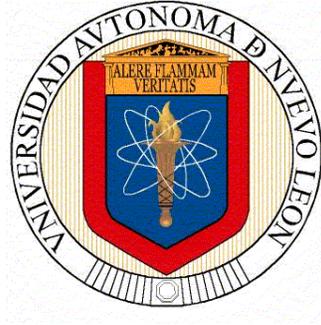


**UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN**  
**FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS**



**TESIS**

**INTEGRACIÓN DE TÉCNICAS DE INTELIGENCIA  
ARTIFICIAL EN LA CALIBRACIÓN DE ABERRACIONES EN  
FRENTE DE ONDA MEDIANTE EL MÉTODO SHACK-HARTMANN**

**PRESENTADA POR  
MARIO ALBERTO JARA REYNA**

**EN OPCIÓN AL GRADO DE  
MAESTRÍA EN ASTROFÍSICA  
PLANETARIA Y TECNOLOGÍAS AFINES**

**2025**

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS



---

**FCFM**

**Integración de Técnicas de Inteligencia  
Artificial en la Calibración de  
Aberraciones en Frente de Onda  
mediante el Método Shack-Hartmann**

*Por*

**Mario Alberto Jara Reyna**

*Tutores*

Dr. Carlos Alberto Guerrero Peña

Dr. Joel Herrera Vázquez

Como requisito para obtener el grado de Maestría en  
Astrofísica Planetaria y Tecnologías Afines

Pedro de Alba, Niños Héroes, Ciudad Universitaria, 66451 San  
Nicolás de los Garza, N.L

*A mis padres, por inculcarme el valor de la educación. Todo lo hago con dedicación para ellos.*

*Al Dr. Carlos Alberto Guerrero Peña, por su constante disposición para atender mis dudas, así como por la paciencia, el tiempo y el conocimiento que me brindó durante este trabajo.*

*Al Dr. Joel Herrera Vázquez, por su disponibilidad, conocimientos e ideas que enriquecieron el desarrollo de esta investigación.*

# Índice general

<b>Resumen</b>	<b>5</b>
<b>Antecedentes</b>	<b>6</b>
<b>1. Introducción</b>	<b>8</b>
<b>2. Planteamiento del Problema</b>	<b>11</b>
2.1. Descripción del Problema . . . . .	11
2.2. Preguntas de Investigación . . . . .	11
2.3. Hipótesis . . . . .	12
2.4. Objetivo General y Objetivos Específicos . . . . .	12
2.5. Justificación del Proyecto . . . . .	13
2.6. Aportación Científica y Tecnológica . . . . .	14
2.7. Delimitaciones . . . . .	14
<b>3. Fundamentos Ópticos</b>	<b>16</b>
3.1. Reflexión y Refracción . . . . .	16
3.2. Reflexión y Refracción . . . . .	16
3.2.1. Ley de la Reflexión . . . . .	17
3.2.2. Ley de la Refracción . . . . .	17
3.3. Lentes y Espejos Simples . . . . .	19
3.3.1. Lentes Convergentes . . . . .	19
3.3.2. Lentes Divergentes . . . . .	20
3.3.3. Espejos Ópticos . . . . .	22
3.4. Formación de Imágenes . . . . .	22
3.5. Sistemas Ópticos . . . . .	24
3.5.1. Matrices de Sistemas Ópticos . . . . .	24
3.5.2. Trazado de Rayos . . . . .	25
3.5.2.1. Trazado de Rayos Matricial . . . . .	25
3.5.2.2. Trazado de Rayos Exacto . . . . .	26

3.6.	Características de los Sistemas Ópticos . . . . .	26
3.6.1.	Pupila del Sistema . . . . .	26
3.6.2.	Aberraciones de sistemas ópticos . . . . .	27
3.6.3.	Superficies Cónicas . . . . .	30
3.7.	Polinomios de Zernike . . . . .	32
3.7.1.	Notación de Noll . . . . .	33
3.8.	Telescopios . . . . .	35
3.8.1.	Telescopios Refractores . . . . .	35
3.8.2.	Telescopios Reflectores . . . . .	35
<b>4.</b>	<b>Análisis y Ajuste del Frente de Onda</b>	<b>38</b>
4.1.	Sensor Shack-Hartmann . . . . .	38
4.2.	Aberración Transversal y la Derivada del Frente de Onda . . . . .	40
4.3.	Ajuste del Frente de Onda . . . . .	41
4.4.	Sistema de Ecuaciones para el Ajuste del Frente de Onda . . . . .	42
4.5.	Mínimos Cuadrados . . . . .	44
4.6.	Gradiente Descendente con Optimizador Adam . . . . .	45
4.7.	Red Hiperdinámica . . . . .	49
<b>5.</b>	<b>Herramientas Avanzadas para la Selección de Hiperparámetros</b>	<b>54</b>
5.1.	Importancia de los Hiperparámetros . . . . .	55
5.1.1.	Hiperparámetros del Gradiente Descendente . . . . .	56
5.1.2.	Hiperparámetros de la Red Hiperdinámica . . . . .	57
5.2.	Definición del Espacio de Búsqueda . . . . .	58
5.2.1.	Selección del Espacio de Búsqueda para el Gradiente Descendente . . . . .	59
5.2.2.	Selección del Espacio de Búsqueda para la Red Hiperdinámica . . . . .	59
5.3.	Ajuste de Hiperparámetros mediante Optimización Bayesiana . . . . .	60
5.4.	Métricas de Evaluación . . . . .	62
5.5.	Visualización y Análisis con Corner Plots . . . . .	64
<b>6.</b>	<b>Generación y Procesamiento de Datos</b>	<b>66</b>
6.1.	Procesamiento de Imágenes . . . . .	67
6.2.	Interpolación y Ajuste de Puntos Faltantes . . . . .	69
6.2.1.	Ejemplo de Interpolación y Ajuste de Puntos Faltantes . . . . .	71
6.2.2.	Evaluación del Método con Diferentes Porcentajes de Puntos Faltantes . . . . .	73
6.3.	Introducción de Ruido en las Medidas . . . . .	74
6.3.1.	Rango de Respuesta Lineal del Sensor . . . . .	76

6.3.2. Análisis Estadístico de las Aberraciones . . . . .	77
<b>7. Entrenamiento y Configuración de Modelos</b>	<b>81</b>
7.1. Entrenamiento Mínimos Cuadrados . . . . .	82
7.2. Gradiente Descendente para Una Sola Muestra . . . . .	82
7.3. Gradiente Descendente para Múltiples Muestras . . . . .	88
7.4. Entrenamiento del Modelo de Red Hiperdinámica . . . . .	92
<b>8. Evaluación del Desempeño y Discusión</b>	<b>98</b>
8.1. Evaluación Comparativa de Desempeño . . . . .	99
8.1.1. Resumen de Resultados . . . . .	99
8.1.2. Análisis en Presencia de Valores Atípicos . . . . .	102
8.1.3. Análisis en el Conjunto de Pruebas . . . . .	103
<b>9. Conclusiones Generales</b>	<b>105</b>
<b>A. Apéndice</b>	<b>111</b>
A.1. Artículo científico resultante de esta investigación . . . . .	111
A.1.1. Introducción . . . . .	111
A.1.2. Manuscrito publicado en la revista <i>Optical Engineering</i>	112
A.2. Estadísticas Detalladas de las Aberraciones . . . . .	135

## Resumen

El presente trabajo desarrolla una metodología para el ajuste de aberraciones en el frente de onda y evalúa distintas técnicas de ajuste en escenarios con una muestra real y múltiples muestras sintéticas. Se comparan tres enfoques principales: el método de mínimos cuadrados, el gradiente descendente y la red hiperdinámica. Cada uno de estos métodos se analiza en términos de precisión, adaptabilidad y costo computacional.

El método de mínimos cuadrados puede ser sensible a valores atípicos y, debido a su dependencia de la pseudoinversa, su precisión puede verse afectada en condiciones de mediciones incompletas o ambiguas, lo que podría llevar a modelar aberraciones inexistentes. Por su parte, el gradiente descendente se presenta como una alternativa adaptable y de bajo costo computacional, siendo adecuado para escenarios donde la flexibilidad es prioritaria, aunque con una precisión potencialmente menor.

La red hiperdinámica emerge como el enfoque más preciso para el ajuste de frentes de onda, minimizando errores en escenarios complejos, como aquellos con puntos atípicos. Sin embargo, su implementación conlleva un mayor costo computacional debido a la complejidad de su estructura y proceso de entrenamiento.

Además de la precisión, este estudio considera la importancia del tiempo de procesamiento en aplicaciones donde la turbulencia atmosférica introduce variaciones rápidas en el frente de onda. Dado que estas fluctuaciones ocurren en escalas de tiempo del orden de milisegundos, es fundamental evaluar la capacidad de cada método para adaptarse dinámicamente a estos cambios y capturar la evolución temporal de las aberraciones.

Los resultados obtenidos permiten analizar las ventajas y limitaciones de cada metodología, proporcionando una evaluación detallada y abriendo nuevas perspectivas para la investigación en ajuste de frentes de onda. Este estudio resalta el potencial de integrar enfoques avanzados que optimicen tanto la precisión como la adaptabilidad en aplicaciones específicas.

# Antecedentes

Históricamente, el ajuste de aberraciones ópticas se ha abordado mediante diversas técnicas, cada una con ventajas y limitaciones específicas.

La **prueba de la navaja** o *Foucault Test*, desarrollada por Léon Foucault en 1858, es uno de los métodos más antiguos para evaluar la calidad óptica de espejos parabólicos. Este método consiste en mover una cuchilla a través del haz de luz reflejado y observar las sombras formadas, las cuales permiten inferir la forma del frente de onda (Malacara, 2007). Aunque su simplicidad es una ventaja, carece de la precisión requerida para aplicaciones modernas.

En la década de 1920, Vasco Ronchi introdujo la **prueba de Ronchi**, que emplea una retícula de líneas finas para detectar aberraciones ópticas. Este método ofreció un enfoque más estructurado que el de Foucault, ya que proporciona patrones visuales más detallados que reflejan la forma del frente de onda (Malacara, 2007). Sin embargo, su resolución sigue siendo limitada para aberraciones complejas.

Los **interferómetros**, como los de Fizeau y Michelson, marcaron un hito importante en el análisis óptico a finales del siglo XIX. El interferómetro de Fizeau, presentado en 1862, utiliza patrones de interferencia para medir aberraciones con alta precisión. Posteriormente, el interferómetro de Michelson, desarrollado en 1881, mejoró este diseño al introducir una mayor sensibilidad para evaluar superficies ópticas (Malacara, 2007). A pesar de su alta precisión, ambos métodos requieren condiciones controladas y son más costosos.

En 1900, Karl Hartmann desarrolló las **pantallas de Hartmann**, que introdujeron un enfoque práctico para medir aberraciones ópticas mediante una placa perforada que proyecta puntos de luz en un detector. Las desviaciones de estos puntos se emplean para calcular las aberraciones del frente de onda (Malacara, 2007). Aunque práctico, este método tiene limitaciones en términos de resolución .

En 1971, Roland Shack y Ben Platt presentaron el **sensor Shack-Hartmann**,

introdujeron el uso de una matriz de microlentes en lugar de la máscara perforada original (Shack, 1971). Este diseño permitió descomponer el frente de onda en subfrentes casi planos, cuya información fue procesada mediante metodologías matemáticas avanzadas para analizar y caracterizar las aberraciones ópticas. Este enfoque innovador incrementó significativamente la precisión y resolución en el análisis del frente de onda, estableciendo las bases para aplicaciones modernas de óptica.

Este trabajo de tesis propone una metodología innovadora basada en el uso de técnicas de inteligencia artificial para mejorar el ajuste de aberraciones ópticas. El objetivo es optimizar tanto la precisión en el cálculo de los coeficientes del frente de onda como analizar posibles beneficios en escenarios específicos, evaluando el impacto de la metodología en métricas relevantes para sistemas ópticos.

# Capítulo 1

## Introducción

El ajuste preciso de aberraciones en el frente de onda es un problema fundamental en sistemas ópticos avanzados, como telescopios adaptativos (Rayces, 1964a). En este trabajo, el término *ajuste* se usará para describir el proceso de estimación de los coeficientes que representan las aberraciones ópticas del frente de onda, a partir de los desplazamientos medidos con el sensor Shack-Hartmann. Esta definición será utilizada de forma consistente a lo largo del documento. Estas aberraciones, generadas por imperfecciones en los componentes ópticos o por efectos externos como la turbulencia atmosférica, afectan directamente la calidad de las imágenes obtenidas (Goel and Ganotra, 2024). La necesidad de desarrollar métodos eficientes y precisos para caracterizar y corregir estas aberraciones ha llevado a la investigación de técnicas avanzadas, incluyendo el uso de sensores como el Shack-Hartmann (SHWFS), y la aplicación de inteligencia artificial (IA) para optimizar el proceso (Liu et al., 2020) (Fowler and Landman, 2023a).

En este contexto, los sensores Shack-Hartmann desempeñan un papel crucial al proporcionar datos de desplazamiento que permiten caracterizar el frente de onda. Sin embargo, traducir estas mediciones en información útil para corregir aberraciones requiere la aplicación de algoritmos robustos y adaptables. Entre las estrategias tradicionales se encuentra el método de mínimos cuadrados, que ha sido ampliamente utilizado debido a su simplicidad y eficacia en escenarios controlados. Sin embargo, este método presenta limitaciones inherentes en casos donde existen datos faltantes o ruido significativo. Para superar estas deficiencias, se han explorado enfoques como el gradiente descendente, que reduce la dependencia de operaciones complejas como la pseudo-inversa. Más recientemente, las redes hiperdinámicas han emergido como una metodología innovadora, ofreciendo una capacidad sin precedentes

para manejar datos complejos (Wetzstein et al., 2020) (Lin et al., 2018).

Esta tesis se enmarca en la investigación y comparación de tres métodos: mínimos cuadrados, gradiente descendente y redes hiperdinámicas, aplicados al ajuste de aberraciones en el frente de onda. El objetivo principal es desarrollar una metodología para el ajuste de aberraciones y evaluar el desempeño de los métodos propuestos en términos de precisión, capacidad de adaptabilidad y tiempo computacional. Además, se analizan sus fortalezas y debilidades en escenarios prácticos.

Para ello, se emplean datos sintéticos generados a partir de datos obtenidos con un sensor SHWFS, incorporando condiciones controladas y ruido que represente la turbulencia atmosférica, para simular los desafíos que enfrentan los sistemas ópticos en aplicaciones reales.

El enfoque de esta investigación también incluye la implementación de estrategias avanzadas como la optimización bayesiana, utilizada para ajustar los hiperparámetros de los modelos y maximizar su desempeño. Además, se realiza un análisis exhaustivo de los residuales, que permite identificar patrones de error y evaluar la consistencia de cada método. Estos análisis no solo contribuyen a entender las limitaciones de cada aproximación, sino también a establecer directrices para futuros desarrollos en el campo.

