \mathbb{R}^4$ , dados por:

$$r_1 = (96.7045, 115.3699, 122.2216, 130.3319),$$

$$r_2 = (15.2341, 18.1745, 19.2538, 20.5315),$$

$$r_3 = (4.4296, 5.2846, 5.5984, 5.9699),$$

son linealmente independientes ya que no existe algún par de números  $\alpha$  y  $\beta$  tales que:

$$r_1 = \alpha r_2 + \beta r_3.$$

Por lo tanto, la matriz  $A$  es de rango completo por renglones, i.e.,  $\dim R_A = 3$ ; y por el Teorema 1.5 se cumple que:

$$\dim R_A = \text{ran}(A) = 3.$$

Puesto que  $A \in M_{3 \times 4}(\mathbb{R})$ , entonces  $A^* = A^t$ . Luego, de la Observación 3.2 inciso (3), se tiene que la matriz pseudoinversa de Penrose de la matriz  $A$  se obtiene por:

$$Y = A^t [AA^t]^{-1}.$$

Haciendo los cálculos para la matriz  $A$  se tiene lo siguiente:

$$A^t = \begin{bmatrix} 96.7045 & 15.2341 & 4.4296 \\ 115.3699 & 18.1745 & 5.2846 \\ 122.2216 & 19.2538 & 5.5984 \\ 130.3319 & 20.5315 & 5.9699 \end{bmatrix};$$

así,

$$AA^t = \begin{bmatrix} 5.4586 & 0.8599 & 0.2500 \\ 0.8599 & 0.1355 & 0.0394 \\ 0.2500 & 0.0394 & 0.0115 \end{bmatrix} \times 10^4.$$

Luego, se tiene que la matriz inversa de  $AA^t$  está dada por:

$$[AA^t]^{-1} = \begin{bmatrix} 0.0078 & -0.0524 & 0.0108 \\ -0.0524 & 0.4473 & -0.3952 \\ 0.0108 & -0.3952 & 1.1238 \end{bmatrix} \times 10^9.$$

Por lo tanto, la matriz pseudoinversa de Penrose de la matriz  $A$  es:

$$Y = A^t [AA^t]^{-1} = \begin{bmatrix} -0.0925 & 0.6226 & -0.1210 \\ -0.0514 & -0.4166 & 2.5555 \\ 0.2348 & -1.4884 & -0.0064 \\ -0.1060 & 1.3026 & -2.1663 \end{bmatrix} \times 10^4.$$

Podemos verificar que se cumplen aproximadamente las condiciones de Penrose:

1.  $AYA = A$  :

$$AYA = \begin{bmatrix} 96.7066 & 115.3725 & 122.2243 & 130.3348 \\ 15.2344 & 18.1749 & 19.2542 & 20.5320 \\ 4.4297 & 5.2847 & 5.5985 & 5.9700 \end{bmatrix} \approx A.$$

2.  $YAY = Y$  :

$$YAY = \begin{bmatrix} -0.0925 & 0.6226 & -0.1209 \\ -0.0514 & -0.4166 & 2.5555 \\ 0.2348 & -1.4884 & -0.0065 \\ -0.1060 & 1.3026 & -2.1663 \end{bmatrix} \times 10^4 \approx Y.$$

3.  $AY = (AY)^t$  :

$$AY = \begin{bmatrix} 1.0000 & 0.0004 & 0.0001 \\ 0.0000 & 1.0000 & 0.0001 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix} \approx I_{3 \times 3} = (AY)^t.$$

4.  $YA = (YA)^t$  :

$$YA = \begin{bmatrix} 0.2817 & 0.2676 & -0.0638 & 0.3559 \\ 0.2676 & 0.9003 & 0.0238 & -0.1326 \\ -0.0638 & 0.0238 & 0.9943 & 0.0316 \\ 0.3560 & -0.1326 & 0.0316 & 0.8236 \end{bmatrix} \approx (YA)^t.$$

Con esto, hemos encontrado la inversa generalizada de la matriz  $A$ , solo nos falta calcular el valor del vector de espectro de turbidez  $\tau$  para hallar una posible solución al problema inverso de estimación de PSD.

**Nota 4.1.** Para hacer más evidentes los resultados de los siguientes cálculos se hacen usando el formato *longEng* de MatLab.

Calculamos el vector de espectro de turbidez  $\tau(\lambda_i)$  dado por la ecuación (2.35) para cada longitud de onda  $\lambda_i$  :

$$\begin{aligned}\tau(\lambda_1) &= \frac{\log_{10}(e)\pi\mathcal{L}}{4} \sum_{j=1}^4 f(D_j)Q_{ext}(x_{1j}, m_i)D_j^2 = 24.7479046889131 \times 10^{12}, \\ \tau(\lambda_2) &= \frac{\log_{10}(e)\pi\mathcal{L}}{4} \sum_{j=1}^4 f(D_j)Q_{ext}(x_{2j}, m_i)D_j^2 = 3.89858946984298 \times 10^{12}, \\ \tau(\lambda_3) &= \frac{\log_{10}(e)\pi\mathcal{L}}{4} \sum_{j=1}^4 f(D_j)Q_{ext}(x_{3j}, m_i)D_j^2 = 1.13358678358970 \times 10^{12}.\end{aligned}$$

De esta manera tenemos que el problema inverso de estimación de distribución de tamaños de partículas está dado por:

$$\begin{aligned}\tau(\lambda_i) &= Af + \epsilon \\ \begin{bmatrix} 24.7479046889131 \times 10^{12} \\ 3.89858946984298 \times 10^{12} \\ 1.13358678358970 \times 10^{12} \end{bmatrix} &= \begin{bmatrix} 96.7045 & 115.3699 & 122.2216 & 130.3319 \\ 15.2341 & 18.1745 & 19.2538 & 20.5315 \\ 4.4296 & 5.2846 & 5.5984 & 5.9699 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} + \epsilon.\end{aligned}\tag{4.9}$$

Debido a que en la práctica generalmente existen errores en las mediciones, es necesario determinar el efecto de los errores en las mediciones, para esto, cambiamos el vector  $\tau(\lambda_i)$  sumándole un valor pequeño  $\epsilon = 1 \times 10^{-4}$ , el cual representa el error en las mediciones (solo para hacer los cálculos de los errores más evidentes consideramos el vector  $\tau$  en centímetros):

$$\begin{aligned}\hat{\tau}(\lambda_i) &= \tau(\lambda_i) + \epsilon \\ &= \begin{bmatrix} 0.247479046889131 \\ 0.0389858946984298 \\ 0.0113358678358970 \end{bmatrix} + \epsilon \\ &= \begin{bmatrix} 0.247579046889131 \\ 0.0390858946984298 \\ 0.0114358678358970 \end{bmatrix}.\end{aligned}$$

Luego, el error absoluto en las mediciones es:

$$\epsilon = \|\tau(\lambda_i) - \hat{\tau}(\lambda_i)\|_2 = 1 \times 10^{-4}.$$

Supongamos que conocemos a  $\tau$  y a la matriz  $A$ , entonces hallemos una solución al problema de la ecuación (4.9). La primera idea que nos surge para resolver el problema es hallar la matriz inversa de la matriz  $A$ , lo cual resulta complicado usando la teoría de matrices inversas, pues solo consideran el caso de matrices cuadradas. Para esto es necesario generalizar la teoría, y usar la matriz inversa generalizada de Moore-Penrose. De esta manera, la primera posibilidad que se plantea para la solución de un problema inverso es construir un operador lineal que mande los datos  $\tau$  a los parámetros estimados  $\hat{f}$ . El operador lineal más general es la matriz inversa generalizada, en este caso, el operador lineal está dado por la matriz de Penrose  $Y$ . Así, una posible solución al problema inverso, según la ecuación (3.10) es:

$$\hat{f} = Y\tau. \quad (4.10)$$

La solución dada por la ecuación (4.10) es posible si consideramos que la matriz  $A$  es bien condicionada, de lo contrario, la estimación  $\hat{f}$  será muy diferente al vector de parámetros verdaderos  $f$ . Expongamos el por qué la matriz  $A$  es mal condicionada, primero consideremos el vector de mediciones libre de errores  $\tau$  y calculamos el valor estimado del vector de PSD  $\hat{f}$ :

$$\hat{f} = Y\tau = \begin{bmatrix} 0.000182477455386 \\ 0.000084237889068 \\ 0.001238859978678 \\ 0.000527147719822 \end{bmatrix}. \quad (4.11)$$

Ahora calculamos la estimación del vector de parámetros  $\hat{f}$  considerando el vector con ruido en las mediciones  $\hat{\tau}$ :

$$\hat{f} = Y\hat{\tau} = \begin{bmatrix} 0.409293417919107 \\ 2.087583172718300 \\ -1.258845877485565 \\ -0.969210418586275 \end{bmatrix}. \quad (4.12)$$

Es posible observar la diferencia entre los vectores de parámetros de PSD  $\hat{f}$  y  $\hat{\hat{f}}$ , sin embargo calculamos los errores en los datos y en los parámetros como:

$$\begin{aligned}\|\hat{f} - \hat{\hat{f}}\|_2 &= 2.655791418318820 \\ \|\tau - \hat{\tau}\|_2 &= 1 \times 10^{-4}\end{aligned}$$

Se observa que el error en los datos es más pequeño que el error en la solución, es decir, pequeños cambios en el vector de mediciones  $\tau$  provoca errores grandes en la solución estimada  $\hat{f}$ , esto quiere decir que el problema inverso es mal condicionado. Esto se debe a que no se cumple la condición de estabilidad, pues la solución no depende continuamente de los datos.

Consideremos la matriz  $A \in M_{3 \times 4}(\mathbb{R})$  como la representación matricial del operador lineal  $K : \mathbb{R}^4 \rightarrow \mathbb{R}^3$  y la matriz transpuesta  $A^t \in M_{4 \times 3}(\mathbb{R})$  como la representación matricial del operador lineal adjunto  $K^* : \mathbb{R}^3 \rightarrow \mathbb{R}^4$ , sobre las bases canónicas de  $\mathbb{R}^3$  y  $\mathbb{R}^4$ . Sabemos que una manifestación de inestabilidad en el sistema (4.9) puede observarse a través del cálculo del número de condición de la matriz del sistema dado por la ecuación (1.3). Para hallar  $c(A)$ , es necesario calcular los valores singulares  $\sigma_k$  ( $k = 1, 2, 3, 4$ ) de la matriz  $A$ . Para esto, primero encontremos los valores propios de la matriz  $A^t A$  dada por:

$$A^t A = \begin{bmatrix} 9603.45947922000 & 11457.0693091600 & 12137.4917044200 & 12942.9044167400 \\ 11457.0693091600 & 13668.4532734200 & 14480.2072625800 & 15441.0765501000 \\ 12137.4917044200 & 14480.2072625800 & 15340.1704035600 & 16358.1046319000 \\ 12942.9044167400 & 15441.0765501000 & 16358.1046319000 & 17443.5863558700 \end{bmatrix}.$$

Para hallar los valores propios de la matriz  $A^t A \in M_{4 \times 4}(\mathbb{R})$  usamos la función  $eig(A^t A)$  en MatLab, la cual nos devuelve el vector  $\beta \in \mathbb{R}^4$  que contiene dichos valores propios:

$$\beta = \begin{bmatrix} 56.0556695120656 \times 10^3 \\ 3.66837970994902 \times 10^{-9} \\ 763.171579628470 \times 10^{-12} \\ 0 \end{bmatrix}. \quad (4.13)$$

Usando la función  $dvs(A)$  de MatLab obtenemos el vector  $\sigma \in \mathbb{R}^4$  que contiene los valores singulares de la matriz  $A \in M_{3 \times 4}(\mathbb{R})$  ordenados de forma decreciente:

$$\sigma = \begin{bmatrix} 236.760785418670 \\ 60.5937041072596 \times 10^{-6} \\ 27.6654149258401 \times 10^{-6} \\ 0 \end{bmatrix}. \quad (4.14)$$

Si consideramos los valores singulares no nulos, se tiene que:

$$\begin{aligned} \max_k |\sigma_k| &= 236.760785418670, \\ \min_k |\sigma_k| &= 27.6654149258401 \times 10^{-6}. \end{aligned}$$

Por lo tanto, el número de condición de la matriz  $A$  es:

$$c(A) = \frac{\max_k |\sigma_k|}{\min_k |\sigma_k|} = \frac{236.760785418670}{27.6654149258401 \times 10^{-6}} = 8.55800594544962 \times 10^6. \quad (4.15)$$

Dado que el número de condición de la matriz  $A$  es muy grande, entonces se tiene que el problema inverso de PSD es mal condicionado. Por lo que no basta con calcular la inversa generalizada de  $A$ ; por esta razón es necesario regularizar el problema, es decir, en lugar de resolver el problema mal condicionado se resolverá un nuevo problema bien condicionado de modo que la solución del nuevo problema sea una buena aproximación de la solución buscada.