El desarrollo de esta tesis se estructura en los siguientes capítulos. En el **Capítulo 3**, se establecen los fundamentos ópticos necesarios para comprender el problema del ajuste de aberraciones en el frente de onda. Se exploran principios clave de la óptica, como la propagación de la luz, aberraciones ópticas y el modelado de sistemas ópticos, proporcionando la base teórica para el análisis posterior.

En el **Capítulo 4**, se introduce el análisis y ajuste del frente de onda, describiendo en detalle el sensor Shack-Hartmann y su papel en la caracterización de aberraciones. Se presenta la relación entre las aberraciones transversales y las derivadas del frente de onda, además de los métodos utilizados para su ajuste, incluyendo la reconstrucción con polinomios de Zernike.

El **Capítulo 5** aborda la selección de hiperparámetros y la evaluación del desempeño de los métodos propuestos. Se describen las métricas utilizadas, el impacto de los hiperparámetros en la precisión de los modelos y la optimización bayesiana como estrategia para encontrar configuraciones óptimas. Además, se introduce el uso de *corner plots* para analizar las relaciones entre hiperparámetros.

En el **Capítulo 6**, se detalla la generación y procesamiento de los datos

utilizados en los experimentos. Se explican las técnicas de preprocesamiento de imágenes, la interpolación de datos faltantes y la generación de datos sintéticos mediante la incorporación de ruido controlado, con el objetivo de evaluar la robustez de los modelos frente a condiciones realistas.

El **Capítulo 7** está dedicado a la implementación y entrenamiento de los modelos de ajuste del frente de onda. Se presentan los tres enfoques estudiados: mínimos cuadrados como método de referencia, gradiente descendente optimizado con Adam y redes hiperdinámicas, detallando sus configuraciones y procedimientos de entrenamiento.

En el **Capítulo 8**, se realiza la evaluación del desempeño de los modelos. Se comparan sus resultados en términos de precisión, estabilidad y capacidad de generalización mediante métricas cuantitativas y análisis visuales. También se examina su comportamiento ante turbulencia atmosférica simulada, permitiendo una discusión integral sobre sus fortalezas y limitaciones.

Finalmente, en el **Capítulo 9**, se presentan las conclusiones generales de la tesis, destacando los hallazgos obtenidos y su impacto en la mejora de los sistemas ópticos adaptativos. Se discuten las contribuciones clave de este trabajo y se proponen posibles líneas de investigación futura para seguir avanzando en el área.

# Capítulo 2

## Planteamiento del Problema

### 2.1. Descripción del Problema

Las aberraciones ópticas son inherentes a todos los sistemas ópticos, desde los más simples hasta los más avanzados, y afectan directamente la calidad de imagen. Estas aberraciones pueden originarse por imperfecciones en los sistemas ópticos o por condiciones externas, como la turbulencia atmosférica, que distorsiona la propagación de la luz y afecta la precisión de su análisis.

Este trabajo de tesis se centra en la evaluación de distintas metodologías de análisis para el ajuste de aberraciones ópticas. Para ello, se utiliza un sensor de frente de onda tipo Shack-Hartmann, un dispositivo ampliamente reconocido por su capacidad para medir deformaciones en el frente de onda. Los datos recopilados y simulados por este sensor se analizan mediante tres enfoques: mínimos cuadrados, gradiente descendente y red hiperdinámica.

El objetivo principal es explorar la precisión de estas metodologías en escenarios definidos, permitiendo una comparación detallada de su desempeño. Este enfoque busca aportar una evaluación sólida de las ventajas y limitaciones de cada método, sin extenderse a implicaciones como la calidad de imagen, pero proporcionando una base sólida para futuras aplicaciones.

### 2.2. Preguntas de Investigación

Este proyecto se guía por las siguientes preguntas clave:

1. ¿Qué tan precisos son los métodos de mínimos cuadrados, gradiente descendente y red hiperdinámica en el ajuste de aberraciones ópticas?

2. ¿Cuáles son las ventajas y limitaciones específicas de cada método?
3. ¿Cómo afecta la presencia de valores atípicos al desempeño de los métodos de mínimos cuadrados, gradiente descendente y red hiperdinámica en el ajuste de aberraciones ópticas?
4. ¿Qué metodologías presentan un mayor potencial para aplicaciones futuras en sistemas ópticos complejos?

Para orientar al lector sobre dónde se abordan las preguntas planteadas, en la Tabla 2.1 se indica la correspondencia entre cada una de ellas y los capítulos respectivos de la tesis.

**Tabla 2.1:** Correspondencia entre las preguntas de investigación y los capítulos donde se abordan.

Pregunta (N° y abreviación)	Capítulo(s) donde se aborda
P1: Precisión de los métodos	Capítulo 7, Sección 8.1, Capítulo 9
P2: Ventajas y limitaciones	Capítulo 4, Capítulo 8, Capítulo 9
P3: Impacto de valores atípicos	Sección 6.2, Sección 8.1.2
P4: Aplicaciones futuras	Sección 2.6, Sección 8.1.3, Capítulo 9

## 2.3. Hipótesis

Se plantea que la implementación de métodos basados en inteligencia artificial permitirá un ajuste más preciso de las aberraciones ópticas en comparación con enfoques tradicionales como el método de mínimos cuadrados. En particular, se espera que el gradiente descendente pueda ofrecer una alternativa viable al método de mínimos cuadrados, destacando por su capacidad para gestionar valores atípicos y su bajo costo computacional. Asimismo, técnicas más avanzadas como la red hiperdinámica podrían sobresalir en la identificación de patrones complejos en los datos, especialmente en escenarios con datos incompletos o afectados por puntos atípicos. Este planteamiento sugiere que los enfoques basados en inteligencia artificial tienen el potencial de superar limitaciones específicas de los métodos tradicionales, ofreciendo una evaluación más precisa y adaptativa.

## 2.4. Objetivo General y Objetivos Específicos

**Objetivo General:**

Desarrollar una metodología basada en inteligencia artificial (IA) para medir y evaluar las aberraciones ópticas en sistemas ópticos, mediante el análisis de datos obtenidos con un sensor de frente de onda tipo Shack-Hartmann.

**Objetivos Específicos:**

1. Implementar y adaptar distintas metodologías para el ajuste de aberraciones ópticas, incluyendo mínimos cuadrados, gradiente descendente y redes hiperdinámicas.
2. Evaluar la precisión, adaptabilidad y tiempo de cómputo de cada metodología en escenarios de muestra única y múltiples muestras.
3. Comparar cuantitativamente los resultados obtenidos con cada enfoque, identificando sus ventajas y limitaciones en términos de desempeño y eficiencia.
4. Establecer una base metodológica que sirva como referencia para futuras investigaciones orientadas al ajuste de aberraciones ópticas mediante técnicas de inteligencia artificial.

Para alcanzar los objetivos planteados, en los capítulos siguientes se explorarán diversos métodos computacionales, incluyendo el ajuste por mínimos cuadrados, el uso del optimizador Adam en algoritmos de gradiente descendente, y redes hiperdinámicas. Cada uno de estos enfoques está alineado con el propósito central de estimar y corregir aberraciones ópticas de manera eficiente, evaluando su desempeño en términos de precisión, adaptabilidad y costo computacional.

## 2.5. Justificación del Proyecto

La evaluación precisa de aberraciones ópticas es fundamental para el desarrollo de sistemas ópticos avanzados, especialmente en aplicaciones astronómicas y aeroespaciales donde la turbulencia atmosférica introduce distorsiones significativas en el frente de onda. Estas aberraciones afectan directamente la calidad de las imágenes captadas por telescopios y sensores, reduciendo la capacidad de resolución y dificultando el análisis detallado de objetos lejanos.

Este proyecto se justifica por la necesidad de explorar metodologías alternativas para el ajuste de aberraciones inducidas por turbulencia atmosférica, mediante el uso de inteligencia artificial. La evaluación de distintos enfoques, como mínimos cuadrados, gradiente descendente y redes hiperdinámicas, per-

mitirá determinar su precisión, adaptabilidad y eficiencia computacional en escenarios simulados. Esta comparación aportará un conocimiento técnico valioso que puede servir como base para el diseño de sistemas ópticos más robustos y adaptativos frente a condiciones atmosféricas variables. Al enfocarse en la caracterización y corrección de aberraciones, este trabajo contribuye al avance de soluciones computacionales aplicadas a la óptica adaptativa y observación astronómica.

## 2.6. Aportación Científica y Tecnológica

La principal contribución de este trabajo radica en el desarrollo y evaluación de metodologías innovadoras para el ajuste de aberraciones ópticas, integrando modelos de inteligencia artificial en el análisis y procesamiento de datos. Desde un enfoque científico, este estudio permitirá identificar fortalezas y limitaciones de estas metodologías en términos de precisión, proporcionando un marco sólido para la selección y aplicación de métodos según las necesidades específicas de cada escenario óptico.

En el ámbito tecnológico, este trabajo no solo evalúa el desempeño de metodologías existentes, sino que también introduce herramientas basadas en inteligencia artificial, diseñadas para abordar desafíos asociados con el ajuste de aberraciones en datos complejos o condiciones variables. Estas herramientas representan un avance significativo al proponer soluciones flexibles y precisas que pueden ser empleadas en sistemas ópticos complejos para abordar desafíos específicos en el ajuste de aberraciones.

Aunque el enfoque principal se centra en la precisión de los métodos, este trabajo abre nuevas perspectivas para investigaciones futuras al establecer una base técnica detallada y generar herramientas que podrían extenderse a aplicaciones prácticas más amplias en la corrección de aberraciones ópticas y otros campos relacionados con la óptica avanzada.

## 2.7. Delimitaciones

La delimitación de este proyecto se enfoca específicamente en telescopios reflectores, excluyendo otros tipos de telescopios como refractores o catadióptricos. Esta decisión se toma para concentrar los esfuerzos y recursos en la optimización de un tipo particular de instrumento astronómico y maximizar la efectividad de la implementación propuesta.

Se limitará la investigación a la utilización de un sensor de frente de onda

tipo Shack-Hartmann como principal herramienta de medición. La elección de este sensor responde a su capacidad para proporcionar mediciones precisas y detalladas de las aberraciones ópticas presentes en el frente de onda, lo cual es fundamental para el objetivo general del proyecto.

Además, la implementación de polinomios de Zernike se considera como un enfoque específico dentro del amplio espectro de técnicas disponibles para el ajuste de aberraciones ópticas. Este método se selecciona por su eficacia en la representación matemática de las aberraciones y su utilidad en la obtención de los coeficientes que mejor ajustan al frente de onda medido.

La delimitación a estos tres elementos —telescopios reflectores, sensor de frente de onda tipo Shack-Hartmann, y polinomios de Zernike— permite asegurar la coherencia y la eficiencia en la ejecución del proyecto.

# Capítulo 3

## Fundamentos Ópticos

En este capítulo se exploran los conceptos fundamentales de la óptica, desde los fenómenos básicos de reflexión y refracción hasta temas avanzados como el trazo de rayos, aberraciones y sistemas ópticos compuestos. Estos fundamentos proporcionan lo necesario para comprender la formación de imágenes y su relevancia en el análisis y diseño de sistemas ópticos, sentando así las bases para abordar el problema principal de este trabajo.

### 3.1. Reflexión y Refracción

La reflexión y refracción son fundamentales en óptica, ya que explican cómo la luz interactúa con distintas superficies y medios, ya sea cambiando de dirección o retornando al medio inicial. Estos conceptos son cruciales para el desarrollo y estudio de sistemas ópticos. Cuando la luz incide sobre una superficie, una fracción de su energía puede reflejarse hacia el medio de origen, mientras que otra parte puede transmitirse al medio adyacente, alterando su dirección (Hecht, 2002). La proporción de luz reflejada o refractada está determinada por las características ópticas de los medios, como sus índices de refracción, y por la configuración geométrica de la interfaz entre ellos (Born and Wolf, 1999).

### 3.2. Reflexión y Refracción

La propagación de la luz se rige por principios fundamentales como la Ley de la Reflexión y la Ley de Snell, ampliamente tratados en la literatura clásica de óptica (Hecht, 2002; Born and Wolf, 1999). A continuación, se retoman

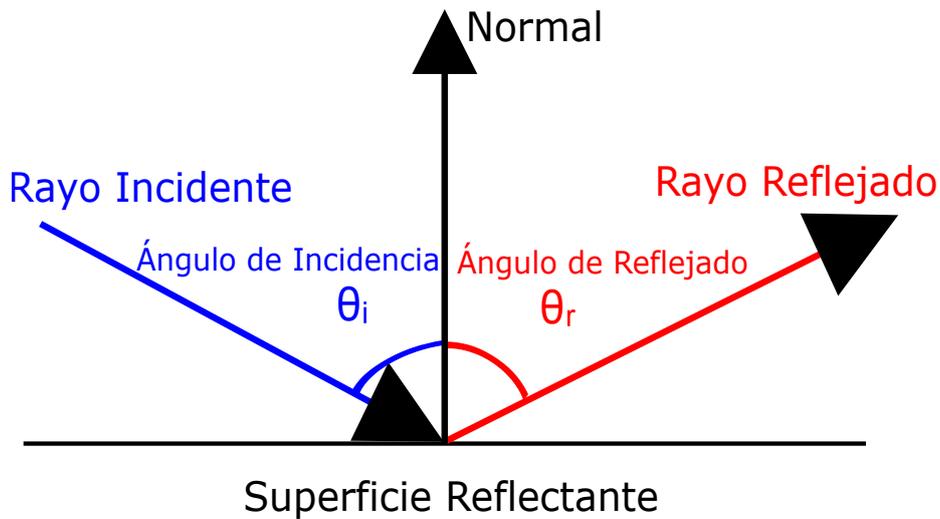
brevemente sus formulaciones, ya que proporcionan la base matemática necesaria para los análisis realizados en capítulos posteriores.

### 3.2.1. Ley de la Reflexión

La Ley de la Reflexión establece que el ángulo de incidencia,  $\theta_i$ , es igual al ángulo de reflexión,  $\theta_r$ , y que ambos rayos, junto con la normal a la superficie, yacen en un mismo plano:

$$\theta_i = \theta_r. \quad (3.1)$$

Este comportamiento se ilustra en la Figura 3.1, donde se muestra un esquema representativo de la reflexión especular.



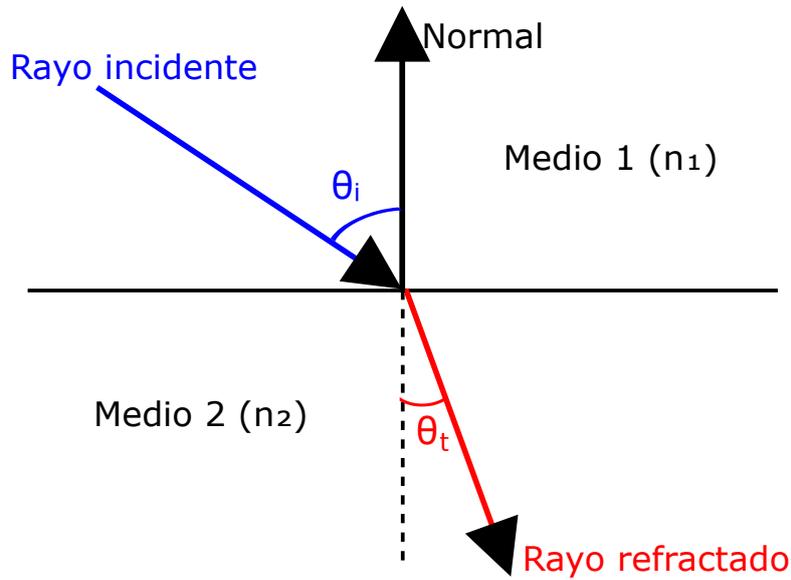
**Figura 3.1:** Esquema ilustrando la Ley de la Reflexión.

### 3.2.2. Ley de la Refracción

Cuando la luz atraviesa una interfaz entre dos medios con diferentes índices de refracción  $n_1$  y  $n_2$ , cambia su dirección de propagación de acuerdo con la Ley de Snell:

$$n_1 \cdot \sin(\theta_i) = n_2 \cdot \sin(\theta_t). \quad (3.2)$$

Este fenómeno se muestra en la Figura 3.2, donde el rayo se curva dependiendo de la relación entre los índices de refracción.



**Figura 3.2:** Esquema ilustrando la Ley de la Refracción.

Además, cuando el rayo incide desde un medio con mayor índice de refracción hacia uno de menor índice, puede ocurrir reflexión interna total. El ángulo crítico para este fenómeno se calcula mediante:

$$\theta_c = \arcsin\left(\frac{n_2}{n_1}\right). \quad (3.3)$$

Estos principios son esenciales en el diseño de elementos ópticos como lentes, espejos y fibras ópticas, y sientan las bases para comprender fenómenos más complejos en la formación y análisis de imágenes, como se discute en la Sección 3.3.

### 3.3. Lentes y Espejos Simples

Cuando observamos un objeto, la luz reflejada en su superficie se dispersa en múltiples direcciones. Los elementos ópticos fundamentales, como lentes y espejos, redirigen estas ondas de luz mediante procesos de refracción y reflexión, respectivamente. Las lentes, ya sean convergentes o divergentes, tienen la capacidad de hacer que la luz converja o diverja, mientras que los espejos utilizan la reflexión para redirigir la luz hacia un punto focal (Hecht, 2002).

El estudio de estos elementos sienta las bases para entender cómo los sistemas ópticos manipulan la luz para cumplir funciones específicas y analizar su comportamiento. Las principales propiedades que describen el comportamiento de las lentes son (Hecht, 2002):

- **Distancia Focal:** La distancia entre el centro de la lente y su punto focal. Es positiva para lentes convergentes y negativa para lentes divergentes.
- **Punto Focal:** El punto donde los rayos convergen o parecen divergir tras pasar por la lente.
- **Plano Focal:** La superficie donde se forma una imagen enfocada tras atravesar la lente.
- **Distancia Objeto-Imagen:** La distancia entre el objeto y su imagen formada, que depende de la posición relativa del objeto respecto a la lente.

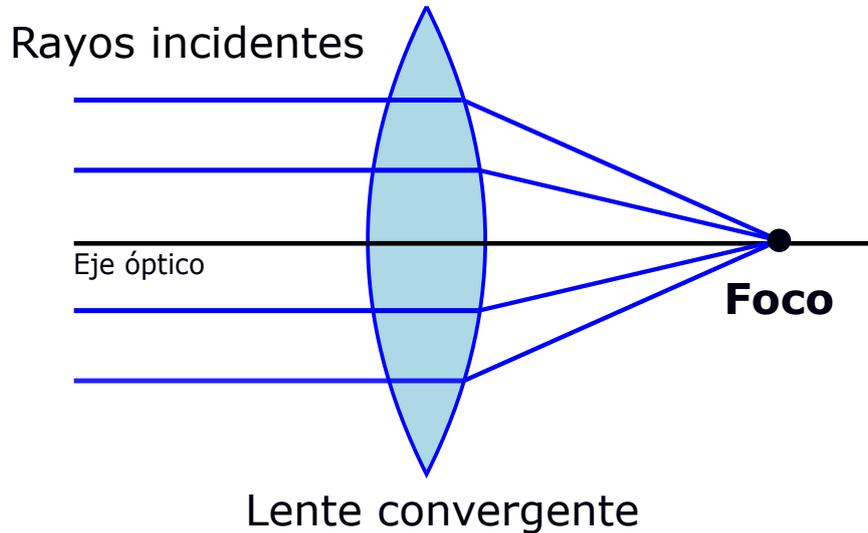
#### 3.3.1. Lentes Convergentes

Las lentes convergentes, más gruesas en el centro que en los bordes, tienen la capacidad de enfocar rayos paralelos hacia un punto focal real después de atravesarlas (Hecht, 2002). La Figura 3.3 ilustra este fenómeno, mostrando cómo los rayos incidentes paralelos a lo largo del eje óptico se refractan en la lente y convergen en un único punto focal.

La relación matemática que describe el comportamiento de estas lentes está dada por la **Ecuación de la Lente Delgada**:

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}, \quad (3.4)$$

donde  $f$  es la distancia focal,  $d_o$  es la distancia del objeto a la lente, y  $d_i$  es



**Figura 3.3:** Esquema de una lente convergente mostrando la convergencia de los rayos incidentes en el foco.

la distancia de la imagen a la lente. Esta Ecuación (3.4) permite determinar la posición de la imagen en función de la ubicación del objeto.

Además, la magnificación ( $M$ ) de la lente, que indica la relación entre el tamaño de la imagen ( $h_i$ ) y el tamaño del objeto ( $h_o$ ), se calcula como:

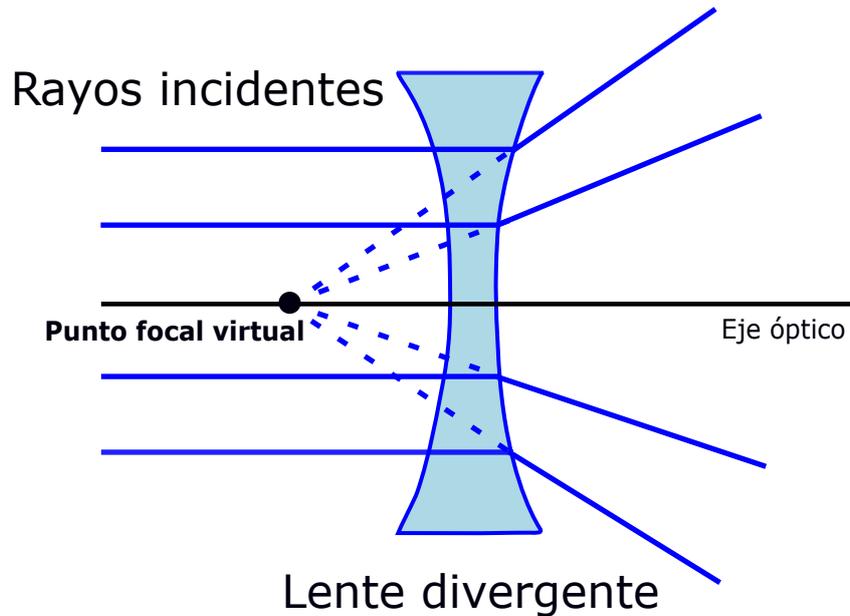
$$M = \frac{h_i}{h_o} = \frac{d_i}{d_o}. \quad (3.5)$$

La Ecuación (3.5) es útil para predecir el tamaño y la orientación de la imagen formada por lentes convergentes (Pedrotti and Pedrotti, 2006).

### 3.3.2. Lentes Divergentes

Las lentes divergentes, más delgadas en el centro que en los bordes, dispersan los rayos de luz que las atraviesan. A diferencia de las lentes convergentes, los rayos emergentes no se cruzan en un punto real, sino que parecen originarse de un punto focal virtual ubicado detrás de la lente (Hecht, 2002). Este

comportamiento óptico hace que las lentes divergentes sean ampliamente utilizadas en la corrección de aberraciones ópticas (Born and Wolf, 1999). La Figura 3.4 ilustra este fenómeno, mostrando el trazado de rayos y el punto focal virtual asociado.



**Figura 3.4:** Esquema de una lente divergente mostrando la dispersión de los rayos y la ubicación del punto focal virtual detrás de la lente.

La ecuación de la lente delgada para lentes divergentes se expresa como:

$$\frac{1}{f} = \frac{1}{d_o} - \frac{1}{d_i}. \quad (3.6)$$

En este caso,  $f$  y  $d_i$  son negativos porque el foco es virtual y está ubicado detrás de la lente.

Además, la magnificación ( $M$ ), que también es negativa, se calcula mediante:

$$M = -\frac{d_i}{d_o}. \quad (3.7)$$

Estas ecuaciones permiten modelar el comportamiento óptico de las lentes divergentes y su aplicación en sistemas que requieren imágenes virtuales (Hecht, 2002).

### 3.3.3. Espejos Ópticos

Los espejos cóncavos y convexos complementan el estudio de lentes al reflejar los rayos de luz y dirigirlos hacia un punto focal. Para los espejos cóncavos, los rayos convergen en un punto focal real, mientras que en los espejos convexos, los rayos parecen divergir desde un punto focal virtual detrás del espejo (Pedrotti and Pedrotti, 2006).

El comportamiento de los espejos puede describirse mediante la misma ecuación que las lentes delgadas (3.4), donde  $f$  es positiva en espejos cóncavos y negativa en espejos convexos (Pedrotti and Pedrotti, 2006). Esta relación permite calcular la posición y características de las imágenes formadas, dependiendo de la distancia del objeto.

El análisis de lentes y espejos simples proporciona una base sólida para comprender cómo los sistemas ópticos manipulan la luz. A partir de estos principios, es posible estudiar en detalle el proceso de formación de imágenes, lo cual será abordado en la Sección 3.4, donde se explorarán los distintos tipos de imágenes generadas y los factores que influyen en su localización, tamaño y orientación.

## 3.4. Formación de Imágenes

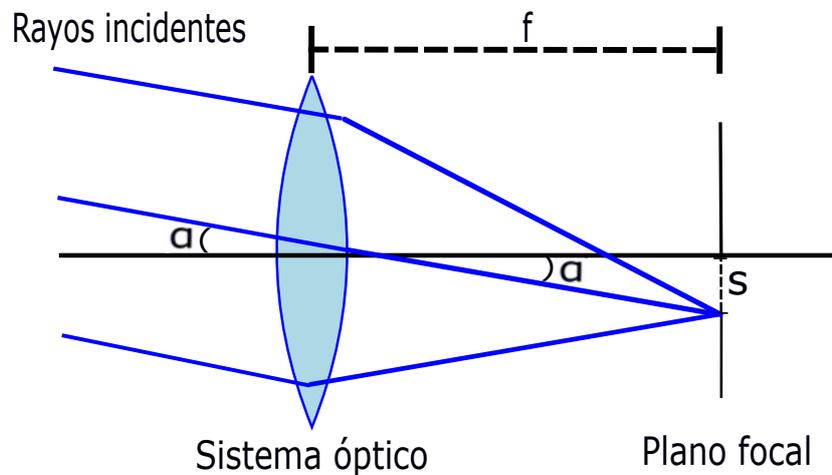
La formación de imágenes es un proceso central en los sistemas ópticos, donde los componentes descritos en la Sección 3.3, como lentes y espejos, interactúan con la luz para proyectarla en el plano focal. Este proceso, determinado por las propiedades geométricas y ópticas de los elementos del sistema, permite generar imágenes precisas que son fundamentales para aplicaciones técnicas y científicas (Born and Wolf, 1999).

Para objetos situados a grandes distancias, la posición de la imagen en el plano focal ( $s$ ) está relacionada con el ángulo subtendido por el objeto ( $\alpha$ ) y la distancia focal del sistema óptico ( $f$ ). Utilizando la aproximación de ángulos pequeños  $\tan \alpha \approx \alpha$ , esta relación se expresa como:

$$s = f\alpha, \tag{3.8}$$

donde la ecuación obtenida es fundamental para determinar la ubicación precisa de las imágenes (Pedrotti and Pedrotti, 2006).

La Figura 3.5 ilustra este proceso, mostrando cómo los rayos de luz provenientes de un objeto inclinado inciden en el sistema óptico y se enfocan en el plano focal. Se observa que los rayos provenientes de distintos puntos del objeto se refractan de manera proporcional a su ángulo de incidencia, formando una imagen en el plano focal cuya posición depende de la distancia focal del sistema óptico.



**Figura 3.5:** Esquema de la formación de imágenes en un sistema óptico, mostrando los rayos incidentes, el sistema óptico y el plano focal.

La escala de placa ( $EP$ ) es un parámetro clave que conecta el tamaño angular de los objetos observados con su representación en el plano focal. Esta escala se define como:

$$EP = \frac{206265}{f}, \quad (3.9)$$

donde  $EP$  se mide en segundos de arco por milímetro, y  $f$  es la distancia focal del sistema óptico. Este concepto, ampliamente utilizado en instrumentación óptica, permite evaluar la resolución angular del sistema (Hecht, 2002).

La eficiencia óptica depende tanto de la apertura del sistema ( $D$ ) como de la distancia focal ( $f$ ), parámetros que determinan la capacidad del sistema para recolectar luz. En este contexto,  $E$  representa la eficiencia óptica, definida

como el cociente entre la intensidad transmitida y la intensidad incidente en el sistema óptico. Esta relación se expresa como:

$$E \propto \left(\frac{D}{f}\right)^2. \quad (3.10)$$

Una mayor apertura y una distancia focal adecuada optimizan la calidad de las imágenes al maximizar la recolección de luz y minimizar los efectos de difracción (Malacara, 2007).

Estas relaciones geométricas y ópticas no solo son esenciales para comprender la formación de imágenes, sino que también establecen las bases para el análisis de sistemas ópticos más complejos. En estos sistemas, la interacción de múltiples componentes, como lentes y espejos, requiere un enfoque sistemático que permita modelar y optimizar su desempeño, tema que se aborda en detalle en la Sección 3.5.

## 3.5. Sistemas Ópticos

Los sistemas ópticos, compuestos por lentes, espejos y otros componentes, son fundamentales para modelar y analizar cómo la luz interactúa con estos elementos (Born and Wolf, 1999). Su estudio requiere herramientas matemáticas que permitan describir y predecir su comportamiento, lo cual es crucial para optimizar la formación de imágenes y garantizar su precisión en aplicaciones técnicas y científicas relevantes para este trabajo.