#### 4.2.2. Solución obtenida por la regularización de Tikhonov

En este caso, vamos a aplicar el método de regularización de Tikhonov, para esto es necesario elegir una colección de problemas bien planteados que dependen del parámetro de regularización. Para cada valor del parámetro, se obtiene un problema distinto, y en consecuencia una solución regularizada. En este caso, las soluciones de los problemas bien planteados están dadas en términos del parámetro de regularización, pues el problema de hallar el parámetro de regularización que mejor ajuste la solución regularizada a la solución del problema mal condicionado queda abierto, porque escapa a los alcances de esta tesis.

Supongamos que los datos dados por el vector de espectro de turbidez  $\tau \in \mathbb{R}^3$  y

la matriz  $A$  son conocidos, y se quiere hallar  $f_\alpha \in \mathbb{R}^4$  que minimice el funcional de Tikhonov dado por:

$$J_\alpha(f) := \|Af - \tau\|_2^2 + \alpha^2 \|f\|_2^2, \quad \forall f \in \mathbb{R}^4, \tau \in \mathbb{R}^3 \text{ y } \alpha > 0, \quad (4.16)$$

donde  $\alpha$  es el factor de Tikhonov que actúa como parámetro de regularización. En este caso, la solución regularizada del problema de PSD es denotada por  $f_\alpha$ . Sabemos que el parámetro de regularización  $\alpha > 0$ , controla el tamaño de la solución regularizada, es decir, se cumple que:

$$\begin{aligned} \|f_\alpha\|_2 &\rightarrow 0 \quad \text{cuando} \quad \alpha \rightarrow \infty, \\ \|f_\alpha\|_2 &\rightarrow \infty \quad \text{cuando} \quad \alpha \rightarrow 0. \end{aligned}$$

Dado que estamos considerando una aplicación sintética la solución exacta denotada por  $f(D) \in \mathbb{R}^4$  es conocida y está dada por:

$$f(D) = \begin{bmatrix} 5.9470 \times 10^{10} \\ 5.4496 \times 10^{10} \\ 5.1977 \times 10^{10} \\ 4.8776 \times 10^{10} \end{bmatrix};$$

con esto, se tiene que  $f(D)$  es la solución exacta la cual se quiere aproximar por la solución regularizada  $f_\alpha$ , además, se cumple que mientras más pequeño sea el tamaño del parámetro de regularización  $\alpha$ , la solución regularizada  $f_\alpha$  se acerca más a la solución exacta  $f(D)$ , es decir:

$$\|f_\alpha - f(D)\|_2 \rightarrow 0 \quad \text{cuando} \quad \alpha \rightarrow 0.$$

De manera general, sabemos que la solución regularizada  $f_\alpha$  se puede obtener resolviendo la ecuación (3.54), en este caso se tiene que:

$$f_\alpha = (\alpha^2 I + A^t A)^{-1} A^t \tau. \quad (4.17)$$

Dado que no conocemos el parámetro de regularización  $\alpha$ , el problema de hallar la matriz inversa  $(\alpha^2 I + A^t A)^{-1}$  resulta difícil. Para esto, consideramos la Descomposición en Valores Singulares de la matriz  $A \in M_{3 \times 4}(\mathbb{R})$ . Luego, por el Teorema 1.15 existe una DVS para la matriz  $A$  y por la Definición 1.23 se tiene que la DVS

está dada por:

$$A = U\Sigma V^t,$$

en su forma matricial se tiene que:

$$A = \underbrace{\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}}_{U_{3 \times 3}} \underbrace{\begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \end{bmatrix}}_{\Sigma_{3 \times 4}} \underbrace{\begin{bmatrix} v_1 & v_2 & v_3 & v_4 \end{bmatrix}^t}_{V_{4 \times 4}^t}.$$

Usando la función  $[U, S, V] = \text{svd}(A)$  de MatLab se obtiene:

$$U = \begin{bmatrix} -986.808413787363 & 160.039077808380 & 24.4263802621486 \\ -155.454187674477 & -894.590195019417 & -418.977778061784 \\ -45.2012169408779 & -417.247979682783 & 907.667876173713 \end{bmatrix} \times 10^{-3},$$

$$S = \begin{bmatrix} 236.760785418670 & 0 & 0 & 0 \\ 0 & 60.5937041072596 \times 10^{-6} & 0 & 0 \\ 0 & 0 & 27.6654149258401 \times 10^{-6} & 0 \end{bmatrix},$$

$$V = \begin{bmatrix} -413.908249327785 & -315.888061077944 & -103.176149702318 & -847.495944615330 \\ -493.798668119822 & -425.291165729621 & 689.642639382938 & 315.726669225716 \\ -523.124812832098 & 831.207539555861 & 172.510449331794 & -75.3299492646644 \\ -557.838029091672 & -168.630315798345 & -695.692357977922 & 419.991301033710 \end{bmatrix} \times 10^{-3}.$$

Sabemos que la solución regularizada usando la DVS de la matriz  $A$  está dada por la ecuación (3.60). En este caso, se cumple que  $r = \text{ran}(A) < n$ , por lo que la suma de la solución regularizada va desde  $i = 1, \dots, r$ . Así, es posible escribir la solución regularizada para el problema inverso de la estimación de PSD usando la DVS de la matriz  $A$ , como:

$$f_\alpha = \sum_{i=1}^r \left[ \frac{\sigma_i}{\alpha^2 + \sigma_i^2} u_i^t \tau \right] v_i. \quad (4.18)$$

Dado que la matriz  $A \in M_{3 \times 4}(\mathbb{R})$  es la representación matricial del operador lineal compacto  $K : \mathbb{R}^4 \rightarrow \mathbb{R}^3$ ,  $A^t \in M_{4 \times 3}$  es la representación matricial del operador adjunto  $K^* : \mathbb{R}^3 \rightarrow \mathbb{R}^4$ , y  $\sigma_1 \geq \sigma_2 \geq \sigma_3 > 0$ , dados por el vector (4.14) son los valores singulares de la matriz  $A$ , entonces según el Teorema 1.17 existen conjuntos

ortonormales  $\{u_i\} \subset \mathbb{R}^3$  y  $\{v_i\} \subset \mathbb{R}^4$  los cuales están dados por las columnas de la matriz  $U$  y la matriz  $V$ , respectivamente. Dichos sistemas ortonormales cumplen con las siguientes propiedades:

$$Av_i = \sigma_i u_i, \quad (4.19)$$

$$A^t u_i = \sigma_i v_i, \quad (4.20)$$

con  $i \in I = \{1, 2, 3\} \subset \mathbb{N}$ . Sabemos que por la DVS de  $A$  que hay más vectores por la derecha  $v_i$  que vectores por la izquierda  $u_i$ , y nos interesa saber que vectores por la derecha nos sirven para hallar la solución regularizada  $f_\alpha$ , por lo que los vectores  $v_i$  que nos sirven deben cumplir que:

$$u_i = \frac{Av_i}{\sigma_i} \quad i = 1, 2, 3.$$

En este caso, se tiene lo siguiente:

$$u_1 = \frac{Av_1}{\sigma_1} = \begin{bmatrix} -986.808413787363 \times 10^{-3} \\ -155.454187674477 \times 10^{-3} \\ -45.2012169408779 \times 10^{-3} \end{bmatrix},$$

$$u_2 = \frac{Av_2}{\sigma_2} = \begin{bmatrix} 160.039078121888 \times 10^{-3} \\ -894.590195019003 \times 10^{-3} \\ -417.247979673665 \times 10^{-3} \end{bmatrix},$$

$$u_3 = \frac{Av_3}{\sigma_3} = \begin{bmatrix} 24.4263804366343 \times 10^{-3} \\ -418.977778115748 \times 10^{-3} \\ 907.667876181879 \times 10^{-3} \end{bmatrix};$$

además, se tiene que:

$$Av_4 = \begin{bmatrix} 18.5739193958117 \times 10^{-15} \\ -2.44859727286224 \times 10^{-15} \\ 1.13275114008876 \times 10^{-15} \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Esto muestra que el sistema ortonormal  $(v_i)$  esta formado por las tres primeras columnas de la matriz  $V$  de la DVS de  $A$ . Luego, es posible calcular solución regularizada  $f_\alpha$  como:

$$\begin{aligned} f_\alpha &= \sum_{i=1}^3 \left[ \frac{\sigma_i}{\alpha^2 + \sigma_i^2} u_i^t \tau \right] v_i \\ &= \left[ \frac{\sigma_1}{\alpha^2 + \sigma_1^2} u_1^t \tau \right] v_1 + \left[ \frac{\sigma_2}{\alpha^2 + \sigma_2^2} u_2^t \tau \right] v_2 + \left[ \frac{\sigma_3}{\alpha^2 + \sigma_3^2} u_3^t \tau \right] v_3. \end{aligned}$$

Para calcular la solución regularizada  $f_\alpha$  solo basta elegir correctamente el parámetro de regularización o factor de Tikhonov  $\alpha > 0$ . Si le damos valores a  $\alpha$  de tal forma que cada vez sea más pequeño obtenemos lo siguiente:

- Para  $\alpha = 1$  :

$$f_\alpha = \begin{bmatrix} 4.38421797581138 \times 10^{10} \\ 5.23043693814276 \times 10^{10} \\ 5.54106671580672 \times 10^{10} \\ 5.90875761045696 \times 10^{10} \end{bmatrix}.$$

- Para  $\alpha = 0.5$  :

$$f_\alpha = \begin{bmatrix} 4.38427660522684 \times 10^{10} \\ 5.23050687725977 \times 10^{10} \\ 5.54114092962522 \times 10^{10} \\ 5.90883665273824 \times 10^{10} \end{bmatrix}.$$

- Para  $\alpha = 0.3$ :

$$f_\alpha = \begin{bmatrix} 4.38428905001994 \times 10^{10} \\ 5.23052170833471 \times 10^{10} \\ 5.54115692854422 \times 10^{10} \\ 5.90885348533342 \times 10^{10} \end{bmatrix}.$$

Con esto hemos encontrado la solución aproximada al problema inverso de estimación de PSD usando la técnica óptica de turbidimetría. Existen métodos para obtener el parámetro de regularización  $\alpha$  de tal forma que sea óptimo, es decir, que la solución regularizada  $f_\alpha$  sea una buena aproximación a la solución exacta  $f(D)$ . El estudio de estos métodos escapa a los alcances de esta tesis, por lo que se deja como un problema abierto.