### 3.5.1. Matrices de Sistemas Ópticos

La representación mediante matrices es una herramienta clave para modelar la transformación de rayos de luz en sistemas ópticos compuestos. Este enfoque simplifica los cálculos necesarios para describir cómo los rayos se propagan a través de diversos elementos, como lentes y espacios libres (Pedrotti and Pedrotti, 2006).

La matriz del sistema  $M$  se utiliza para relacionar las propiedades iniciales y finales de un rayo al atravesar un elemento óptico. Por ejemplo, para un sistema con distancia focal  $f$ , la matriz se define como:

$$M = \begin{bmatrix} 1 & -\frac{1}{f} \\ 0 & 1 \end{bmatrix}. \quad (3.11)$$

Para modelar la propagación en el espacio libre, se emplea la matriz de distancia  $D$ , que describe cómo los rayos se desplazan axialmente en una distancia  $d$ :

$$D = \begin{bmatrix} 1 & 0 \\ d & 1 \end{bmatrix}. \quad (3.12)$$

La combinación de estas matrices permite calcular la transformación total de los rayos en sistemas ópticos compuestos. Por ejemplo, un sistema con distancia focal  $f$  y un desplazamiento axial  $d$  tiene una matriz total  $M_{\text{total}}$  descrita como:

$$M_{\text{total}} = \begin{bmatrix} 1 & 0 \\ d & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -\frac{1}{f} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{f} \\ d & 1 \end{bmatrix}. \quad (3.13)$$

Esta formulación es esencial para diseñar y evaluar sistemas ópticos complejos, facilitando el cálculo de trayectorias, distancias focales y magnificaciones ([Hecht, 2002](#)).

### 3.5.2. Trazado de Rayos

El trazado de rayos es una técnica fundamental para analizar cómo la luz interactúa con los sistemas ópticos. Este método permite evaluar el comportamiento de los rayos a medida que atraviesan los componentes del sistema, considerando tanto transformaciones matriciales como soluciones exactas ([Born and Wolf, 1999](#)).

#### 3.5.2.1. Trazado de Rayos Matricial

El trazado matricial utiliza matrices para describir las transformaciones de los rayos al interactuar con elementos ópticos. Este enfoque es particularmente útil para sistemas compuestos debido a su capacidad de modelar múltiples interacciones mediante multiplicaciones matriciales.

La matriz de transferencia general para un sistema óptico se expresa como:

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad (3.14)$$

donde los coeficientes  $A$ ,  $B$ ,  $C$  y  $D$  dependen de las propiedades específicas de los elementos del sistema ([Pedrotti and Pedrotti, 2006](#)).

### 3.5.2.2. Trazado de Rayos Exacto

En situaciones donde se requiere mayor precisión, el trazado exacto utiliza ecuaciones paramétricas para calcular la trayectoria de los rayos. Para un rayo que parte del punto  $(x_0, y_0, z_0)$  con dirección definida por los cosenos directores  $(l, m, n)$ , las ecuaciones paramétricas son:

$$x = x_0 + lt, \quad (3.15)$$

$$y = y_0 + mt, \quad (3.16)$$

$$z = z_0 + nt, \quad (3.17)$$

donde  $t$  representa la distancia recorrida por el rayo (Hecht, 2002). Estas ecuaciones permiten determinar con precisión los puntos de intersección con superficies y calcular los cosenos directores tras la interacción (Born and Wolf, 1999).

La combinación de matrices y trazado de rayos permite diseñar sistemas ópticos optimizados y minimizar aberraciones (Malacara, 2007). Los detalles sobre pupila del sistema, aberraciones e impacto de superficies cónicas se abordan en la Sección 3.6.

## 3.6. Características de los Sistemas Ópticos

Una vez descrito el sistema óptico, analizaremos su desempeño considerando características fundamentales como la pupila, las aberraciones y la geometría de las superficies ópticas utilizadas. Estos elementos determinan la eficiencia en la recolección de luz, la calidad de las imágenes formadas y la capacidad del sistema para minimizar los efectos no ideales, proporcionando una base sólida para sus aplicaciones.

### 3.6.1. Pupila del Sistema

La pupila del sistema juega un papel crucial en la formación de imágenes al regular la cantidad de luz que entra y sale del sistema óptico. La pupila de entrada controla la luz incidente, mientras que la pupila de salida define el ángulo máximo bajo el cual los rayos pueden emerger (Smith, 2007). La relación entre estas pupilas se expresa como:

$$\frac{D_o}{d_o} = \frac{D_i}{d_i}, \quad (3.18)$$

donde  $D_o$  y  $D_i$  son los diámetros de las pupilas de entrada y salida, y  $d_o$  y  $d_i$  las distancias asociadas. Además, la apertura numérica ( $NA$ ), que caracteriza la capacidad de resolución del sistema, se define como:

$$NA = \frac{D_i}{2z}, \quad (3.19)$$

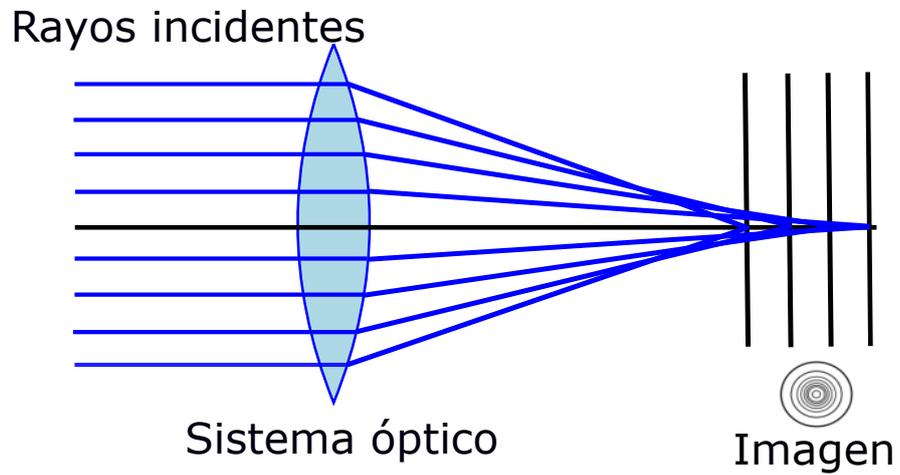
donde  $z$  es la distancia focal. Una apertura numérica más alta implica una mayor capacidad para resolver detalles finos (Pedrotti and Pedrotti, 2006).

### 3.6.2. Aberraciones de sistemas ópticos

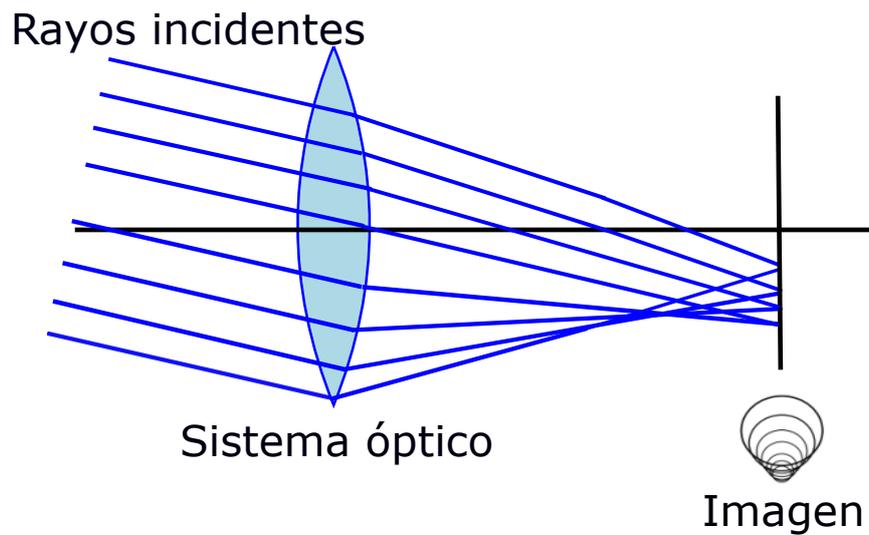
Las aberraciones representan desviaciones del comportamiento ideal de un sistema óptico, afectando la nitidez y claridad de las imágenes formadas (Hecht, 2002). Estas imperfecciones pueden degradar significativamente la calidad de una imagen, incluso en sistemas bien diseñados.

Una de las aberraciones más comunes es la aberración *esférica*, la cual ocurre cuando los rayos de luz que inciden en las zonas periféricas de una lente no convergen en el mismo punto focal que los rayos cercanos al eje óptico. Esto genera una imagen borrosa y con halos concéntricos alrededor del punto focal ideal. En la Figura 3.6, se observa cómo los rayos periféricos y axiales no coinciden en un único foco, lo que da lugar a un patrón característico en el plano imagen. Esta aberración es más pronunciada en lentes esféricas simples y puede minimizarse mediante el uso de lentes asféricas o combinaciones de lentes correctoras (Hecht, 2002).

Otra aberración significativa es la *coma*, que afecta a los objetos situados fuera del eje óptico. En este caso, los rayos no convergen en un punto único, sino que generan una imagen distorsionada con una forma característica de "cometa", alargando el punto de luz y causando una pérdida de simetría. Este efecto se debe a variaciones en el ángulo de incidencia de los rayos con respecto a la óptica y es un problema frecuente en telescopios y sistemas de gran apertura (Hecht, 2002). La Figura 3.7 muestra cómo la imagen de un punto fuera del eje se deforma, provocando que la luz se distribuya de manera desigual en el plano focal.

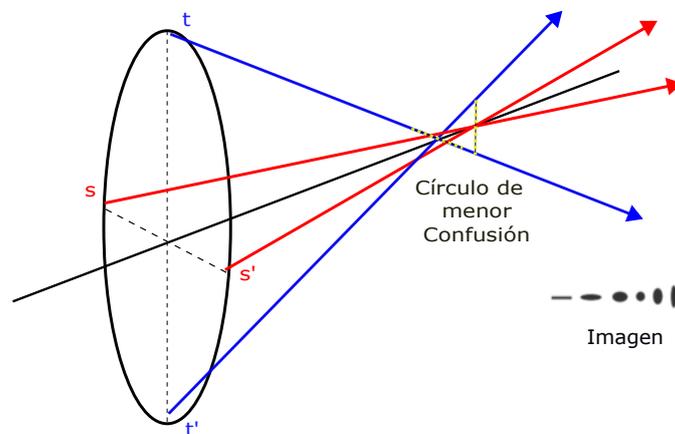


**Figura 3.6:** Representación de la aberración esférica. Se observan anillos de intensidad debido a la diferencia de foco entre los rayos periféricos y axiales.



**Figura 3.7:** Aberración de coma. Los rayos incidentes que llegan en ángulos oblicuos generan una imagen elongada con una estructura en forma de "cometa".

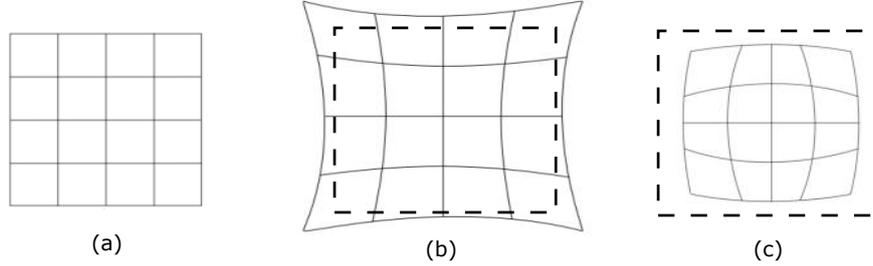
El *astigmatismo* es otra aberración importante que ocurre cuando los rayos provenientes de un objeto fuera del eje no convergen en un único punto focal, sino que forman dos planos de enfoque distintos en direcciones perpendiculares (Hecht, 2002). Como resultado, una imagen enfocada en un plano aparecerá desenfocada en el otro. Este fenómeno es particularmente problemático en sistemas ópticos que operan con ángulos de incidencia grandes (Hecht, 2002). En la Figura 3.8, se representa esta aberración, mostrando cómo la imagen varía dependiendo de la orientación del plano de enfoque.



**Figura 3.8:** Astigmatismo óptico. Se observan dos planos de enfoque distintos, tangencial y sagital, lo que impide una imagen única y nítida.

Por último, la *distorsión* afecta la forma de la imagen sin modificar su enfoque. Existen dos tipos principales: la *distorsión en barril*, en la que las líneas rectas parecen curvarse hacia afuera, y la *distorsión en cojín*, en la que las líneas rectas se curvan hacia adentro. La Figura 3.9 ilustra ambas formas de distorsión junto con una cuadrícula ideal como referencia. Estas deformaciones son comunes en lentes de cámaras fotográficas y proyectores y se deben a variaciones en la magnificación de la imagen en diferentes regiones del campo visual (Hecht, 2002).

Estas aberraciones representan limitaciones fundamentales en el diseño de sistemas ópticos. Un análisis detallado de estas aberraciones se puede realizar mediante polinomios de Zernike, que se abordarán en la Sección 3.6.



**Figura 3.9:** Distorsión óptica. (a) Cuadrícula sin distorsión. (b) Distorsión en barril. (c) Distorsión en cojín.

### 3.6.3. Superficies Cónicas

Las superficies cónicas, derivadas de la intersección de un cono con un plano en distintos ángulos, desempeñan un papel fundamental en la corrección de aberraciones ópticas. Dependiendo de la inclinación del plano de corte, se obtienen distintas curvas cónicas, como la circunferencia, la elipse, la parábola y la hipérbola, las cuales presentan propiedades geométricas y ópticas particulares. La Figura 3.10 muestra estas secciones cónicas y su clasificación en función de la geometría del corte.

Matemáticamente, estas superficies se describen mediante la ecuación sagital:

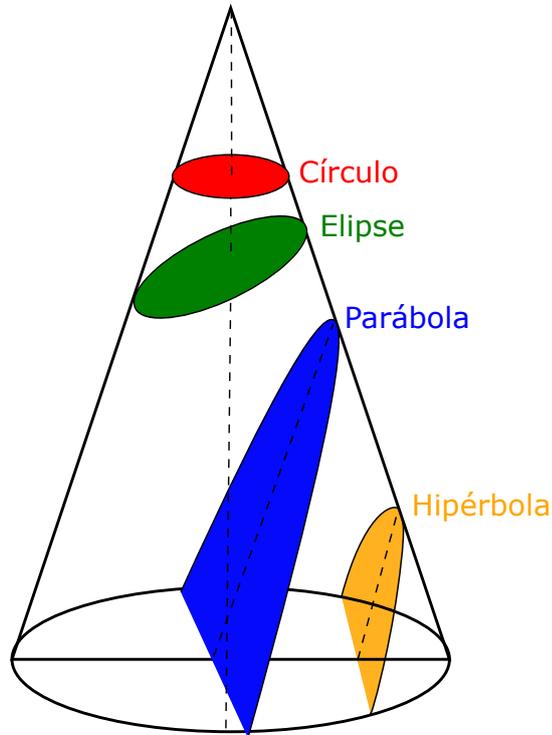
$$z = \frac{cs^2}{1 + \sqrt{1 - (K + 1)c^2s^2}}, \quad (3.20)$$

donde  $z$  representa la altura de la superficie,  $s$  la distancia radial,  $c = 1/r$  la curvatura paraxial y  $K$  la constante de conicidad (Hecht, 2002). Según el valor de  $K$ , las superficies se clasifican como se muestra en la Tabla 3.1.

**Tabla 3.1:** Clasificación de las superficies cónicas según la constante  $K$ .

Superficie	Valor de $K$
Hipérbola	$K < -1$
Parábola	$K = -1$
Elipse prolata	$-1 < K < 0$
Circunferencia	$K = 0$
Elipse oblata	$K > 0$

Otro parámetro importante en el análisis de superficies cónicas es la excen-



**Figura 3.10:** Intersección de un cono con un plano en distintos ángulos, generando diferentes curvas cónicas: circunferencia, elipse, parábola e hipérbola.

tricidad ( $e$ ), que describe la forma geométrica de las curvas asociadas. Para una elipse, la excentricidad se calcula como:

$$e = \sqrt{1 - \frac{b^2}{a^2}}, \quad (3.21)$$

donde  $a$  y  $b$  son los semiejes mayor y menor, respectivamente. Una excentricidad  $e = 0$  corresponde a una circunferencia, mientras que valores más cercanos a  $e = 1$  indican una mayor elongación.

El uso de superficies cónicas en óptica proporciona soluciones efectivas para minimizar aberraciones. Los espejos parabólicos, por ejemplo, eliminan la aberración esférica al concentrar rayos paralelos en un único punto focal, mientras que los espejos hiperbólicos permiten corregir aberraciones ópticas

avanzadas gracias a sus propiedades geométricas ([Born and Wolf, 1999](#)).

El análisis detallado de estas superficies es clave en el diseño de sistemas ópticos avanzados. En la siguiente sección, se introducirán los polinomios de Zernike, herramientas matemáticas esenciales para cuantificar las aberraciones ópticas de manera precisa.

### 3.7. Polinomios de Zernike

Los polinomios de Zernike son herramientas fundamentales en la descripción y ajuste de aberraciones ópticas. Aunque inicialmente se pueden abordar en coordenadas cartesianas, su representación en coordenadas polares resulta más conveniente para describir fenómenos ópticos, ya que reflejan de manera natural las simetrías inherentes a los sistemas ópticos ([Born and Wolf, 1999](#)). Este enfoque facilita el modelado y análisis de aberraciones en el contexto de este estudio.

En coordenadas cartesianas, la transición a polares  $(\rho, \theta)$  se define mediante:

$$\rho = \sqrt{x^2 + y^2}, \quad \theta = \arctan\left(\frac{y}{x}\right). \quad (3.22)$$

Esta transformación permite mapear cualquier función  $f(x, y)$  al dominio circular unitario, representándola como  $f(\rho, \theta)$ . En este contexto, los polinomios de Zernike, propuestos por Frits Zernike en 1934, ofrecen una representación ortogonal de aberraciones ([Zernike, 1934](#)).

En coordenadas polares  $(\rho, \theta)$ :

- $\rho$  es la coordenada radial normalizada, con  $0 \leq \rho \leq 1$ .
- $\theta$  es la coordenada azimutal, con  $0 \leq \theta < 2\pi$ .

La expresión general de los polinomios de Zernike es:

**Caso para  $m$  par:**

$$Z_n^m(\rho, \theta) = R_n^m(\rho) \cos(m\theta), \quad (3.23)$$

**Caso para  $m$  impar:**

$$Z_n^m(\rho, \theta) = R_n^m(\rho) \sin(m\theta), \quad (3.24)$$

donde:

- $Z_n^m(\rho, \theta)$  es el polinomio de Zernike de orden radial  $n$  y frecuencia azimutal  $m$ .
- $R_n^m(\rho)$  es la parte radial, definida como:

$$R_n^m(\rho) = \sum_{k=0}^{(n-m)/2} \frac{(-1)^k \cdot \rho^{n-2k}}{k! \cdot \left(\frac{n+m}{2} - k\right)! \cdot \left(\frac{n-m}{2} - k\right)!}. \quad (3.25)$$

### 3.7.1. Notación de Noll

La Notación de Noll es una convención que asigna índices enteros  $\mathcal{N}$  a los polinomios de Zernike, lo que facilita su identificación y uso en aplicaciones (Noll, 1976). Esta notación permite mapear  $n$  y  $m$  de manera única a un índice secuencial  $j$ :

$$\mathcal{N} = \frac{n(n+2) + m}{2}. \quad (3.26)$$

Para garantizar la ortogonalidad en el dominio circular unitario, los polinomios de Zernike normalizados toman la forma:

- Para  $m \neq 0$ :

$$Z_n^m(\rho, \theta) = \sqrt{2(n+1)} \cdot R_n^m(\rho) \cos(m\theta), \quad (\text{par}) \quad (3.27)$$

$$Z_n^m(\rho, \theta) = \sqrt{2(n+1)} \cdot R_n^m(\rho) \sin(m\theta), \quad (\text{impar}) \quad (3.28)$$

- Para  $m = 0$ :

$$Z_n^0(\rho) = \sqrt{n+1} \cdot R_n^0(\rho). \quad (3.29)$$

Al trabajar directamente en coordenadas polares, se simplifican los cálculos computacionales y se asegura la representación precisa de las aberraciones. Además, su ortogonalidad facilita la deconvolución de componentes individuales de aberración (Noll, 1976).

Presentamos la Tabla 3.2, que resume los primeros 37 polinomios de Zernike según la nomenclatura de Noll. Esta tabla proporciona una referencia directa para los coeficientes y sus respectivas representaciones matemáticas utilizadas en este estudio.

**Tabla 3.2:** Lista de los primeros 37 polinomios de Zernike en la nomenclatura de Noll.

Coef.	Nombre	Polinomio
0	Piston	1
1	Tilt $x$	$\sqrt{4p} \cos(\varphi)$
2	Tilt $y$	$\sqrt{4p} \sin(\varphi)$
3	Power (Focus)	$\sqrt{3}(2p^2 - 1)$
4	Astigmatismo $y$	$\sqrt{6p^2} \sin(2\varphi)$
5	Astigmatismo $x$	$\sqrt{6p^2} \cos(2\varphi)$
6	Coma $y$	$\sqrt{8}(3p^3 - 2p) \sin(\varphi)$
7	Coma $x$	$\sqrt{8}(3p^3 - 2p) \cos(\varphi)$
8	Trefoil $y$	$\sqrt{8p^3} \sin(3\varphi)$
9	Trefoil $x$	$\sqrt{8p^3} \cos(3\varphi)$
10	Esférica primaria	$\sqrt{5}(6p^4 - 6p^2 + 1)$
11	Astigmatismo secundario $x$	$\sqrt{10}(4p^4 - 3p^2) \cos(2\varphi)$
12	Astigmatismo secundario $y$	$\sqrt{10}(4p^4 - 3p^2) \sin(2\varphi)$
13	Tetrafoil $x$	$\sqrt{10p^4} \cos(4\varphi)$
14	Tetrafoil $y$	$\sqrt{10p^4} \sin(4\varphi)$
15	Coma secundaria $x$	$\sqrt{12}(10p^5 - 12p^3 + 3p) \cos(\varphi)$
16	Coma secundaria $y$	$\sqrt{12}(10p^5 - 12p^3 + 3p) \sin(\varphi)$
17	Trefoil secundario $x$	$\sqrt{12}(5p^5 - 4p^3) \cos(3\varphi)$
18	Trefoil secundario $y$	$\sqrt{12}(5p^5 - 4p^3) \sin(3\varphi)$
19	Pentafoil $x$	$\sqrt{12p^5} \cos(5\varphi)$
20	Pentafoil $y$	$\sqrt{12p^5} \sin(5\varphi)$
21	Esférica secundaria	$\sqrt{7}(20p^6 - 30p^4 + 12p^2 - 1)$
22	Astigmatismo terciario $y$	$\sqrt{14}(15p^6 - 20p^4 + 6p^2) \sin(2\varphi)$
23	Astigmatismo terciario $x$	$\sqrt{14}(15p^6 - 20p^4 + 6p^2) \cos(2\varphi)$
24	Tetrafoil secundario $y$	$\sqrt{14}(6p^6 - 5p^4) \sin(4\varphi)$
25	Tetrafoil secundario $x$	$\sqrt{14}(6p^6 - 5p^4) \cos(4\varphi)$
26	Hexafoil $y$	$\sqrt{14p^6} \sin(6\varphi)$
27	Hexafoil $x$	$\sqrt{14p^6} \cos(6\varphi)$
28	Coma terciaria $y$	$\sqrt{16}(35p^7 - 60p^5 + 30p^3 - 4p) \sin(\varphi)$
29	Coma terciaria $x$	$\sqrt{16}(35p^7 - 60p^5 + 30p^3 - 4p) \cos(\varphi)$
30	Trefoil terciario $y$	$\sqrt{16}(21p^7 - 30p^5 + 10p^3) \sin(3\varphi)$
31	Trefoil terciario $x$	$\sqrt{16}(21p^7 - 30p^5 + 10p^3) \cos(3\varphi)$
32	Pentafoil secundario $y$	$\sqrt{16}(7p^7 - 6p^5) \sin(5\varphi)$
33	Pentafoil secundario $x$	$\sqrt{16}(7p^7 - 6p^5) \cos(5\varphi)$
34	Septafoil $y$	$\sqrt{16p^7} \sin(7\varphi)$
35	Septafoil $x$	$\sqrt{16p^7} \cos(7\varphi)$
36	Esférica terciaria	$3^4 \sqrt{9}(70p^8 - 140p^6 + 90p^4 - 20p^2 + 1)$

## 3.8. Telescopios

Los telescopios son instrumentos esenciales para recolectar, redirigir y analizar la luz proveniente de objetos celestes. A través de diferentes configuraciones ópticas, permiten formar imágenes y caracterizar frentes de onda, siendo una herramienta indispensable en aplicaciones científicas y tecnológicas. En esta sección se describen los telescopios refractores y reflectores, destacando sus características clave y su relevancia en el contexto de este estudio.

### 3.8.1. Telescopios Refractores

Los telescopios refractores utilizan lentes para redirigir y enfocar la luz mediante el fenómeno de refracción. La trayectoria de la luz está descrita por la Ley de Snell, presentada en la Ecuación (3.2). Aunque ofrecen imágenes nítidas, los telescopios refractores presentan limitaciones debido a aberraciones cromáticas, ya que la luz de diferentes longitudes de onda se curva de manera desigual al atravesar las lentes (Born and Wolf, 1999).

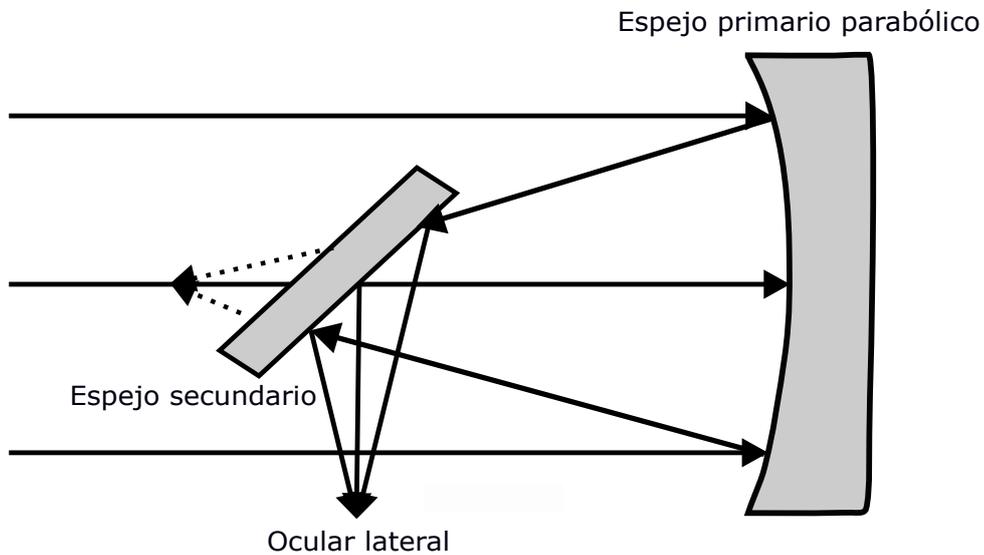
Para mitigar estos efectos, se pueden emplear lentes acromáticas, que combinan materiales con diferentes índices de refracción (Hecht, 2002). Sin embargo, la fabricación de lentes grandes y precisas es costosa, lo que limita su uso en investigaciones avanzadas. A pesar de estas restricciones, los telescopios refractores son útiles en aplicaciones terrestres y observaciones astronómicas básicas.

### 3.8.2. Telescopios Reflectores

A diferencia de los telescopios refractores, los reflectores utilizan espejos curvos para recolectar y redirigir la luz, como se describe mediante la Ley de Reflexión en la Ecuación (3.1). Este diseño elimina las aberraciones cromáticas y permite alcanzar aperturas significativamente mayores, esenciales para observar objetos débiles y distantes (Karttunen et al., 2017).

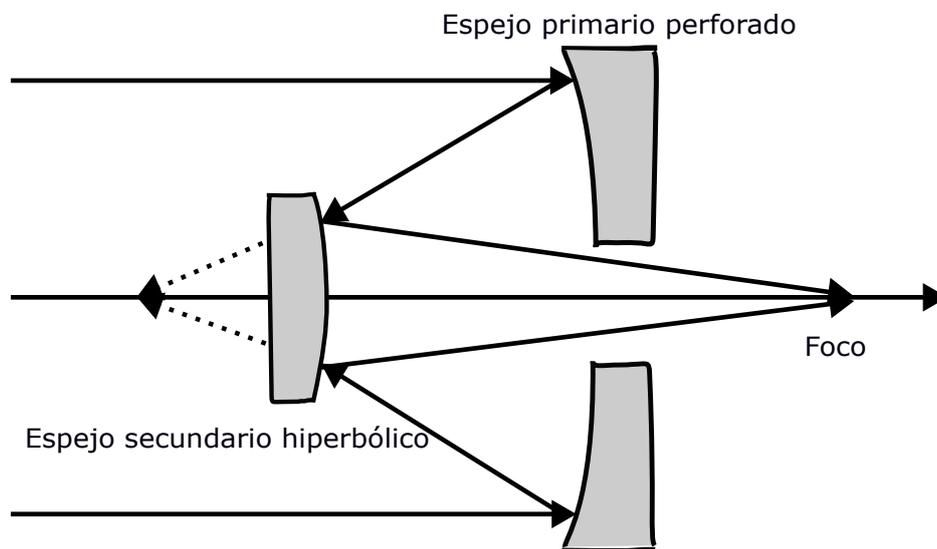
Entre las configuraciones más utilizadas se encuentran (Karttunen et al., 2017):

**Telescopio Newtoniano:** utiliza un espejo primario parabólico y un espejo secundario plano que desvía la luz hacia un ocular lateral. Es ampliamente usado en astronomía amateur por su simplicidad y bajo costo. Su configuración óptica se ilustra en la Figura 3.11.