# Conclusiones

En esta tesis se propone una metodología para resolver el problema inverso de estimación de PSD mediante la técnica óptica de Turbidimetría. Para esto, se resolvió un problema inverso mal condicionado descrito por una ecuación integral de Fredholm, la cual relaciona las mediciones de espectro de turbidez con la función de distribución de tamaños de partículas a partir de la teoría de Mie. Se estudió teóricamente un experimento de turbidimetría y con ello los conceptos físicos más importantes relacionados con la teoría de Rayleigh-Mie como la dispersión de luz, el índice de refracción de una partícula y del medio en la que está inmersa, la absorción y extinción de luz y los coeficientes de Mie.

Dado que resolver la ecuación integral de Fredholm que modela el problema inverso resulta complicado, en su lugar se resolvió el problema inverso discreto que resulta en un sistema de ecuaciones lineales mal condicionado, el cual surge tomando en cuenta un conjunto finito de diámetros promedios de partículas y de longitudes de onda. Para la solución del problema inverso discreto se estudiaron diferentes métodos, entre ellos la inversa generalizada implementada por la inversa de Moore-Penrose y la regularización de Tikhonov implementada mediante la descomposición en valores singulares de la matriz del sistema.

Para comprender mejor el problema inverso de estimación de PSD, se realizó una aplicación sintética considerando diámetros promedios de partículas de NBR entre 80 y 85 nanómetros y longitudes de onda entre 300 y 600 nanómetros. Primero se resuelve el problema directo para el cual se supone conocida la función de distribución de tamaños de partículas de tipo logarítmica normal; y se realizan los cálculos para hallar los coeficientes de la matriz del sistema. La solución del problema directo arroja como resultado una aproximación muy cercana al espectro de turbidez original. Después se resuelve el problema inverso usando la inversa generalizada de Penrose, la cual no es suficiente para hallar la solución aproximada, pues el problema es mal condicionado. Por esto el problema inverso fue regula-

rizado usando el método de regularización de Tikhonov. La solución encontrada depende del parámetro de regularización de Tikhonov, por lo que fue necesario darle valores a dicho parámetro para hallar la aproximación más cercana a la solución original del problema. En este trabajo solo se consideran los valores 1, 0.5 y 0.3, con los cuales se obtuvieron soluciones muy cercanas a la original, lo que comprueba que la distribución de tamaños de partículas es del tipo logarítmica normal. El problema de hallar un parámetro de regularización que mejor aproxime la solución aproximada a la original escapa de los alcances de esta tesis, por lo que queda como un problema abierto, al igual que el problema de considerar conjuntos de diámetros y longitudes de onda más grandes de tal forma que la matriz del sistema sea de mayores dimensiones. Una de las limitaciones que se encontraron al escribir esta tesis fue el de hallar datos reales que encajaran con el modelo estudiado, ya que no se contaba con un espectrofotómetro para realizar las mediciones, y así poder validar los cálculos matemáticos obtenidos.

El objetivo de esta tesis es brindar la teoría necesaria para comprender el problema inverso discreto de estimación de PSD basado en mediciones de turbidimetría, y aplicar dichos conocimientos adquiridos en un problema sintético; este trabajo aporta la implementación de la regularización de Tikhonov para solución del problema inverso de estimación de PSD, ya que en la literatura dicho problema es abordado por medio de mínimos cuadrados y redes neuronales.

# Apéndice A

## Conceptos físicos básicos

En este apéndice hablaremos principalmente sobre los conceptos básicos de la luz que nos permitirán comprender mejor su naturaleza.

La luz es esencial para casi todo en la vida en la tierra. La luz es el principal medio por el cual es posible transmitir y recibir información hacia y desde objetos situados en el entorno inmediato y en todo el universo. Muchos de los fenómenos de la vida diaria dependen de las propiedades de la luz. El color azul del cielo diurno es resultado del fenómeno óptico de la dispersión de la luz por las moléculas del aire, como lo son los colores rojo y anaranjado de los amaneceres y atardeceres. Nuestra imagen se puede observar en un espejo debido al fenómeno de la reflexión de la luz. Al usar anteojos o lentes de contacto es necesario que ocurra el fenómeno de la refracción de la luz para una visión clara. Los colores del arcoiris resultan de la dispersión de la luz conforme pasa a través de las gotas de lluvia que flotan en el cielo después de una tormenta.

Este apéndice es un resumen del capítulo 35 “Naturaleza de la luz y leyes de óptica geométrica” del libro [37].

### Naturaleza de la luz

En el siglo *XIX*, la luz era considerada un flujo de partículas que eran emitidas por un objeto observado. Isaac Newton, principal arquitecto del modelo de las partículas de la luz, afirmaba que estas eran emitidas por una fuente luminosa y que estimulaban el sentido de la vista al entrar en los ojos del observador. Con esta idea pudo explicar la reflexión y la refracción. En 1678, el físico y astrónomo holandés Christian Huygens demostró que una teoría de ondas de luz podría tam-

bién explicar la reflexión y la refracción. En 1801, Thomas Young dio la primera demostración clara de la naturaleza ondulatoria de la luz. Demostró que, bajo condiciones apropiadas, los rayos de luz se interfieren unos con otros. Tal comportamiento no podía ser explicado en aquel tiempo por una teoría de partículas porque no había forma imaginable en que dos o más partículas pudieran unirse y cancelarse entre sí. En 1873, James Maxwell afirmó que la luz era una forma de onda electromagnética de alta frecuencia. A pesar de estudiar la luz como una onda, dicho modelo ondulatorio de la luz, no puede explicar ciertos experimentos. El más notable de estos es el efecto fotoeléctrico, descubierto por Hertz: cuando incide luz sobre una superficie metálica, a veces se expulsan electrones de la superficie. Una de las dificultades más importantes que surgieron, fue que algunos experimentos demostraron que la energía cinética de un electrón expulsado es independiente de la intensidad de la luz. Este hallazgo contradujo el modelo de onda, que sostenía que un haz luminoso más intenso adiciona más energía al electrón. Einstein propuso una explicación del efecto fotoeléctrico en 1905 aplicando un modelo de acuerdo con el concepto de cuantización desarrollado por Max Planck en 1900. El modelo de cuantización supone que la energía de una onda luminosa está presente en partículas llamadas fotones; por tanto, se dice que la energía está cuantizada. Según la teoría de Einstein, la energía de un fotón es proporcional a la frecuencia de la onda electromagnética:

$$E = hf \tag{A.1}$$

donde la constante de proporcionalidad  $h = 6.63 \times 10^{-34} J \cdot s$  es la constante de Planck. En vista de estos desarrollos, debe considerarse que la luz tiene doble naturaleza: en algunos casos exhibe características de una onda y en otras de una partícula.

### Aproximación de un rayo en óptica geométrica

El campo de la óptica geométrica abarca el estudio de la propagación de la luz, a partir del supuesto de que la luz se desplaza en una dirección fija y en línea recta cuando pasa por un medio uniforme, y cambia su dirección en el momento en que se encuentra con la superficie de un medio diferente o si las propiedades ópticas del medio no son uniformes ya sea en espacio o en tiempo. Para comprender la aproximación de un rayo primero observe que los rayos de una onda determinada

son líneas rectas perpendiculares a los frentes de onda, como se ilustra en la Figura A.1 para una onda plana.

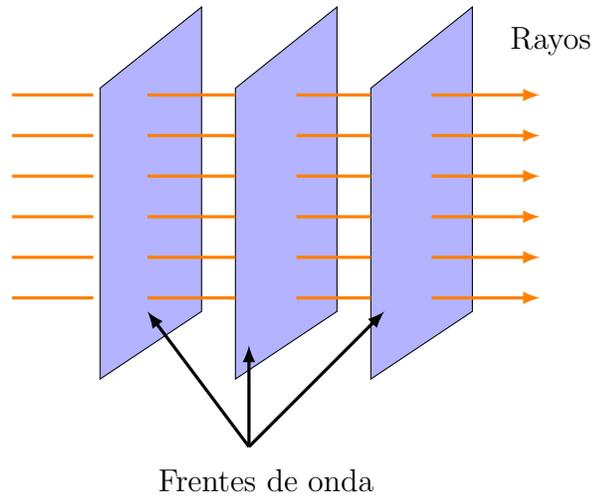


Figura A.1: Óptica geométrica.

En la aproximación de un rayo, una onda que se mueve en un medio se desplaza en línea recta en la dirección de sus rayos. Si la onda se encuentra con una barrera en la que hay una abertura circular cuyo diámetro ( $d$ ) es mucho mayor que la longitud de onda ( $\lambda$ ), como se muestra en la Figura A.2 la onda que emerge de la abertura continúa moviéndose en línea recta. En este caso, cuando  $\lambda \ll d$ , el cuerpo arroja una sombra nítida. La aproximación de un rayo y la suposición de que  $\lambda \ll d$ , es muy buena para el estudio de espejos, lentes, prismas e instrumentos ópticos asociados, por ejemplo telescopios, cámaras y anteojos.

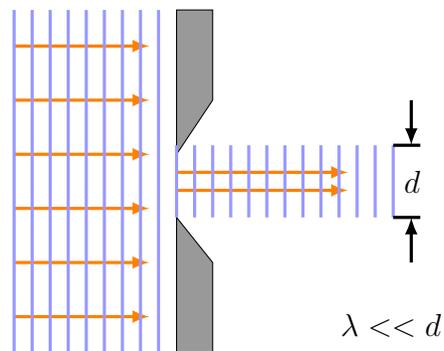


Figura A.2: Trayectoria en línea recta.

Si el diámetro de la abertura es del orden de una longitud de onda, como se muestra en la Figura A.3 las ondas se extienden desde la abertura en todas direcciones. Este efecto se llama difracción.

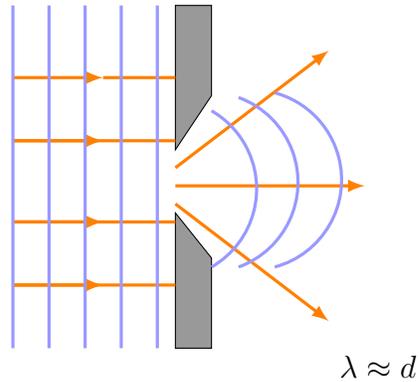


Figura A.3: Difracción de ondas.

Por último, si la abertura es mucho menor que la longitud de onda, la abertura se aproxima como una fuente puntual de ondas como se muestra en la Figura A.4. Efectos similares se observan cuando las ondas se encuentran con un objeto opaco de dimensión  $d$ .

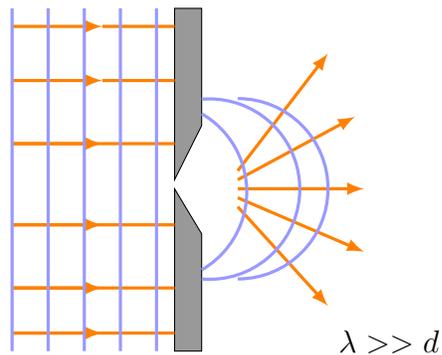


Figura A.4: Fuente puntual de ondas

## La onda bajo reflexión

Una onda corresponde a la propagación de una perturbación de un lugar a otro. Se puede afirmar que la perturbación se propaga por el medio material, la forma de la perturbación es conservada y la velocidad de propagación es constante.

El modelo de onda progresiva describe ondas que viajan a través de un medio uniforme sin interactuar con algo más en el camino. Consideremos como una onda progresiva es afectada cuando encuentra un cambio en el medio. Por ejemplo, considere un pulso que viaja en una cuerda que esta rígidamente unida a un soporte en un extremo, como se muestra en la Figura A.5.

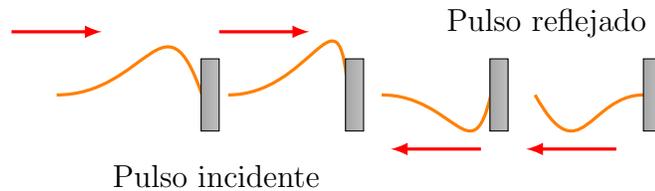


Figura A.5: Reflexión de un pulso con extremo fijo.