**Figura 3.11:** Diseño óptico de un telescopio reflectante tipo Newtoniano.

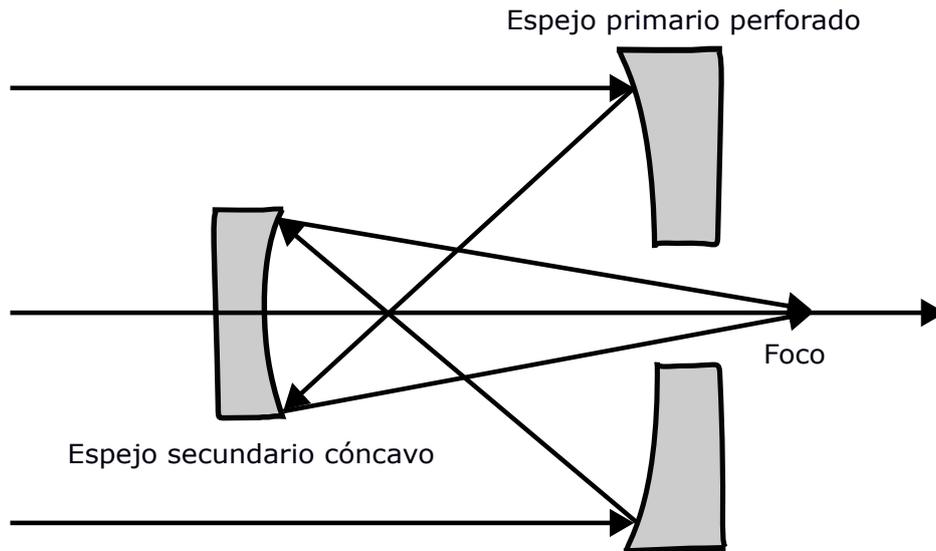
**Telescopio Cassegrain:** emplea un espejo secundario hiperbólico que refleja la luz hacia un foco ubicado detrás del espejo primario perforado. Su diseño compacto y larga distancia focal lo hacen ideal para investigación astronómica. Véase Figura 3.12.



**Figura 3.12:** Diseño óptico del telescopio Cassegrain.

**Telescopio Gregoriano:** similar al Cassegrain, pero con un espejo secun-

dario cóncavo. Aunque menos común, ofrece ventajas en ciertas aplicaciones ópticas, como aquellas que requieren una mejor corrección de aberración *esférica*, ya que su configuración permite utilizar espejos con curvaturas específicas que compensan este tipo de distorsión. Se muestra en la Figura 3.13.



**Figura 3.13:** Diseño óptico del telescopio Gregoriano.

**Telescopio Ritchey-Chrétien:** variante avanzada del diseño Cassegrain que utiliza espejos hiperbólicos tanto en el primario como en el secundario. Es ampliamente utilizado en telescopios profesionales y espaciales debido a su corrección de aberraciones fuera del eje.

Con los conceptos fundamentales de óptica y las configuraciones principales de telescopios reflectores ya descritos, se han sentado las bases necesarias para abordar el estudio del frente de onda. Los telescopios reflectores, en particular, son esenciales para generar frentes de onda bien definidos, lo cual resulta clave en la caracterización óptica y el análisis de aberraciones.

Los principios discutidos en este capítulo proporcionan la base conceptual para la interpretación de mediciones realizadas con sensores de frente de onda, como se analiza en el siguiente capítulo. En particular, se explorará el uso de herramientas avanzadas para medir y ajustar estos frentes de onda, comenzando con una descripción detallada del sensor Shack-Hartmann.

# Capítulo 4

## Análisis y Ajuste del Frente de Onda

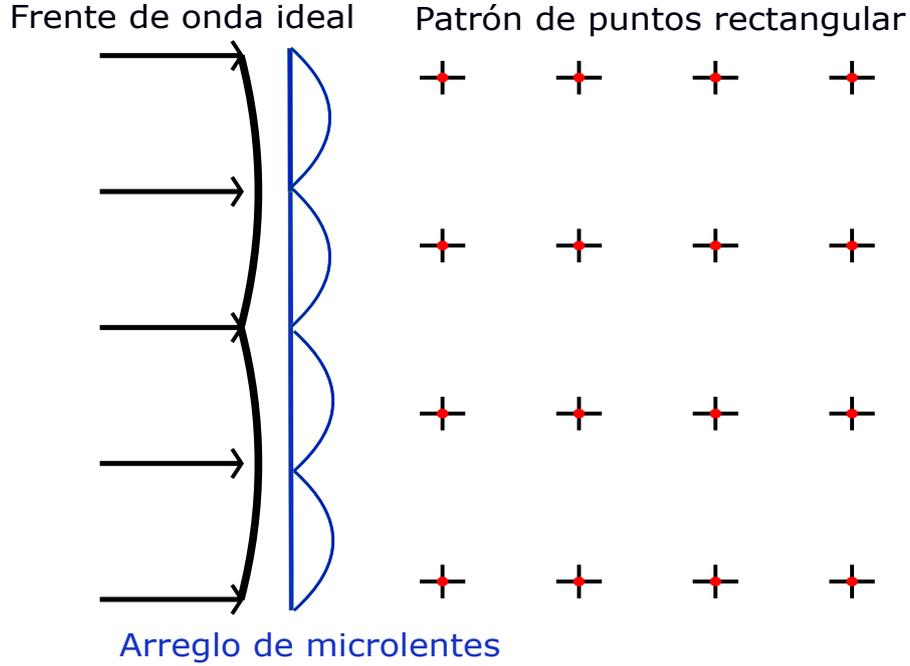
El análisis del frente de onda es una herramienta indispensable para la caracterización de sistemas ópticos. En este capítulo se estudiará la reconstrucción y evaluación de frentes de onda mediante el uso del sensor Shack-Hartmann, un dispositivo ampliamente utilizado en óptica adaptativa, control de calidad óptica y diseño experimental.

Los conceptos teóricos presentados en el capítulo anterior servirán como base para interpretar los desplazamientos de los puntos focales medidos por el sensor y vincularlos con aberraciones específicas. A lo largo del capítulo, se establecerá la formulación matemática que permite traducir estos desplazamientos en representaciones cuantitativas del frente de onda, las cuales serán fundamentales para los métodos de ajuste que se desarrollarán en los capítulos siguientes.

Un frente de onda ideal, ya sea plano o esférico, puede verse afectado por imperfecciones en los elementos ópticos o por perturbaciones externas, como la turbulencia atmosférica, generando aberraciones que degradan la calidad de las imágenes (Goel and Ganotra, 2024). La Figura 4.1 ilustra un frente de onda ideal al pasar por un sensor Shack-Hartmann, donde los spots formados en el detector permanecen centrados y uniformemente distribuidos.

### 4.1. Sensor Shack-Hartmann

El sensor Shack-Hartmann es una herramienta ampliamente utilizada en óptica para medir aberraciones del frente de onda. Desarrollado por Roland



**Figura 4.1:** Frente de onda ideal pasando por un sensor Shack-Hartmann. Los spots formados en el detector se mantienen alineados en una distribución regular.

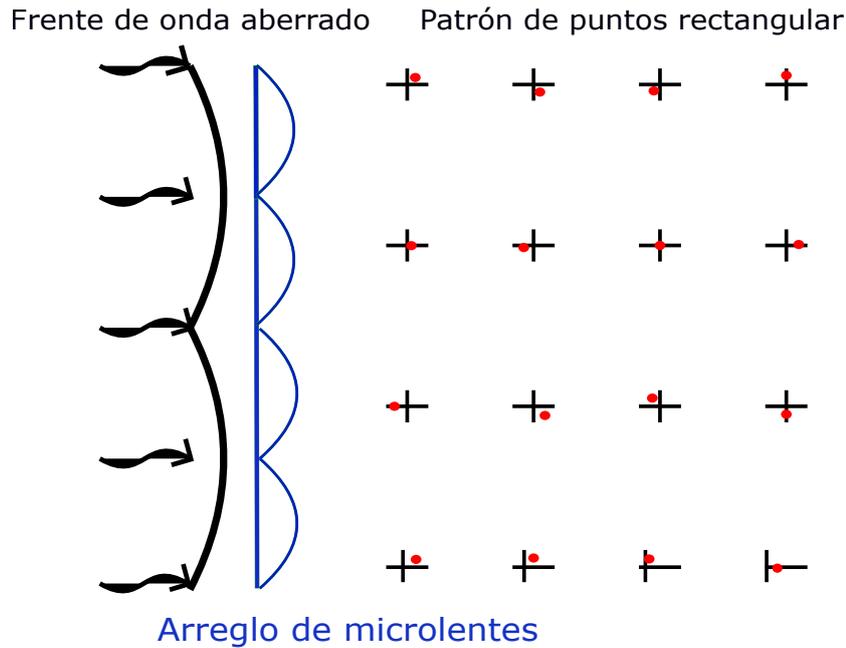
Shack y Ben Platt en 1971 ([Shack, 1971](#)), este sensor perfeccionó la pantalla de Hartmann mediante la incorporación de una matriz de microlentes, lo que permitió realizar mediciones locales del frente de onda con alta precisión.

El principio de funcionamiento del sensor consiste en dividir el frente de onda incidente en múltiples subfrentes utilizando una matriz de microlentes. Cada microlente focaliza su subfrente en un punto (spot) sobre un detector, típicamente una cámara CCD, cuyas posiciones varían en función de las pendientes locales del frente de onda ([Tyson, 2015](#)). Las desviaciones de las posiciones medidas  $(P_x, P_y)$  respecto a las posiciones ideales  $(L_x, L_y)$  están relacionadas con las derivadas del frente de onda mediante la ecuación:

$$\frac{\partial W}{\partial x} = \frac{L_x - P_x}{f_m}, \quad \frac{\partial W}{\partial y} = \frac{L_y - P_y}{f_m}, \quad (4.1)$$

donde  $f_m$  es la distancia focal de las microlentes ([Malacara, 2007](#)). Estas relaciones permiten medir las pendientes del frente de onda y, en última instancia, reconstruirlo.

Cuando el frente de onda presenta aberraciones, las posiciones de los spots en el detector se desplazan de su ubicación ideal. La Figura 4.2 muestra un ejemplo de un frente de onda aberrado al pasar por un sensor Shack-Hartmann, donde los spots presentan desviaciones significativas.



**Figura 4.2:** Frente de onda aberrado pasando por un sensor Shack-Hartmann. Se observan desplazamientos en los spots formados en el detector debido a las aberraciones ópticas presentes.

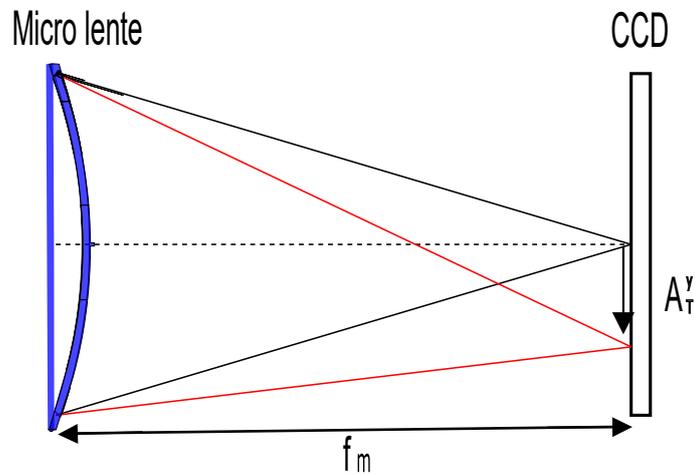
## 4.2. Aberración Transversal y la Derivada del Frente de Onda

Como se describió en la Sección 3.6.2, las aberraciones ópticas provocan desviaciones en la trayectoria de los rayos de luz. En el contexto del sensor Shack-Hartmann, estas desviaciones se manifiestan como aberraciones transversales, observables como desplazamientos en las posiciones de los *spots* generados en el detector. La relación entre estas aberraciones transversales ( $A_T^x, A_T^y$ ) y las derivadas del frente de onda está dada por:

$$\frac{\partial W}{\partial x} = \frac{-A_T^x}{f_m}, \quad \frac{\partial W}{\partial y} = \frac{-A_T^y}{f_m}, \quad (4.2)$$

donde  $\frac{\partial W}{\partial x}$  y  $\frac{\partial W}{\partial y}$  son las pendientes del frente de onda (Rayces, 1964b).

La Figura 4.3 muestra un ejemplo visual de la aberración transversal, donde los spots en el detector del sensor Shack-Hartmann presentan desplazamientos característicos debido a estas desviaciones (Novak et al., 2007).



**Figura 4.3:** Ejemplo de aberración transversal en un sensor Shack-Hartmann. Se observan desplazamientos característicos en los spots generados en el detector.

Estas pendientes proporcionan información directa sobre la forma del frente de onda y las aberraciones presentes. Por ejemplo, aberraciones como coma y astigmatismo se manifiestan como patrones característicos en las distribuciones de los spots. Medir y analizar estas desviaciones es esencial para interpretar las aberraciones y desarrollar correcciones ópticas adecuadas (Southwell, 1980).

### 4.3. Ajuste del Frente de Onda

En consecuencia, el ajuste del frente de onda constituye el proceso mediante el cual se reconstruye un frente de onda aberrado a partir de las mediciones realizadas con el sensor Shack-Hartmann. Esta reconstrucción se realiza mediante una expansión en polinomios de Zernike, descritos en la Sección 3.7,

los cuales forman una base ortogonal sobre el disco unitario y son ideales para describir aberraciones ópticas (Chai et al., 2024).

La representación del frente de onda en términos de polinomios de Zernike se expresa como:

$$W(x_j, y_j) = \sum_{i=1}^N A_i Z_i(x_j, y_j), \quad (4.3)$$

donde  $Z_i(x_j, y_j)$  son los polinomios de Zernike evaluados en las coordenadas cartesianas  $(x_j, y_j)$  del plano pupilar, y  $A_i$  son los coeficientes que representan la contribución de cada polinomio al frente de onda (Noll, 1976). Estos coeficientes proporcionan información detallada sobre las aberraciones presentes en el sistema óptico, como coma, astigmatismo o aberración esférica.

El número de términos  $N$  en la expansión de Zernike se elige para balancear la precisión del modelo y la factibilidad computacional. En esta tesis se emplearon los 37 términos descritos en la Tabla 3.2, lo que permite capturar una amplia variedad de aberraciones manteniendo un costo computacional razonable (Chai et al., 2024).

## 4.4. Sistema de Ecuaciones para el Ajuste del Frente de Onda

La construcción de las matrices  $Z$  y  $W$  es un paso clave para el ajuste del frente de onda. Estas matrices encapsulan la información necesaria para resolver el sistema de ecuaciones que permite determinar los coeficientes  $A_i$  de los polinomios de Zernike, descritos en la Sección 4.3.

La matriz  $Z$  se compone de las derivadas parciales de los polinomios de Zernike con respecto a  $x$  y  $y$ , evaluadas en las posiciones  $(P_{xi}, P_{yi})$  correspondientes a las subaperturas del sensor Shack-Hartmann. Cada columna está asociada a un término  $A_n$  de la expansión de Zernike, mientras que cada fila contiene las derivadas evaluadas en un punto específico:

$$Z = \begin{bmatrix} \frac{\partial Z(A_0, P_{x0}, P_{y0})}{\partial x} & \dots & \frac{\partial Z(A_n, P_{x0}, P_{y0})}{\partial x} \\ \vdots & \ddots & \vdots \\ \frac{\partial Z(A_0, P_{xn}, P_{yn})}{\partial x} & \dots & \frac{\partial Z(A_n, P_{xn}, P_{yn})}{\partial x} \\ \frac{\partial Z(A_0, P_{x0}, P_{y0})}{\partial y} & \dots & \frac{\partial Z(A_n, P_{x0}, P_{y0})}{\partial y} \\ \vdots & \ddots & \vdots \\ \frac{\partial Z(A_0, P_{xn}, P_{yn})}{\partial y} & \dots & \frac{\partial Z(A_n, P_{xn}, P_{yn})}{\partial y} \end{bmatrix}. \quad (4.4)$$

Esta matriz se construye utilizando las derivadas de los polinomios de Zernike y las mediciones del sensor, lo que permite relacionar las pendientes locales del frente de onda con las contribuciones de cada término de la expansión.

La matriz  $W$  contiene las pendientes locales del frente de onda, calculadas utilizando las posiciones de los spots procesados. Las derivadas parciales del frente de onda se organizan en un vector columna:

$$W = \begin{bmatrix} \frac{\partial W(P_{x0}, P_{y0})}{\partial x} \\ \vdots \\ \frac{\partial W(P_{xn}, P_{yn})}{\partial x} \\ \frac{\partial W(P_{x0}, P_{y0})}{\partial y} \\ \vdots \\ \frac{\partial W(P_{xn}, P_{yn})}{\partial y} \end{bmatrix}. \quad (4.5)$$

Las derivadas se obtienen mediante la Ecuación (4.1), relacionando directamente las posiciones medidas en el sensor con las pendientes del frente de onda.

Una vez construidas las matrices  $Z$  y  $W$ , el ajuste del frente de onda se realiza resolviendo el sistema de ecuaciones:

$$Z \cdot A = W, \quad (4.6)$$

donde  $A$  es el vector columna de coeficientes de los polinomios de Zernike. Este sistema constituye la base para la reconstrucción del frente de onda aberrado.

En este trabajo, exploramos tres métodos para resolver este sistema de ecuaciones: mínimos cuadrados, un enfoque basado en gradiente descendente

y el uso de una red hiperdinámica. A continuación, se describen en detalle cada uno de estos métodos, evaluando sus ventajas, limitaciones y aplicabilidad al ajuste de frentes de onda en sistemas ópticos.

## 4.5. Mínimos Cuadrados

El método de mínimos cuadrados es una técnica fundamental para resolver sistemas de ecuaciones sobredeterminados, donde el número de ecuaciones es mayor que el número de incógnitas. En el contexto del ajuste de frentes de onda, este método permite estimar los coeficientes de los polinomios de Zernike que describen las aberraciones ópticas, minimizando el error entre las derivadas medidas del frente de onda y las derivadas predichas por el modelo.

El problema se basa en el sistema matricial descrito en la Sección 4.3, representado por la Ecuación (4.6), donde  $Z$  es la matriz que encapsula las derivadas parciales de los polinomios de Zernike,  $A$  es el vector de coeficientes que se busca determinar, y  $W$  es el vector que contiene las derivadas medidas del frente de onda. Este sistema establece la base para el ajuste, minimizando la diferencia entre los valores medidos y los predichos por el modelo.

El método de mínimos cuadrados busca encontrar  $A$  de manera que el error cuadrático medio sea mínimo, definiendo este error como la norma cuadrática del residual:

$$\text{Error Residual} = \|Z \cdot A - W\|^2. \quad (4.7)$$

La minimización de este error lleva a la resolución del sistema de ecuaciones normal asociado:

$$(Z^T Z) \cdot A = Z^T W, \quad (4.8)$$

donde  $Z^T$  es la transpuesta de  $Z$ . Si  $Z^T Z$  es invertible, la solución para  $A$  se expresa como:

$$A = (Z^T Z)^{-1} Z^T W. \quad (4.9)$$

El método de mínimos cuadrados representa una herramienta clásica y ampliamente utilizada para el ajuste del frente de onda, ofreciendo ventajas como su simplicidad y la existencia de soluciones analíticas directas en sistemas bien condicionados. No obstante, en la práctica, su desempeño puede

verse comprometido en escenarios donde las mediciones están afectadas por ruido, valores atípicos o datos incompletos (Noel et al., 2023). La sensibilidad a valores atípicos, en particular, puede provocar que el modelo ajuste aberraciones inexistentes, lo que deteriora la precisión general. Asimismo, la dependencia del método en la pseudoinversa lo hace vulnerable ante matrices mal condicionadas, como suele ocurrir en sensores Shack-Hartmann con pérdida parcial de datos (Barakat and Newsam, 1980; Noel et al., 2023).

Estas limitaciones se vuelven especialmente críticas en aplicaciones ópticas donde el frente de onda cambia rápidamente, como en presencia de turbulencia atmosférica, desalineaciones ópticas o ruido del sensor (Pérez et al., 2024; Guo et al., 2022). En estos entornos tiempo-dependientes, como en sistemas de óptica adaptativa, el método de mínimos cuadrados carece de mecanismos para adaptarse dinámicamente a nuevas condiciones, ya que trata cada conjunto de mediciones de forma independiente y no aprovecha la información histórica para mejorar progresivamente sus estimaciones (Herrmann, 1980). Esta falta de adaptabilidad puede reducir su capacidad para ofrecer una reconstrucción precisa y estable en condiciones variables.

Frente a estas limitaciones, proponemos el uso de enfoques alternativos basados en aprendizaje supervisado, los cuales permiten ajustar modelos a partir de ejemplos previos y mejorar su rendimiento en condiciones variables (Guo et al., 2022). Este tipo de métodos ofrece una mayor flexibilidad ante mediciones incompletas o condiciones cambiantes, ya que pueden aprender patrones temporales y optimizar sus estimaciones mediante estrategias iterativas (Fowler and Landman, 2023b). En la siguiente sección se explorará un enfoque representativo de esta categoría: el método de gradiente descendente, el cual permite aproximar los coeficientes de aberración minimizando el error de forma eficiente, incluso cuando las condiciones de medición no son ideales.

## 4.6. Gradiente Descendente con Optimizador Adam

Como alternativa a los métodos analíticos, el gradiente descendente ofrece un enfoque iterativo para minimizar el error entre las predicciones del modelo  $Z \cdot A$  y las mediciones reales  $W$ . Este método ajusta los coeficientes  $A$  de forma progresiva para optimizar una función de pérdida, en este caso el error absoluto medio (MAE, por sus siglas en inglés). A diferencia de los mínimos cuadrados, el gradiente descendente permite trabajar con datos incompletos

o ambiguos, reduciendo el impacto de los valores atípicos en el ajuste del frente de onda.

El proceso de gradiente descendente se basa en la evaluación del gradiente de la función de pérdida  $J(A)$ , que describe la dirección y magnitud del cambio necesario en los coeficientes para minimizar el error:

$$\nabla J(A) = \left( \frac{\partial J}{\partial A_1}, \frac{\partial J}{\partial A_2}, \dots, \frac{\partial J}{\partial A_n} \right), \quad (4.10)$$

donde  $\nabla J(A)$  representa el vector de derivadas parciales de  $J(A)$  respecto a cada coeficiente  $A_i$ . Los coeficientes se actualizan iterativamente en la dirección opuesta al gradiente de la función de pérdida:

$$A_{\text{new}} = A_{\text{previous}} - \alpha \nabla J(A), \quad (4.11)$$

donde  $\alpha$  es la tasa de aprendizaje, un hiperparámetro que determina el tamaño del paso en cada iteración. Una selección adecuada de este hiperparámetro es crucial para garantizar la convergencia del modelo.

Aunque el gradiente descendente estándar es efectivo, puede enfrentar dificultades en problemas donde los gradientes son extremadamente pequeños o grandes, como ocurre en este caso debido a las características de las pantallas de microlentes en el sensor Shack-Hartmann. Estas pantallas generan desplazamientos mínimos en las posiciones de los spots, lo que resulta en gradientes muy pequeños y dificulta la convergencia eficiente del algoritmo. Para abordar estas limitaciones, se utiliza el optimizador Adam ("*Adaptive Moment Estimation*"), que adapta dinámicamente la tasa de aprendizaje basándose en el historial acumulado de los gradientes y sus cuadrados (Kingma and Ba, 2014). Adam introduce dos momentos clave: el momento de primer orden  $m_t$ , que acumula los gradientes, y el momento de segundo orden  $v_t$ , que acumula los cuadrados de los gradientes:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla L(A^{(t)}), \quad (4.12)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla L(A^{(t)}))^2, \quad (4.13)$$

donde  $\beta_1$  y  $\beta_2$  son factores de decaimiento exponencial. Ambos momentos se corrigen para compensar su inicialización en cero:

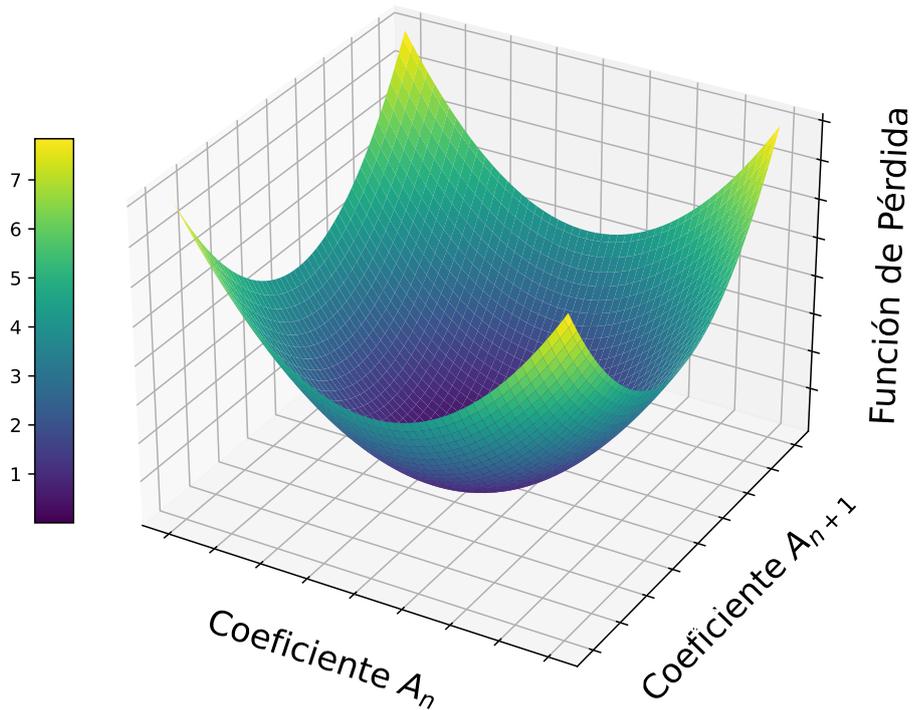
$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}. \quad (4.14)$$

Finalmente, los coeficientes  $A$  se actualizan según:

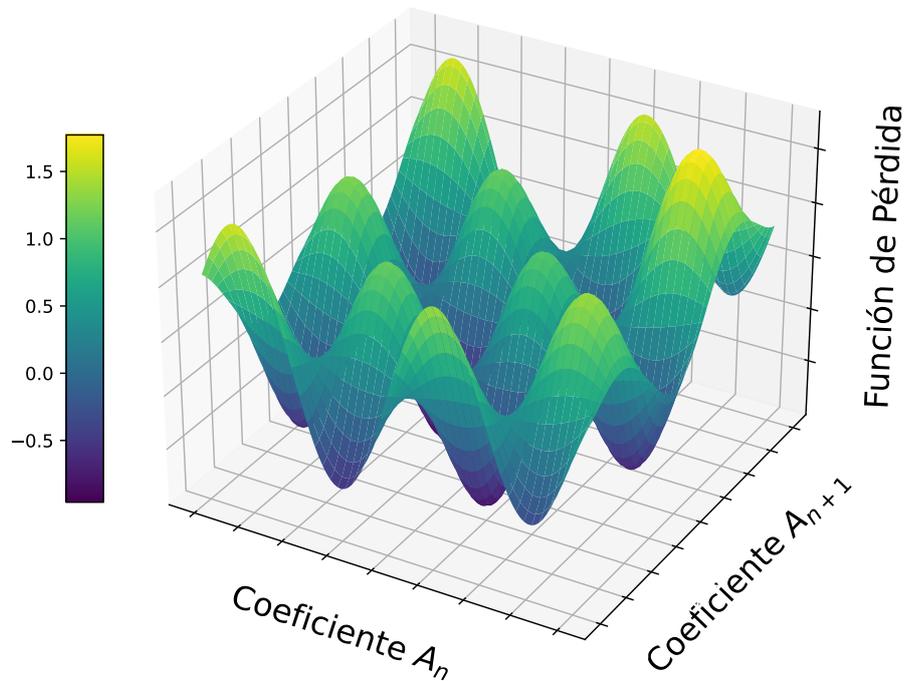
$$A^{(t+1)} = A^{(t)} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}, \quad (4.15)$$

donde  $\eta$  es la tasa de aprendizaje, y  $\epsilon$  es un hiperparámetro pequeño para estabilidad numérica.

Para visualizar mejor el tipo de problemas que enfrentamos al utilizar gradiente descendente, a continuación se muestran dos ejemplos de funciones de pérdida:



**Figura 4.4:** Representación de superficie de la Función de Pérdida con Mínimo Global Único.



**Figura 4.5:** Representación de superficie de la Función de Pérdida con Múltiples Mínimos Locales.

En la Figura 4.4, se muestra un escenario con un único mínimo global, lo que permite que el algoritmo de gradiente descendente encuentre la solución óptima de manera eficiente. Sin embargo, en la Figura 4.5, observamos un problema más complejo con múltiples mínimos locales, lo que dificulta que el gradiente descendente encuentre el mínimo global. Este tipo de superficies son comunes en problemas no convexos y destacan la importancia de utilizar optimizadores avanzados como Adam, que ayudan a evitar caer en mínimos locales.