Cuando el pulso alcanza el soporte, se presenta un cambio severo en el medio, es decir, la cuerda termina. Como resultado, el pulso experimenta reflexión; esto es, el pulso se mueve de regreso a lo largo de la cuerda en la dirección opuesta. Note que el pulso reflejado está invertido. Esta inversión puede ser explicada de la siguiente manera: cuando el pulso alcanza el extremo fijo de la cuerda, esta produce una fuerza hacia arriba sobre el soporte. Por la tercera ley de Newton<sup>1</sup> el soporte debe ejercer sobre la cuerda una fuerza de reacción de igual magnitud y con dirección opuesta (hacia abajo). Esta fuerza hacia abajo hace que el pulso se invierta en la reflexión.

Como el ejemplo anterior, donde consideramos ondas sobre una cuerda, cuando un rayo de luz que se desplaza en un medio encuentra una frontera con otro medio, parte de la luz incidente se refleja: para ondas sobre una cuerda unidimensional, la onda reflejada necesariamente debe restringirse a una dirección a lo largo de la cuerda, como en la Figura A.5; para ondas de luz que viajan en el espacio tridimensional, no se aplican tales restricciones y las ondas de luz reflejadas pueden estar en direcciones distintas de la dirección de las ondas incidentes.

La Figura A.6 muestra varios rayos de un haz de luz incidente en una superficie reflectora lisa, semejante a un espejo. Los rayos reflejados son paralelos entre sí. La dirección de un rayo reflejado está en el plano perpendicular a la superficie reflectora que contiene al rayo incidente. La reflexión de la luz desde esta superficie

<sup>1</sup>Tercera ley de Newton: cuando un cuerpo ejerce una fuerza sobre otro, este ejerce sobre el primero una fuerza igual y de sentido opuesto.

lisa se denomina reflexión especular. Una superficie se comporta como superficie lisa mientras las variaciones de superficie son mucho menores que la longitud de onda de la luz incidente.

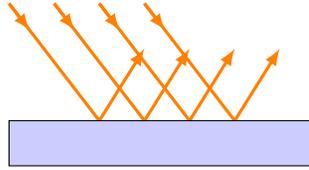


Figura A.6: Reflexión especular.

Si la superficie reflectora es rugosa, como muestra la Figura A.7 la superficie refleja los rayos no como un conjunto paralelo sino en varias direcciones. La reflexión desde cualquier superficie rugosa se conoce como reflexión difusa.

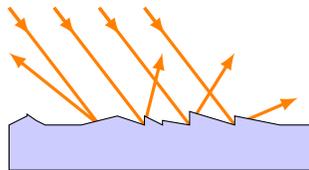


Figura A.7: Reflexión difusa.

En lo posterior, se analizará la reflexión especular y se usará el término reflexión, al referirnos ha dicho tipo de reflexión. Considere un rayo de luz que viaja en el aire y que incide a un ángulo en una superficie plana y lisa, como se ve en la Figura A.8.

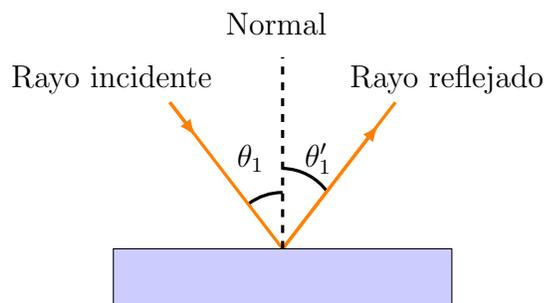


Figura A.8: Reflexión especular.

Los rayos incidente y reflejado forman los ángulos  $\theta_1$ <sup>2</sup> y  $\theta'_1$ , respectivamente, donde los ángulos se observan entre la normal<sup>3</sup> y los rayos. Experimentos y teoría muestran que el ángulo de reflexión es igual ángulo de incidencia:

$$\theta'_1 = \theta_1. \quad (\text{A.2})$$

La correspondencia se denomina ley de reflexión. Porque la reflexión de ondas desde una interfaz<sup>4</sup> entre dos medios es un fenómeno común, se identifica un modelo de análisis para esta situación: la onda bajo reflexión.

## La onda bajo refracción

Además del fenómeno de la reflexión para las ondas sobre cuerdas, también se encontró que parte de la energía de la onda incidente se transmite en el nuevo medio. De igual modo, cuando un rayo de luz que se mueve por un medio transparente encuentra una frontera que lleva a otro medio de igual característica, como se muestra en la Figura A.9 parte de la energía se refleja y parte penetra al segundo medio. De igual manera que en la reflexión, la dirección de la onda transmitida muestra un comportamiento interesante debido a la naturaleza tridimensional de las ondas de luz. El rayo que penetra al segundo medio se dobla en la frontera y se dice que se refracta.

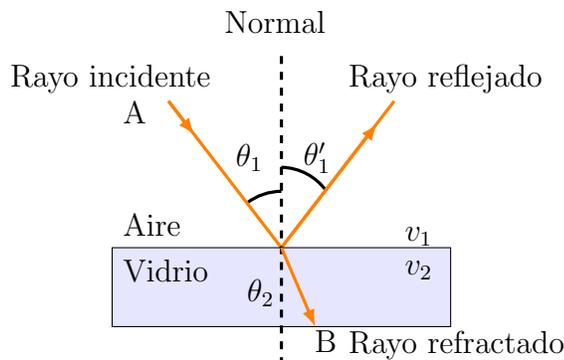


Figura A.9: Refracción.

<sup>2</sup>Notación: El subíndice 1 se utiliza para indicar los parámetros para la luz en el medio inicial. Cuando la luz pasa de un medio a otro, se usa el subíndice 2 para los parámetros asociados con la luz en el nuevo medio.

<sup>3</sup>Es una línea con trazo perpendicular a la superficie en el punto donde el rayo incidente cae en la superficie.

<sup>4</sup>Zona de comunicación o acción de un sistema sobre otro.

El rayo incidente, el rayo reflejado y el rayo refractado todos se encuentran en el mismo plano. La relación entre el ángulo de refracción  $\theta_2$ , las propiedades de los dos medios y el ángulo de incidencia se da a través de la siguiente correspondencia:

$$\frac{\text{sen } \theta_2}{\text{sen } \theta_1} = \frac{v_2}{v_1}; \quad (\text{A.3})$$

donde  $v_1$  es la rapidez de la luz en el primer medio y  $v_2$  es la rapidez de la luz en el segundo. La trayectoria de un rayo de luz que pasa por una superficie refractaria es reversible. Por ejemplo, el rayo que se ilustra en la Figura A.9 pasa del punto  $A$  al punto  $B$ . Si el rayo se originó en  $B$ , viajaría a lo largo de la recta  $BA$  hasta llegar al punto  $A$ , y la parte reflejada apuntaría hacia abajo y a la izquierda del vidrio. Por la ecuación (A.3) es posible inferir que cuando la luz se mueve de un material en el que su rapidez es alta a un material en el que su rapidez es menor, como se observa en la Figura A.10 el ángulo de refracción  $\theta_2$  es menor que el ángulo de incidencia  $\theta_1$ , y el rayo se dobla hacia la normal.

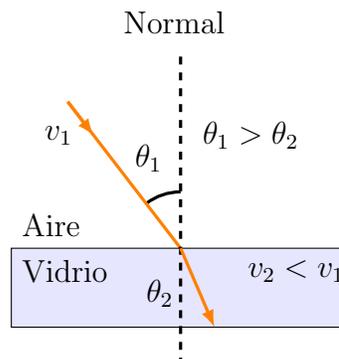


Figura A.10: Cuando un haz de luz pasa del aire al vidrio, la luz disminuye su velocidad al entrar a éste y su trayectoria se dobla hacia la normal.

Si la luz se mueve en el aire, su rapidez es de  $3 \times 10^8 \text{ m/s}$ , pero esta rapidez se reduce a casi  $2 \times 10^8 \text{ m/s}$  si la luz entra a un bloque de vidrio. Cuando la luz emerge de nuevo hacia el aire, su rapidez aumenta en forma instantánea a su valor original de  $3 \times 10^8 \text{ m/s}$ . Si el rayo se mueve de un material en el que la luz se mueve con más lentitud hacia un material en el que se mueve con más rapidez, como se muestra en la Figura A.11 el ángulo  $\theta_2$  es mayor que  $\theta_1$  y el rayo se dobla alejándose de la normal.

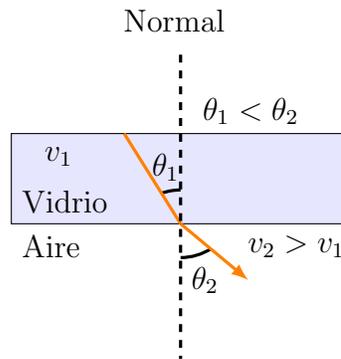


Figura A.11: Cuando el haz de luz se mueve del vidrio al aire, la luz aumenta su velocidad al entrar al aire y su trayectoria se dobla alejándose de la normal.

Para ver por qué la luz se comporta como lo hace, considere la Figura A.12 que representa un haz de luz entrando en un trozo de vidrio desde la izquierda. Una vez dentro del vidrio, la luz puede encontrar un electrón ligado a un átomo, indicado como un punto  $A$ . Suponga que la luz es absorbida por el átomo; esto hace que el electrón oscile. El electrón oscilante en tal caso actúa como una antena y emite el haz de luz hacia un átomo en  $B$ , donde la luz es absorbida de nuevo.

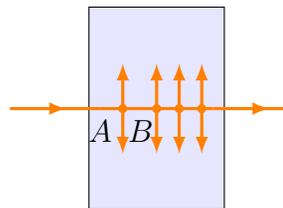


Figura A.12: Luz que pasa de un átomo a otro en un medio. Los puntos son electrones y las flechas verticales representan sus oscilaciones.

Aún cuando la luz se mueve de un átomo a otro a  $3 \times 10^8 \text{ m/s}$ , la absorción y la radiación que tienen lugar hacen que el promedio de la rapidez de la luz que pasa por el material disminuya a unos  $2 \times 10^8 \text{ m/s}$ . Una vez que la luz emerge hacia el aire, la absorción y la radiación cesan y la luz viaja con una rapidez constante de  $3 \times 10^8 \text{ m/s}$ .

## Índice de refracción

En general, la rapidez de la luz en cualquier material es menor que en el vacío. De hecho, la luz se desplaza a su máxima rapidez en el vacío. Es conveniente definir el índice de refracción<sup>5</sup>  $n$  de un medio como la siguiente relación:

$$n = \frac{\text{rapidez de la luz en el vacío}}{\text{rapidez de la luz en un medio}} \equiv \frac{c}{v}. \quad (\text{A.4})$$

Por esta definición, queda claro que el índice de refracción es un número sin dimensiones mayor que la unidad por que  $v$  siempre es menor que  $c$ . Además, el índice de refracción  $n$  es igual a la unidad para el vacío.

Cuando la luz pasa de un medio a otro, su frecuencia no cambia, pero sí lo hace su longitud de onda. Para ver como ocurre esto, consideremos la Figura A.13. Las ondas pasan junto a un observador situado en el punto  $A$  en el medio 1 con cierta frecuencia e inciden en la frontera entre el medio 1 y el medio 2. La frecuencia a la que pasan las ondas junto a un observador situado en el punto  $B$  en el medio 2 debe ser igual a la frecuencia a la que pasa en el punto  $A$ . Si este no fuera el caso, la energía se acumularía o desaparecería en la frontera. Como no hay mecanismo para que esto se presente, la frecuencia debe ser constante cuando un rayo de luz pasa de un medio a otro.

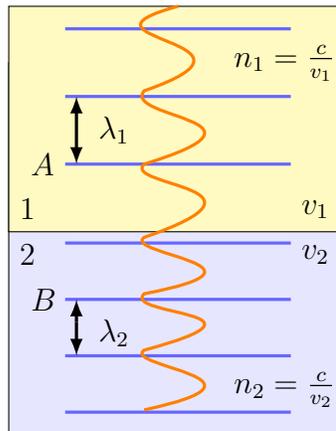


Figura A.13: Cuando una onda se mueve del medio 1 al medio 2, cambia su longitud de onda, pero su frecuencia permanece constante.

<sup>5</sup>El índice de refracción  $n$  aquí no es un entero.

Recordemos que la rapidez de una onda sinusoidal se define como:

$$v = \lambda f; \quad (\text{A.5})$$

donde,  $\lambda$  es la longitud de onda y  $f$  la frecuencia.

Dado que la correspondencia de la ecuación (A.5) debe ser válida en ambos medios, y como  $f_1 = f_2 = f$ , se tiene que:

$$v_1 = \lambda_1 f \quad y \quad v_2 = \lambda_2 f. \quad (\text{A.6})$$

Como  $v_1 \neq v_2$ , se deduce que  $\lambda_1 \neq \lambda_2$ , como se muestra en la Figura A.13. Al dividir la primera ecuación (A.5) entre la segunda, obtenemos que:

$$\frac{\lambda_1}{\lambda_2} = \frac{v_1}{v_2};$$

luego, al usar la ecuación (A.4) se obtiene una correspondencia entre el índice de refracción y la longitud de onda:

$$\frac{\lambda_1}{\lambda_2} = \frac{v_1}{v_2} = \frac{c/n_1}{c/n_2} = \frac{n_2}{n_1}. \quad (\text{A.7})$$

La expresión anterior nos da:

$$\lambda_1 n_1 = \lambda_2 n_2.$$

Si el medio 1 es el vacío, o aire, entonces  $n_1 = 1$ . Por esto, se deduce por la ecuación (A.7) que el índice de refracción de cualquier medio se expresa como:

$$n = \frac{\lambda}{\lambda_n}; \quad (\text{A.8})$$

donde  $\lambda$  es la longitud de onda de la luz en el vacío y  $\lambda_n$  es la longitud de onda de la luz en el medio cuyo índice de refracción es  $n$ .

Dado que el índice de refracción  $n$  es un número sin dimensiones mayor que la unidad ( $n > 1$ ), entonces  $\lambda > \lambda_n$ .

Ahora podemos expresar la ecuación (A.3) de una manera diferente. Al sustituir el término  $v_2/v_1$  de la ecuación (A.3) con  $n_1/n_2$  de la ecuación (A.7), se obtiene:

$$\frac{\text{sen } \theta_2}{\text{sen } \theta_1} = \frac{n_1}{n_2}.$$

Por lo tanto, esta última expresión se convierte en la ley de refracción de Snell, dada por:

$$n_1 \operatorname{sen} \theta_1 = n_2 \operatorname{sen} \theta_2 \quad (\text{A.9})$$

Este descubrimiento experimental de esta correspondencia suele acreditarse a Willebrord Snell(1591-1627). La refracción de las ondas en una interfaz entre dos medios, es un fenómeno común y se puede identificar un modelo de análisis para esta situación: la onda bajo refracción. La ecuación (A.9) es la representación matemática de este modelo para la radiación electromagnética.

## Principio de Huygens

En esta sección se desarrollan las leyes de reflexión y de la refracción mediante el uso de un método geométrico propuesto por Huygens en 1678. El principio de Huygens es una construcción geométrica para usar el conocimiento de un frente de onda anterior, para determinar la posición de un frente de onda nuevo en algún instante. En la construcción de Huygens todos los puntos en un frente de onda determinado se toman como fuentes puntuales para la producción de ondas esféricas secundarias, llamadas trenes de onda, que se propagan hacia afuera por un medio con magnitudes de velocidad características de ondas en ese medio. Después de algún intervalo de tiempo transcurrido la nueva posición del frente de onda es la superficie tangente a los trenes de onda.

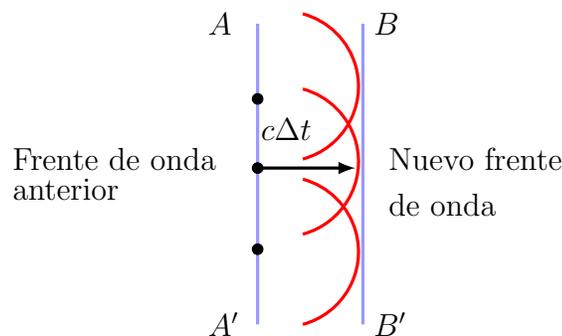


Figura A.14: Construcción de Huygens para una onda plana que se propaga a la derecha.

Consideremos una onda plana que se mueve en el espacio libre, como se muestra en la Figura A.14. En  $t = 0$ , el frente de onda está indicado por el plano marcado como  $AA'$ . En una construcción de Huygens, cada punto de este frente de onda

es considerado fuente puntual. Con estos puntos como fuentes para los trenes de ondas, se trazan círculos, cada uno con radio  $c \Delta t$ , donde  $c$  es la rapidez de la luz en el vacío y  $\Delta t$  es cierto intervalo de tiempo durante el cual se propaga la onda. La superficie trazada tangente a estos trenes de onda es el plano  $BB'$ , que es el frente de onda en un tiempo posterior, y es paralelo a  $AA'$ .

## Principio de Huygens aplicado a la reflexión

En esta sección se deduce la ley de reflexión con el uso del principio de Huygens. Para la ley de reflexión, consideremos la Figura A.15. La recta  $AB$  representa un frente de onda plana de la luz incidente precisamente cuando el rayo 1 incide en la superficie. En ese instante, la onda en  $A$  envía un tren de ondas de Huygens (el círculo en rojo con centro en  $A$ ). La luz reflejada se propaga hacia  $D$ . Al mismo tiempo, la onda en  $B$  emite un tren de ondas de Huygens (el círculo en rojo con centro  $B$ ) con la propagación de luz hacia  $C$ . La Figura A.15 muestra estos trenes de ondas después de un intervalo  $\Delta t$ , después del cual el rayo 2 incide en la superficie. Como los rayos 1 y 2 se mueven a la misma rapidez, se debe obtener  $AD = BC = c\Delta t$ .

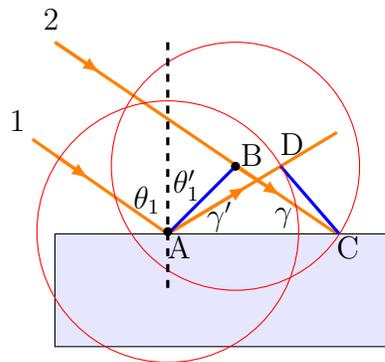


Figura A.15: Construcción de Huygens para demostrar la ley de reflexión.

Observe que los dos triángulos,  $ABC$  y  $ADC$ , son congruentes por que tienen la misma hipotenusa  $AC$  y porque  $AD = BC$ . Luego, tendremos que:

$$\cos \gamma = \frac{BC}{AC} \quad \text{y} \quad \cos \gamma' = \frac{AD}{AC};$$

donde  $\gamma = 90^\circ - \theta_1$  y  $\gamma' = 90^\circ - \theta'_1$ . Como  $AD = BC$ , tendremos que:

$$\cos \gamma = \cos \gamma'.$$

Por lo tanto, se debe cumplir que:

$$\begin{aligned} \gamma &= \gamma', \\ 90^\circ - \theta_1 &= 90^\circ - \theta'_1; \end{aligned}$$

así,

$$\theta_1 = \theta'_1, \tag{A.10}$$

que es la ley de reflexión.

## Principio de Huygens aplicado a la refracción

Para la deducción de la ley de refracción de Snell, usemos el principio de Huygens y la Figura A.16. Concentrémonos en el instante en que el rayo 1 incide sobre la superficie y en el intervalo de tiempo consecutivo hasta que el rayo 2 hace lo mismo. Durante este intervalo de tiempo, la onda en  $A$  envía un tren de ondas de Huygens (el círculo rojo con centro en  $A$ ) y la luz refracta hacia  $D$ . En el mismo intervalo de tiempo, la onda en  $B$  envía un tren de ondas de Huygens (el círculo rojo con centro en  $B$ ) y la luz continúa su propagación hacia  $C$ . Ya que estos dos trenes de onda se desplazan en medios diferentes, los radios de los trenes de ondas son diferentes. El radio del tren de ondas desde  $A$  es  $AD = v_2 \Delta t$ , donde  $v_2$  es la rapidez de la onda en el segundo medio. El radio del tren de ondas desde  $B$  es  $BC = v_1 \Delta t$ , donde  $v_1$  es la rapidez de la onda en el medio original. A partir de los triángulos  $ABC$  y  $ADC$ , se tiene que:

$$\text{sen } \theta_1 = \frac{BC}{AC} = \frac{v_1 \Delta t}{AC} \quad \text{y} \quad \text{sen } \theta_2 = \frac{AD}{AC} = \frac{v_2 \Delta t}{AC}.$$

Si divide la primera ecuación entre la segunda, se obtiene:

$$\frac{\text{sen } \theta_1}{\text{sen } \theta_2} = \frac{v_1}{v_2}.$$

Pero por la ecuación (A.4) sabemos que  $v_1 = c/n_1$  y  $v_2 = c/n_2$ . Por lo tanto,

$$\frac{\text{sen } \theta_1}{\text{sen } \theta_2} = \frac{c/n_1}{c/n_2} = \frac{n_2}{n_1}$$

y así,

$$n_1 \text{ sen } \theta_1 = n_2 \text{ sen } \theta_2$$

que es la ley de refracción de Snell.

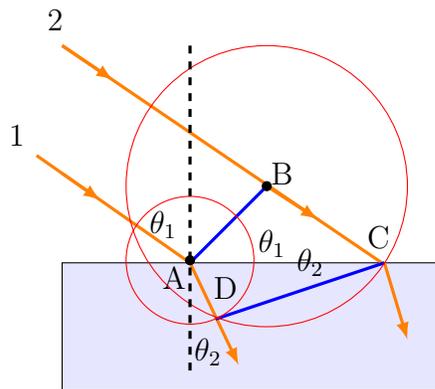


Figura A.16: Construcción de Huygens para demostrar la ley de la refracción de Snell.

## Dispersión

Una propiedad importante del índice de refracción  $n$  es que, para un material determinado, el índice varía con la longitud de onda de la luz que pase por el material. Este comportamiento se denomina dispersión. Como  $n$  es una función que depende de la longitud de onda, la ley de la refracción de Snell indica que luces de diferentes longitudes de onda se refractan a diferentes ángulos cuando inciden sobre un material. El índice de refracción generalmente disminuye con una longitud de onda creciente. Suponga que un haz de luz blanca (combinación de todas las longitudes de onda visibles) incide en un prisma. Claramente, el ángulo de desviación  $\delta$  depende de la longitud de onda. Los rayos que emergen se dispersan en una serie de colores conocida como espectro visible. Estos colores, en orden de longitud de onda decreciente son rojo, naranja, amarillo, verde, azul y violeta. Newton demostró que cada color tiene un ángulo particular de desviación y que los colores se pueden recombinar para formar la luz blanca original. La dispersión

de luz en un espectro se comprueba con mayor claridad en la naturaleza con la formación de un arco iris, el cual es visto frecuentemente por un observador que esta situado entre el sol y una zona con lluvia. para comprender como se forma considere el siguiente ejemplo.

**Ejemplo A.1.** (*El cálculo de los arcos iris*)

Desde los tiempos del filósofo Aristóteles, la formación de los arcos iris han fascinado e inspirado a la humanidad para realizar múltiples intentos de explicación científica. Los arcos iris son fenómenos ópticos que consisten en la aparición de arcos en el cielo de luz multicolor, dicho fenómeno es originado por la descomposición de la luz solar en el espectro visible, lo cual es producido por refracción, cuando las diferentes longitudes de onda de la luz solar atraviesan gotas de agua esféricas contenidas en la atmósfera. Según ideas de Newton y Descartes, las gotas de lluvia dispersan la luz solar en una diferente gama de longitudes de onda.

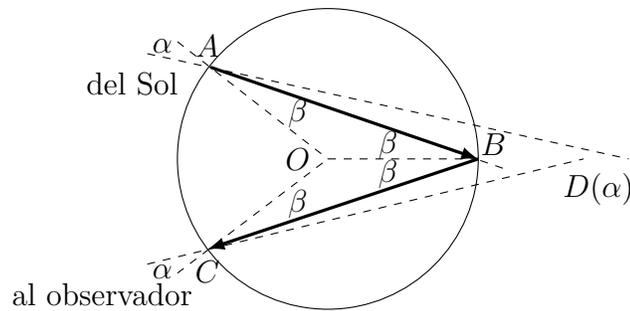


Figura A.17: Formación del arco iris primario.

En la Figura A.17 mostramos un rayo de luz solar que entra en una gota de lluvia esférica en el punto  $A$ . La recta  $AB$  muestra la trayectoria del rayo de luz solar que entra en la gota de lluvia. Con base en la ley de Snell la luz se refracta hacia la recta normal  $AO$  y establece que:

$$\text{sen}(\alpha) = k\text{sen}(\beta),$$

donde:

- $\alpha$  es el ángulo de incidencia.
- $\beta$  es el ángulo de refracción.
- $k \approx \frac{4}{3}$  es el índice de refracción para el agua.

En el punto  $B$  parte de la luz del rayo solar pasa a través de la gota de lluvia y se refracta hacia el aire, pero nos interesa la parte de luz que se refleja en la recta  $BC$ . Sabemos por las leyes de reflexión que el ángulo de incidencia es igual al ángulo de reflexión. En el punto  $C$ , parte del rayo se refleja, pero nos interesa la parte del rayo que sale de la gota de lluvia esférica. El ángulo de desviación  $D(\alpha)$  es la magnitud de la rotación en el sentido de las manecillas del reloj que describe el rayo de luz solar al entrar en la gota. Dicha desviación se define como:

$$D(\alpha) = (\alpha - \beta) + (\pi - 2\beta) + (\alpha - \beta).$$

Al simplificar se tiene que:

$$D(\alpha) = \pi + 2\alpha - 4\beta. \quad (\text{A.11})$$

Estamos interesados en el valor mínimo de la desviación  $D(\alpha)$  y el valor del ángulo de incidencia del rayo de luz solar  $\alpha$ , en el que se da dicho valor mínimo.

De la ley de Snell se tiene que:

$$\text{sen}\beta = \frac{1}{k}\text{sen}\alpha;$$

luego, se cumplirá que:

$$\beta = \arcsen\left(\frac{\text{sen}(\alpha)}{k}\right). \quad (\text{A.12})$$

Al sustituir el valor de  $\beta$  en la función de desviación  $D(\alpha)$ , se obtiene:

$$D(\alpha) = \pi + 2\alpha - 4 \arcsen\left(\frac{\text{sen}(\alpha)}{k}\right).$$

Recordemos que para encontrar el valor mínimo de una función continua es necesario encontrar un punto crítico a partir del cálculo de la derivada de  $D(\alpha)$ . Antes de realizar el cálculo de la derivada, recordemos lo siguiente:

$$\frac{d}{du}\left(\arcsen\left(\frac{u}{a}\right)\right) = \frac{du}{\sqrt{a^2 - u^2}}.$$

De aquí, se tiene que la derivada de  $D(\alpha)$  esta dada por:

$$\frac{d}{d\alpha}D(\alpha) = 2 - \frac{4 \frac{d}{d\alpha}(\text{sen}(\alpha))}{\sqrt{k^2 - \text{sen}^2(\alpha)}}.$$

Al simplificar se obtiene lo siguiente:

$$\frac{d}{d\alpha}D(\alpha) = \frac{2\sqrt{k^2 - \text{sen}^2(\alpha)} - 4 \cos(\alpha)}{\sqrt{k^2 - \text{sen}^2(\alpha)}}.$$

Ahora, es necesario encontrar los puntos críticos de la función  $D(\alpha)$ , es decir, necesitamos encontrar los puntos en los que se cumple que:

$$\frac{d}{d\alpha}D(\alpha) = 0.$$

Notemos que  $\sqrt{k^2 - \text{sen}^2(\alpha)} > 0$ , para cualquier valor del ángulo  $\alpha$ . Por lo que, la igualdad solo se cumplirá, si y solo si,

$$\begin{aligned} 2\sqrt{k^2 - \text{sen}^2(\alpha)} - 4 \cos(\alpha) &= 0, \\ \Rightarrow \sqrt{k^2 - \text{sen}^2(\alpha)} &= 2 \cos(\alpha), \\ \Rightarrow k^2 - \text{sen}^2(\alpha) &= 4 \cos^2(\alpha), \\ \Rightarrow k^2 &= 4 \cos^2(\alpha) + \text{sen}^2(\alpha). \end{aligned}$$

Luego, al sustituir la identidad pitagórica,  $\cos^2(\alpha) = 1 - \text{sen}^2(\alpha)$ , en la última ecuación, se sigue que:

$$\begin{aligned} k^2 &= 4(1 - \text{sen}^2(\alpha)) + \text{sen}^2(\alpha) \\ &= 4 - 4 \text{sen}^2(\alpha) + \text{sen}^2(\alpha) \\ &= 4 - 3 \text{sen}^2(\alpha). \end{aligned}$$

Así, obtenemos que:

$$\text{sen}^2(\alpha) = \frac{4 - k^2}{3}.$$

Por lo tanto,

$$\text{sen}(\alpha) = \sqrt{\frac{4 - k^2}{3}}.$$

Dado que el índice de refracción del agua es  $k \approx \frac{4}{3}$ , tendremos la siguiente igualdad:

$$\text{sen}(\alpha) = 0.86,$$

la cual nos permite obtener el valor del ángulo de incidencia  $\alpha$ , dado por:

$$\alpha = \arcsen(0.86)$$

$$\therefore \alpha = 59.4^\circ.$$

Si sustituimos el valor del ángulo de incidencia  $\alpha$ , en la ecuación (A.12) se obtiene que el ángulo de refracción  $\beta$ , está dado por:

$$\beta = \arcsen\left(\frac{\sen(59.4^\circ)}{k}\right)$$

$$\therefore \beta = 40.21^\circ$$

Por último, al sustituir el valor de  $\alpha$  y  $\beta$  en la función de desviación determinada por la ecuación (A.11) se obtiene el valor de la desviación mínima  $D(\alpha)$  :

$$D(\alpha) = \pi + 2(59.4) - 4(40.21)$$

$$\therefore D(\alpha) \approx 138^\circ.$$

El significado del resultado anterior nos indica que la desviación mínima del rayo de luz solar se da cuando el valor del ángulo de incidencia es  $\alpha \approx 59.4^\circ$ , esto es, que el cambio de la desviación  $D(\alpha)$  con respecto al ángulo con el que entra el rayo de luz en la gota de lluvia es aproximadamente cero. Esto nos indica que muchos rayos de luz solar con un ángulo de incidencia de  $59.4^\circ$ , resultan ser desviados en más o menos la misma cantidad. La concentración de los rayos de luz solar que provienen de las cercanías de la desviación mínima  $D(\alpha)$ , crea la brillantez del arco iris primario. Una duda que surge con el estudio de la aparición de arcos iris en el cielo, es de dónde provienen sus colores. Según el físico inglés Isaac Newton, en su experimento del prisma, la luz solar comprende una gama de longitudes de onda de color rojo, naranja, amarillo, verde, azul, índigo y violeta. En el año 1666, Newton usó prismas triangulares por los cuales hacía pasar luz blanca a través de una de las caras del prisma, la cual se refracta<sup>6</sup> hasta descomponerse en los colores básicos. Dichos prismas actúan de este modo debido a que la luz cambia de velocidad cuando pasa del aire al cristal del prisma. Una vez separados los colores, Newton descubrió que el índice de refracción es diferente para cada color, a este efecto se le conoce como dispersión.

---

<sup>6</sup>Refracción: Es el cambio de dirección de un rayo de luz u otra radiación que se produce al pasar oblicuamente de un medio a otro de distinta densidad.

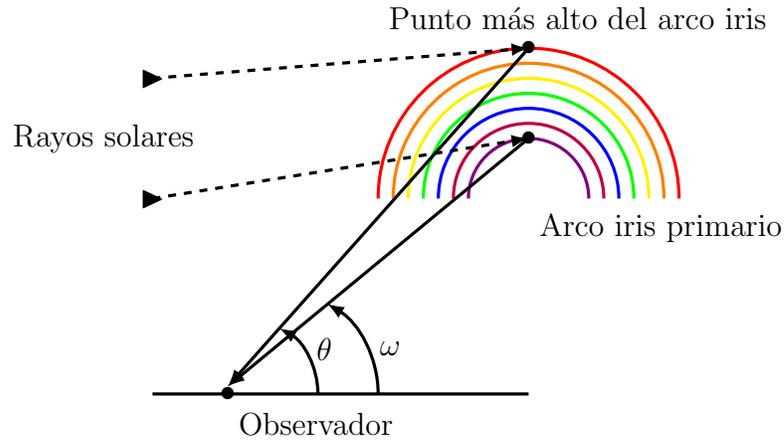


Figura A.18: Ángulo del arco iris.

En la Figura A.18 se observa el ángulo de elevación desde el observador hasta el punto más alto del arco iris, dicho ángulo es llamado ángulo del arco iris y se define como:

$$\pi - D(\alpha). \quad (\text{A.13})$$

Dado que el rayo de luz solar se descompone en siete colores básicos, se tendrá que cada color formará un arco, el cual tendrá un diferente ángulo de refracción y en consecuencia, cada arco tendrá un ángulo de arco iris diferente. Esto último se debe a que cada color de luz que compone la gama de longitudes de onda del rayo solar, tiene un diferente índice de refracción, en el caso de la luz roja, la cual refracta el primer arco del arco iris primario, tiene por índice de refracción  $k \approx 1.3318$  y para la luz violeta que forma el último arco, tiene por índice de refracción  $k \approx 1.3435$ . Al realizar los mismos cálculos que en la primera parte de este ejemplo, tendremos que el ángulo de incidencia para el índice de refracción de la luz roja y la luz violeta varían muy poco o casi nada en comparación con el índice de refracción del agua. Al hacer los cálculos, se tiene que para la luz roja:

- $\alpha = \arcsen\left(\sqrt{\frac{4-(1.3318)^2}{3}}\right) = 59.48^\circ$ .
- $\beta = \arcsen\left(\frac{\text{sen}(59.48^\circ)}{1.3318}\right) = 40.30^\circ$ .
- $D(\alpha) = \pi - 2(59.48) - 4(40.30) = 137.76^\circ$ .

En consecuencia, el ángulo del arco iris es:

$$\theta = 180^\circ - 137.76^\circ = 42.24^\circ.$$

Esto nos indica que el ángulo del arco iris para el arco de luz roja es alrededor de  $42.3^\circ$ , el cual es el ángulo de elevación desde el observador, hacia el arco de luz roja del arco iris.

Realizamos los mismos cálculos para el arco de luz violeta:

- $\alpha = \arcsen\left(\sqrt{\frac{4-(1.3435)^2}{3}}\right) = 58.80^\circ.$
- $\beta = \arcsen\left(\frac{\text{sen}(58.80^\circ)}{1.3435}\right) = 39.54^\circ.$
- $D(\alpha) = \pi - 2(58.80) - 4(39.54) = 139.44^\circ.$

De aquí, se tiene que:

$$\omega = 180^\circ - 139.44^\circ = 40.56^\circ.$$

Luego, tendremos que el ángulo del arco iris para el arco de luz violeta es alrededor de  $40.6^\circ$ , el cual es el ángulo de elevación del observador hacia al arco de luz violeta del arco iris primario. Por lo tanto, se puede decir que el arco iris en realidad, consta de siete arcos separados que corresponden a los siete colores básicos en los que se dispersa la luz del rayo solar que entra en las gotas de lluvia esféricas.

Normalmente al ver un arco iris primario en el cielo se puede observar arriba de este, otro arco iris más tenue, dicho arco iris es llamado arco iris secundario. La formación del arco iris secundario se produce cuando el rayo solar entra en la gota de lluvia esférica y se refleja dos veces antes de salir de ella. Esto es, el rayo solar entra la gota de lluvia y se refracta en el punto *A*, luego se refleja en el punto *B* y luego en el punto *C*. Por último, el rayo solar se refracta en el punto *D* al salir de la gota.

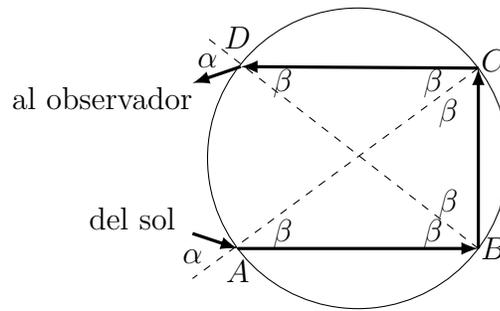


Figura A.19: Formación del arco iris secundario.

En la Figura A.19 se puede observar el movimiento de un rayo solar al entrar en una gota de agua. Al contrario que para el arco iris primario, se tiene que  $D(\alpha)$  es la magnitud total de la rotación en sentido contrario al movimiento de las manecillas del reloj que describe el rayo al entrar a la gota. Notemos que el rayo incidente cambia de dirección un ángulo  $(\alpha - \beta)$ . Al entrar en el interior de la gota el rayo solar incidente se refleja dos veces en la superficie de la gota, cambiando de dirección en cada reflexión  $(\pi - 2\beta)$ . Al salir de la gota el rayo solar cambia de dirección un ángulo  $\alpha - \beta$ . Este trayecto de cuatro etapas es descrito por la siguiente ecuación:

$$\begin{aligned} D(\alpha) &= (\alpha - \beta) + (\pi - 2\beta) + (\pi - 2\beta) + (\alpha - \beta) \\ &= 2(\alpha - \beta) + 2(\pi - 2\beta) \\ &= 2\alpha - 2\beta + 2\pi - 4\beta. \end{aligned}$$

Por lo tanto, el ángulo de desviación  $D(\alpha)$ , se define como:

$$D(\alpha) = 2\alpha - 6\beta + 2\pi. \quad (\text{A.14})$$

Al sustituir el valor de  $\beta$ , dado por la ecuación (A.12) en la ecuación anterior tendremos lo siguiente:

$$D(\alpha) = 2\alpha + 2\pi - 6 \arcsen\left(\frac{\text{sen}(\alpha)}{k}\right).$$

Al realizar la derivada de  $D(\alpha)$  con respecto al ángulo de incidencia  $\alpha$ , tendremos

que:

$$\begin{aligned}\frac{d}{d\alpha}D(\alpha) &= \frac{d}{d\alpha} \left( 2\alpha + 2\pi - 6 \operatorname{arcsen} \left( \frac{\operatorname{sen}(\alpha)}{k} \right) \right) \\ &= 2 - \frac{6 \cos(\alpha)}{\sqrt{k^2 - \operatorname{sen}^2(\alpha)}}.\end{aligned}$$

Al simplificar se obtiene:

$$\frac{d}{d\alpha}D(\alpha) = \frac{2\sqrt{k^2 - \operatorname{sen}^2(\alpha)} - 6 \cos(\alpha)}{\sqrt{k^2 - \operatorname{sen}^2(\alpha)}}.$$

Notemos que para cualquier valor del ángulo de incidencia  $\alpha$ , siempre se tiene que  $\sqrt{k^2 - \operatorname{sen}^2(\alpha)} > 0$ . Por lo que, la igualdad:

$$\frac{d}{d\alpha}D(\alpha) = 0,$$

se cumplirá, si y solo si,

$$\begin{aligned}2\sqrt{k^2 - \operatorname{sen}^2(\alpha)} - 6 \cos(\alpha) &= 0, \\ \Rightarrow \sqrt{k^2 - \operatorname{sen}^2(\alpha)} &= 3 \cos(\alpha), \\ \Rightarrow k^2 - \operatorname{sen}^2(\alpha) &= 9 \cos^2(\alpha), \\ \Rightarrow k^2 &= 9 \cos^2(\alpha) + \operatorname{sen}^2(\alpha).\end{aligned}$$

Usando la identidad trigonométrica:

$$\operatorname{sen}^2(\alpha) = 1 - \cos^2(\alpha),$$

y sustituyendo en la última ecuación obtenemos:

$$\begin{aligned}k^2 &= 9 \cos^2(\alpha) + 1 - \cos^2(\alpha), \\ \Rightarrow k^2 &= 8 \cos^2(\alpha) + 1, \\ \Rightarrow \frac{k^2 - 1}{8} &= \cos^2(\alpha).\end{aligned}$$

Por lo tanto, la desviación mínima o el valor mínimo de  $D(\alpha)$ , sucede cuando se cumple que:

$$\cos(\alpha) = \sqrt{\frac{k^2 - 1}{8}}.$$

Al sustituir el índice de refracción para el agua  $k \approx \frac{4}{3}$ , podemos obtener el valor del ángulo de incidencia cuando ocurre la desviación mínima, como:

$$\begin{aligned} \cos(\alpha) &= \sqrt{\frac{\left(\frac{4}{3}\right)^2 - 1}{8}} = 0.31, \\ \Rightarrow \alpha &= \arccos(0.31), \\ \therefore \alpha &= 71.83^\circ \end{aligned}$$

Por otro lado, si sustituimos el valor de  $\alpha$  en la ecuación (A.12) obtendremos el valor mínimo del ángulo de refracción  $\beta$ , el cual se obtiene como:

$$\begin{aligned} \beta &= \arcsen\left(\frac{\text{sen}(71.83^\circ)}{1.33}\right), \\ \therefore \beta &= 45.59^\circ. \end{aligned}$$

Sustituyendo el valor de  $\alpha$  y  $\beta$  en el ángulo de desviación  $D(\alpha)$ , tendremos lo siguiente:

$$\begin{aligned} D(\alpha) &= 2(71.83^\circ) - 6(45.59^\circ) + 2\pi, \\ \therefore D(\alpha) &= 230.12^\circ. \end{aligned}$$

Recordemos que el ángulo de desviación  $D(\alpha)$  es la magnitud total de la rotación en sentido contrario al movimiento de las manecillas del reloj, el cual describe el rayo al entrar en la gota de agua. Por lo que podemos decir que el ángulo de desviación mínima es:

$$360^\circ - 230.12^\circ = 129.88^\circ,$$

así tendremos que la desviación mínima es alrededor de  $130^\circ$ . Por lo tanto, al sustituir el valor del ángulo de desviación mínima en la ecuación (A.14) tendremos que el ángulo del arco iris secundario es:

$$\zeta \approx 50^\circ.$$

En la Figura A.20 se observa el ángulo del arco iris primario  $\theta$  y el ángulo del arco iris secundario  $\zeta$ .

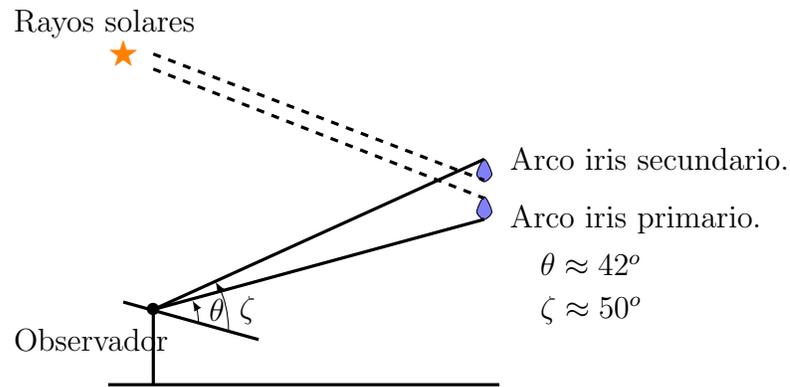


Figura A.20: Ángulo del arco iris primario y secundario.

En la Figura A.19 se observa el trayecto del rayo al entrar en el interior de la gota después de una reflexión y la trayectoria del rayo después de dos reflexiones. El arco iris secundario es formado por la reflexión doble del rayo solar al entrar en la gota, esto provoca que los colores del arco iris secundario sean más tenues y estén en el orden inverso a los colores del arco iris primario. Es decir, en el arco de luz color violeta se encuentra el punto más alto del arco iris.

En general, el rayo solar incidente en el interior de la gota se puede reflejar  $n$ -veces en la superficie de la gota, cambiando de dirección en cada reflexión  $(\pi - 2\beta)$ . Además, siempre se cumplirá que el ángulo incidente cambia de dirección un ángulo  $\alpha - \beta$ , el cual es igual al cambio de dirección que tiene el rayo cuando sale de la gota. Así, podemos escribir el ángulo de desviación  $D(\alpha)$  como:

$$D(\alpha) = 2\alpha - n(\pi - 2\beta) + (\alpha - \beta).$$

Por lo tanto, el ángulo de desviación  $D(\alpha)$  tiene como expresión general:

$$D(\alpha) = n\pi + 2\alpha - 2(n + 1)\beta; \tag{A.15}$$

donde  $n$  es el número de veces que el rayo solar se refleja en el interior de la gota.

## Reflexión interna total

Un efecto interesante denominado reflexión interna total se presenta al dirigir luz desde un medio con índice de refracción determinado hacia otro que tenga un índice de refracción menor. Considere la Figura A.21 en la cual un rayo de luz se

desplaza en el medio 1 y se encuentra la frontera entre el medio 1 y el medio 2, donde  $n_1$  es mayor que  $n_2$ . Los niveles 1 al 5 indican varias posibles direcciones del rayo, consistente con el modelo de una onda bajo refracción. Los rayos refractados están doblados alejándose de la normal porque  $n_1$  es mayor que  $n_2$ .

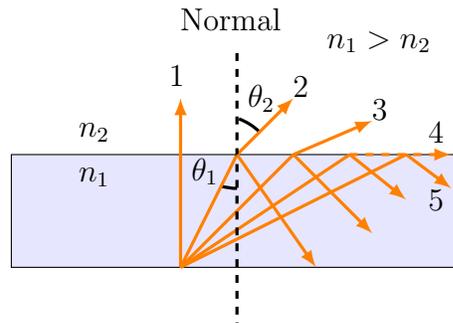


Figura A.21: Los rayos se desplazan de un medio de índice de refracción  $n_1$  hacia un medio de índice de refracción  $n_2$ , donde  $n_2 < n_1$ . Cuando el ángulo de incidencia aumenta  $\theta_1$ , el ángulo de refracción  $\theta_2$  aumenta hasta que  $\theta_2$  es de  $90^\circ$  (rayo 4). La línea discontinua indica que no hay energía que se propague en esta dirección. Para ángulos de incidencia incluso mayores, se presenta reflexión interna total (rayo 5).

En algún ángulo particular de incidencia  $\theta_c$ , denominado ángulo crítico, el rayo de luz refractado se mueve paralelo a la frontera, de modo que  $\theta_2 = 90^\circ$ , como se observa en la Figura A.22. Para ángulos de incidencia mayores a  $\theta_c$ , el rayo se refleja por completo en la frontera, como lo muestra la Figura A.21. Usemos la ley de refracción de Snell para hallar el valor del ángulo crítico.

Cuando  $\theta_1 = \theta_c$ ,  $\theta_2 = 90^\circ$  y la ecuación (A.9) da:

$$n_1 \text{ sen } \theta_c = n_2 \text{ sen } (90^\circ) = n_2$$

$$\text{sen } \theta_c = \frac{n_2}{n_1} \quad (\text{para } n_1 > n_2) \quad (\text{A.16})$$

Esta ecuación solo se usa cuando  $n_1$  es mayor que  $n_2$ . Es decir, la reflexión interna total se presenta solo cuando la luz se dirige de un medio de índice de refracción conocido hacia un medio de índice de refracción menor. Si  $n_1 < n_2$ , la ecuación (A.16) daría  $\text{sen } \theta_c > 1$ ; el cual es un resultado sin sentido porque el seno de un ángulo nunca puede ser mayor a la unidad.

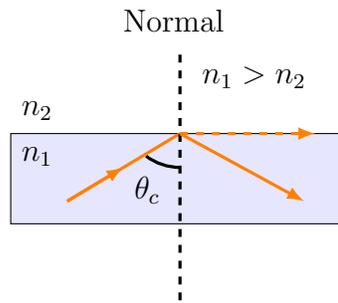


Figura A.22: El ángulo de incidencia que produce un ángulo de refracción igual a  $90^\circ$  es el ángulo crítico  $\theta_c$ , con este ángulo de incidencia se refleja toda la energía de la luz incidente.

El ángulo crítico para la reflexión interna total es pequeño cuando  $n_1$  es considerablemente mayor a  $n_2$ . Por ejemplo, el ángulo crítico para un diamante en el aire es de  $24^\circ$ . Cualquier rayo dentro del diamante que se aproxime a la superficie a un ángulo mayor de  $24^\circ$  se refleja por completo de regreso hacia el interior del cristal. Esta propiedad, combinada con un correcto tallado de caras, hace que el diamante brille. Los ángulos de las caras se cortan de modo que la luz sea atrapada dentro del cristal por medio de múltiples reflexiones internas. Estas dan a la luz una trayectoria larga en el medio, y se presenta una dispersión de colores considerable. En el momento en el que la luz sale por la superficie superior del cristal, los rayos asociados con diferentes colores ya han sido ampliamente separados unos de otros.

# Bibliografía

- [1] Acebo González, D. & Hernández García, Armando T. (2013). *Los métodos turbidimétricos y sus aplicaciones en las ciencias de la vida*. Revista CENIC. Ciencias Biológicas, vol. 44. Ciudad de la Habana, Cuba. Disponible en: <http://www.redalyc.org/articulo.oa?id=181226886003>
- [2] Asmar Charris, Abraham J. (1995). *Tópicos en Teoría de Matrices*. Universidad Nacional de Colombia, Medellín.
- [3] Bohren, C. & Huffman, D. (1983). *Absorption and Scattering of Light by Small Particles*. Wiley, New York.
- [4] Bunge, M. (2006). *Problemas directos e inversos*. Grupo Bunge. <http://grupobunge.wordpress.com>
- [5] Burden, R.L. & Douglas Faires, J. (1985). *Análisis Numérico*. Grupo Editorial Iberoamericana. México.
- [6] Cervantes, M. *Clasificación de la materia*. Andalucía, España. Disponible en: [www.juntadeandalucia.es](http://www.juntadeandalucia.es)
- [7] Clementi, Luis A. (2011). *Caracterización de partículas nanométricas por técnicas de dispersión de luz. Estimación de distribuciones de tamaños en látex poliméricos*. Tesis para obtener el grado académico de Doctor en Tecnología Química. Instituto de Desarrollo Tecnológico para la Industria Química (INTEC) dependiente de la Universidad Nacional del Litoral. Facultad de Ingeniería Química. Argentina.
- [8] Demmel, J. (1997). *Applied Numerical Linear Algebra*. SIAM, Philadelphia.
- [9] Detección de endotoxinas mediante el ensayo de LAL, método turbidimétrico. North Chesterfield FUJIFILM Wako Chemical U.S.A Corporation (2014). Disponible en: [www.wakopyrostar.com](http://www.wakopyrostar.com)

- [10] *Equipos y Laboratorios de Colombia, Artículos*. Consultado: 05/09/2022 a las 17:07 pm en el sitio web: <https://www.equiposylaboratorio.com/portal/articulo-ampliado/quE-es-y-usos-del-espectrofotometro>
- [11] Friedberg, Stephen H.; Insel, Arnold J. & Spence, Lawrence E. (1982) *Algebra Lineal*. Illinois State University.
- [12] Galpin, T.; Chartier, R.; Levergood, N. & Greenslade, Margaret E. (2017). *Refractive index retrievals for polystyrene latex spheres in the spectral range 220-420 nm*. Aerosol Science and Technology.
- [13] Giménez, M. (2019). *A Guidebook to Particle Size Analysis*. Horiba Scientific. Disponible en: [www.horiba.com/us/particle](http://www.horiba.com/us/particle)
- [14] González Garcés, J. (2016). *Análisis de la distribución de tamaños de partícula mediante difracción láser: aplicación para la mejora de mezclas de especias y aromas*. Trabajo de fin de grado de Ingeniería en Química. Universidad Politécnica de Valencia, España.
- [15] Groetsch, C. (2001). *Inverse problems: the other two-thirds of the story*. Questions Mathematical, Supplementary Issue.
- [16] Grossman, Stanley I. *Álgebra Lineal*. Sexta Edición. Ma Graw Hill.
- [17] Hecht, E. (2000). *Optics*. Fifth edition. Adelphi University.
- [18] Izaguirre, María F.; Clementi, Luis A.; Vega, Jorge R. & Gugliotta, Luis M. (2012). *Estimación de la distribución de tamaños de partícula por turbidimetría. Comparación de tres técnicas de resolución del problema inverso*. Asociación Argentina de Mecánica Computacional, Argentina.
- [19] Kerker, M. (1969). *The Scattering of Light and other Electro-Magnetic Radiation*. Academic Press, New York.
- [20] Kirsch, A. (1996). *An Introduction to the Mathematical Theory of Inverse Problems*. Second edition. Springer-Verlag, New York.
- [21] Kourti, T. *Polymer Latexes: Production by Homogeneous Nucleation and Methods for Particle Size Determination*. A thesis submitted to the school of graduate studies in partial fulfilment of the requirements for the degree Doctor of Philosophy. McMaster University. July 1989.

- [22] Laven, P. *The optics of a water drop: Mie scattering and the Debye series*.
- [23] Lines, Claire D. (2003). *Inverse Scattering by Unbounded Rough Surfaces*. Tesis presentada para el grado de Doctor en Filosofía. Brunel University.
- [24] Lloset, Mario A.; Gugliotta, Luis M. & Meira, Gregorio R. *Particle Size Distribution of SBR and NBR Latexes by UV-VIS Turbidimetry Near the Rayleigh Region*. CONICET y Universidad Nacional del Litoral. Santa Fe, Argentina.
- [25] Lo Giudice Fernández, E.; Hernández Sanchez, M. & Abad Hernández, J. (2012). *Solución del problema inverso de dispersión angular de luz para partículas de látex*. Instituto Nacional de Investigaciones en Metrología. Ciudad de la Habana, Cuba.
- [26] López Camino, R. (2003) *Aplicaciones Lineales*. Departamento de Geometría y Topología. Universidad de Granada.
- [27] Luenberger, David G. (1969). *Optimization by Vector Space Methods*. Wiley Professional Paperback Series, Canada.
- [28] Mancilla Aguilar, J.L. *Descomposición en valores singulares*. Notas para los cursos 21 y 22.
- [29] *Matrices y Determinantes: Rango de una matriz*. Universidad Europea de Madrid. Sitio web: Cartagena99. Disponible en : [https://www.cartagena99.com/recursos/alumnos/apuntes/DYCRE\\_1\\_T1\\_MaEm\\_U5L03.pdf](https://www.cartagena99.com/recursos/alumnos/apuntes/DYCRE_1_T1_MaEm_U5L03.pdf)
- [30] Mazo Vivar, A. *Esparcimiento de Rayleigh*. Revista Eureka sobre Enseñanza y Divulgación de las Ciencias, vol.13, núm. 2, abril, 2016, pp. 505-510. Asociación de Profesores Amigos de la Ciencia: EUREKA. Cádiz, España.
- [31] Méndez Cruz, I. (2017). *Problemas Lineales Discretos Mal Planteados y sus Aplicaciones en Restauración de Imágenes Digitales*. Tesis para optar por el grado de Maestro en Ciencias. Universidad Nacional Autónoma de México, México.
- [32] Nakos G. & Joyner D. (1999) *Álgebra Lineal con Aplicaciones*. International Thomson Editores.

- [33] Otero, Fernando A. (2012). *Caracterización de Sistemas de Partículas por Dispersión de Luz. Estimación de Parámetros y Problema Inverso*. Tesis para optar el grado de doctor en Ingeniería de Orientación Electrónica. Universidad Nacional del Mar de Plata, Argentina.
- [34] Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill, Inc.
- [35] Salas, J.; Torrente, A. & Villaseñor, E.J. *Pseudo Inversa y descomposición en valores singulares*. Álgebra Lineal, tema 14. Universidad Carlos III de Madrid.
- [36] San Fabián, E. (2021) *Química Cuántica y Espectroscopía*. Universidad de Alicante, Departamento de Química y Física.
- [37] Serway, Raymond A. & Jewett, John W. Jr. *Física para ciencias e ingeniería con Física moderna*. Vol.2. Séptima Edición. CENGAGE Learning.
- [38] Schiebener, P.; Straub, J.; Levelt Sengers, J.H. & Gallagher, J.S. (1990). *Journal of Physical and Chemical Reference*. Disponible en:<https://doi.org/10.1063/1.555859>
- [39] Skiba, Yuri N. *Número de condición de una matriz y métodos de su evaluación*. Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México.
- [40] Thome, N. (2019). *La inversa generalizada de Moore-Penrose y Aplicaciones*. Publicaciones electrónicas. Sociedad Matemática Mexicana. Vol 21.
- [41] Toledo, M. *Espectroscopía ultravioleta-visible: conceptos básicos*. España.
- [42] Van de Geijn, R. & Myers, M. *Advanced Linear Algebra: Foundations to Frontiers*. Consultado: 12/09/2022 a las 14:35 en el sitio web: <https://www.cs.utexas.edu/users/flame/laff/alaff/chapter01-induced-matrix-norms.html>
- [43] Vargas Ubera, J. (2007). *Métodos numéricos de inversión en el problema de estimación de tamaño de partículas a partir de difracción de luz láser*. Tesis para optar por el grado de Doctor en Óptica. Instituto Nacional de Astrofísica, Óptica y Electrónica. Tonantzintla, Puebla.
- [44] White, Hearvey E. (1965). *Física Moderna*. Quinta Edición. UTEHA.

- [45] Yamamoto, G. & Tanaka, M. (1969). *Determination of Aerosol Size Distribution from Spectral Attenuation Measurements*. Applied Optics (vol.8).
- [46] Zueco Jordán, J. (2003). *Solución de problemas inversos en conducción de calor mediante el método de simulación por redes*. Tesis para optar por el grado de Doctor en Física Aplicada. Universidad Politécnica de Cartagena. Cartagena, Colombia.