El optimizador Adam mejora significativamente la eficiencia y precisión en comparación con el gradiente descendente estándar (Keskar and Socher, 2017). Esto lo convierte en una herramienta adecuada para problemas como el ajuste de frentes de onda, donde las mediciones pueden contener ruido significativo o variaciones en los gradientes. Además, Adam permite trabajar con datos incompletos o ambiguos sin comprometer la precisión.

Para cerrar esta sección, cabe destacar que, si bien el gradiente descendente ofrece una solución eficiente, existen alternativas más flexibles basadas en aprendizaje profundo e inteligencia artificial generativa. En la siguiente sección se explorará una de estas propuestas, que permite una mayor adaptabilidad en escenarios complejos y variables para el ajuste de frentes de onda.

## 4.7. Red Hiperdinámica

Como último método, dada la complejidad del problema, decidimos crear una combinación de redes neuronales que permita una mayor flexibilidad y capacidad de adaptación. Este enfoque, al que denominamos red hiperdinámica, consiste en una arquitectura compuesta por dos redes: una red auxiliar generativa y una red primaria. La red auxiliar se encarga de generar dinámicamente los coeficientes  $A$ , que representan los parámetros de aberración, mientras que la red primaria se encarga de producir las mediciones  $W$  predichas. Este enfoque nos permite obtener los coeficientes deseados de manera indirecta, utilizando los pesos generados por la red auxiliar, y al mismo tiempo comparar las mediciones predichas con los valores reales para evaluar el desempeño del modelo.

Una red neuronal simple es un modelo matemático inspirado en el funcionamiento de las neuronas biológicas. Consiste en varias capas: una capa de entrada, una o más capas ocultas y una capa de salida. Cada neurona en una capa está conectada a las neuronas de la siguiente capa mediante pesos (Haykin, 1998).

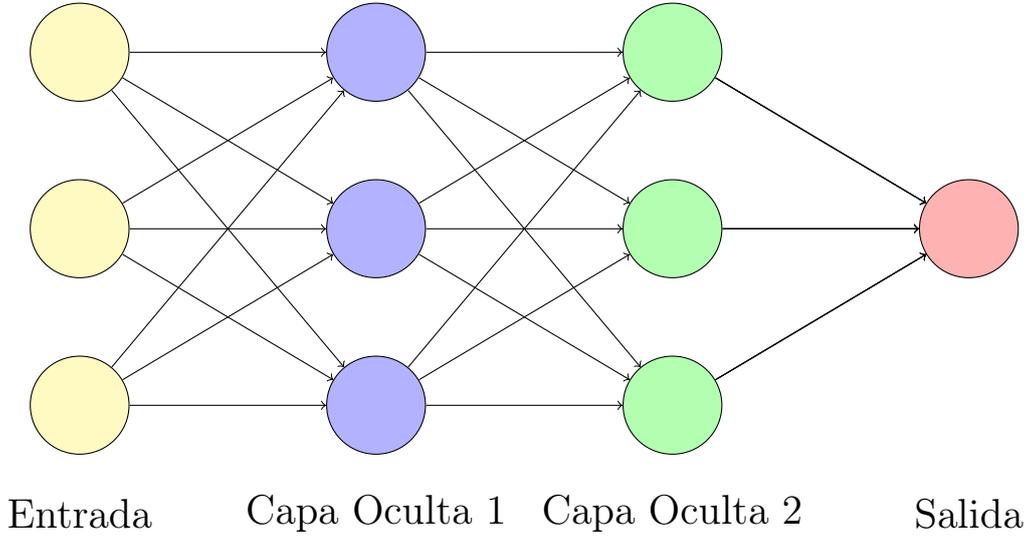
La salida de una neurona individual se calcula como:

$$y = f \left( \sum_{i=1}^n w_i x_i + b \right), \quad (4.16)$$

donde  $x_i$  representa las entradas,  $w_i$  los pesos asociados a cada entrada,  $b$  es el término de sesgo, y  $f(\cdot)$  corresponde a la función de activación, la cual introduce no linealidad en la salida de la neurona.

El proceso de entrenamiento de una red neuronal consiste en ajustar los pesos  $w_i$  y los sesgos  $b$  para minimizar una función de pérdida  $\mathcal{L}$ , la cual mide la diferencia entre los valores predichos y los valores reales.

El siguiente diagrama muestra una red neuronal simple:



**Figura 4.6:** Diagrama de una red neuronal simple con dos capas ocultas.

En nuestro problema, el objetivo es ajustar coeficientes de aberraciones de frente de onda utilizando datos de sensores Shack-Hartmann. Aplicar directamente una red neuronal simple puede ser insuficiente debido a la variabilidad y complejidad de los datos de entrada. Para abordar esta situación, proponemos una variante de las *hypernetworks* (Duvall et al., 2024; Chauhan et al., 2024), a la que denominamos red hiperdinámica, compuesta por dos elementos principales: la red primaria y la red auxiliar generativa.

La red auxiliar generativa toma los datos de entrada y produce los coeficientes  $A_{gen}$  generados de manera indirecta:

$$A_{gen} = G(Z, W; \phi), \quad (4.17)$$

donde  $G$  es la red auxiliar generativa parametrizada por  $\phi$ .

La red auxiliar tiene como función principal extraer las características de los polinomios de Zernike, identificando patrones y relaciones útiles entre las matrices de entrada y las mediciones. Esta red busca generar de manera dinámica los coeficientes  $A_{gen}$  que mejor representen el comportamiento del frente de onda, permitiendo que la red primaria pueda hacer predicciones más precisas. Al generar los coeficientes de forma adaptativa según las condiciones de entrada, la red auxiliar mejora significativamente la capacidad de generalización del modelo y facilita el ajuste de aberraciones.

La red primaria, utilizando los coeficientes generados por la red auxiliar,

junto con las matrices  $Z$  y  $W$ , calcula las mediciones  $W_{pred}$  mediante:

$$W_{pred} = P(Z, A_{gen}, W; \psi), \quad (4.18)$$

donde  $P$  es la red primaria parametrizada por  $\psi$ .

La red primaria combina los coeficientes generados  $A_{gen}$  con la matriz  $Z$  y utiliza las mediciones reales  $W$  para producir las mediciones predichas  $W_{pred}$ . Esta red actúa como un mecanismo de validación, ya que las mediciones predichas se comparan directamente con los valores reales para evaluar la precisión del modelo. Además, permite realizar un ajuste fino del sistema, asegurando que las predicciones se mantengan dentro de un rango aceptable de error, lo que es crucial en nuestro problema.

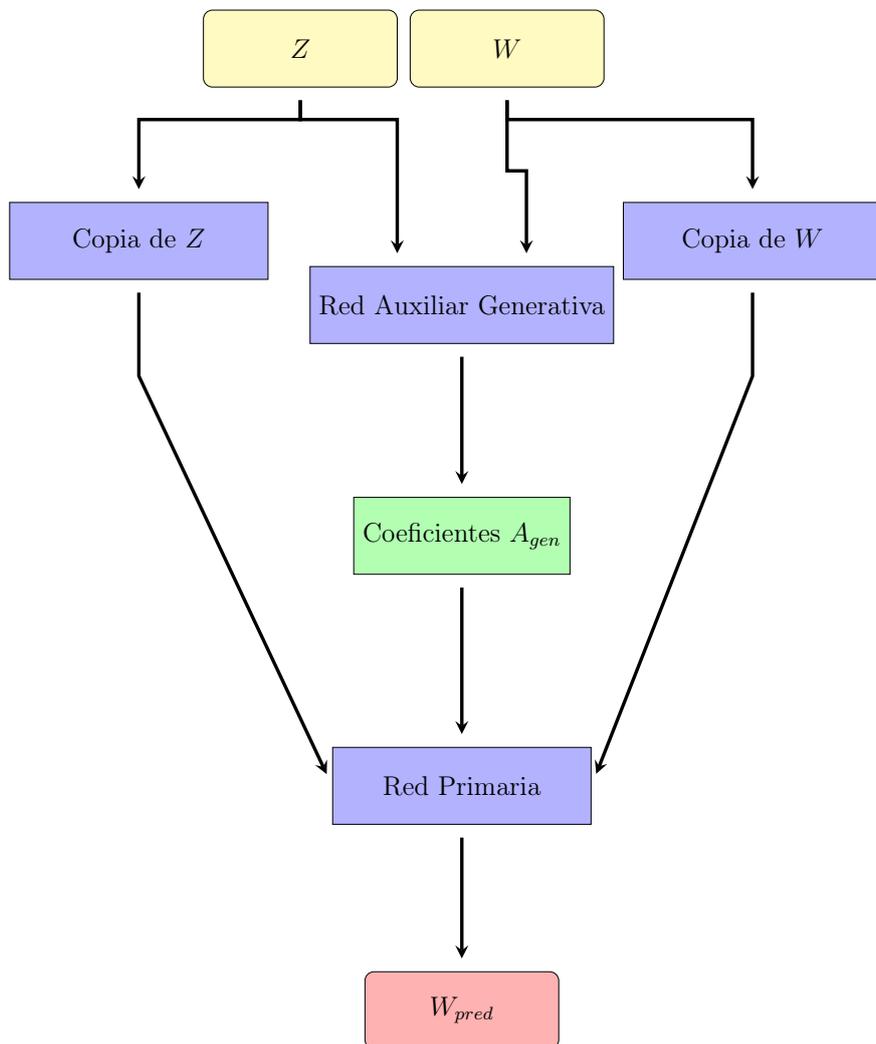
Matemáticamente, la red hiperdinámica completa se modela como:

$$W_{pred} = P(Z, G(Z, W; \phi), W; \psi), \quad (4.19)$$

donde:

- $G(Z, W; \phi)$  es la red auxiliar generativa parametrizada por  $\phi$ , que produce los coeficientes  $A_{gen}$ ,
- $P(Z, A_{gen}, W; \psi)$  es la red primaria parametrizada por  $\psi$ , que utiliza los coeficientes  $A_{gen}$  generados por  $G$ , junto con las matrices  $Z$  y  $W$ , para calcular las mediciones  $W_{pred}$ ,
- $Z$  es la matriz de derivadas parciales de los polinomios de Zernike,
- $W$  es el vector que contiene las mediciones reales de pendientes del frente de onda,
- $W_{pred}$  son las mediciones predichas por la red primaria.

La arquitectura de la red hiperdinámica permite obtener los coeficientes de aberración de manera indirecta y validar las predicciones mediante la comparación de las mediciones  $W$  con los valores reales obtenidos de los sensores. El siguiente diagrama muestra la arquitectura propuesta de la red hiperdinámica, donde las matrices  $Z$  y  $W$  se utilizan tanto en la red auxiliar generativa como en la red primaria, incluyendo copias de las matrices originales:



**Figura 4.7:** Arquitectura de la red hiperdinámica, mostrando cómo las matrices  $Z$  y  $W$  se utilizan en la red auxiliar generativa para producir los coeficientes  $A_{gen}$ , y cómo sus copias se emplean en la red primaria para calcular las mediciones predichas  $W_{pred}$ .

El entrenamiento de la red hiperdinámica implica ajustar los parámetros  $\theta$ ,  $\phi$  y  $\psi$  para minimizar una función de pérdida que mide la discrepancia entre los valores predichos  $W_{pred}$  y los valores reales  $W$  utilizados por la red primaria. Aunque la función de pérdida se aplica sobre  $W_{pred}$ , este proceso indirectamente ajusta y mejora los coeficientes  $A$  generados por la red auxiliar generativa. La función de pérdida utilizada es el error absoluto medio (MAE):

$$L(W_{\text{pred}}, W) = \frac{1}{n} \sum_{i=1}^n |W_{\text{pred},i} - W_i|. \quad (4.20)$$

El algoritmo Adam se emplea únicamente para optimizar los parámetros de la red hiperdinámica ( $\theta$ ,  $\phi$  y  $\psi$ ) utilizando momentos acumulados para mejorar la convergencia. Dado que el funcionamiento del optimizador Adam ya ha sido detallado en la Sección 4.15, aquí simplemente se aplica el mismo principio, utilizando una tasa de aprendizaje adaptativa basada en los momentos  $\hat{m}_t$  y  $\hat{v}_t$  descritos en las Ecuaciones (4.12) y (4.13).

Entre las ventajas potenciales de esta configuración de red hiperdinámica que proponemos, destaca su adaptabilidad, ya que ajusta los pesos en función de las entradas. Esto podría permitir un mejor rendimiento en condiciones variables y una reducción del costo computacional al evitar el cálculo explícito de soluciones analíticas. Asimismo, la posibilidad de reentrenarse con los resultados obtenidos ofrece la oportunidad de mejorar iterativamente la precisión del modelo. Sin embargo, este enfoque debe implementarse con cautela, ya que un reentrenamiento excesivo podría llevar al sobreajuste y comprometer la capacidad de generalización.

Dado su diseño, esta configuración podría mostrar mayor robustez frente a puntos atípicos en comparación con los métodos tradicionales, lo que la haría adecuada para problemas complejos como el ajuste de aberraciones ópticas. No obstante, presenta retos importantes, como la complejidad inherente a su arquitectura, la necesidad de grandes cantidades de datos para un entrenamiento eficaz y la dependencia de hardware especializado para asegurar su eficiencia.

Los tres enfoques presentados: mínimos cuadrados, gradiente descendente y la configuración de red hiperdinámica propuesta ofrecen alternativas complementarias para resolver el problema de ajuste de aberraciones ópticas. Sin embargo, tanto el gradiente descendente como la red hiperdinámica dependen críticamente de la correcta selección de hiperparámetros para lograr un rendimiento óptimo.

Dado que la correcta selección y análisis de hiperparámetros desempeña un papel fundamental en la efectividad de estos métodos, el siguiente capítulo explorará herramientas avanzadas para este propósito. Se analizarán las métricas utilizadas para evaluar el desempeño de los modelos, así como las técnicas más sofisticadas, como la optimización bayesiana y los corner plots, que permiten comprender y optimizar el comportamiento de los modelos propuestos.

## Capítulo 5

# Herramientas Avanzadas para la Selección de Hiperparámetros

En este capítulo se describen las herramientas avanzadas utilizadas para la selección de hiperparámetros y la evaluación del desempeño de los modelos propuestos. Se comienza con una explicación de la importancia de los hiperparámetros y su impacto crítico en el rendimiento de los métodos de ajuste. Posteriormente, se definen las métricas de evaluación que permitirán cuantificar de manera objetiva la calidad de las soluciones obtenidas y comparar la efectividad de los enfoques propuestos. A continuación, se delimita el espacio de búsqueda para garantizar una exploración eficiente y dirigida. Como técnica principal, se detalla la optimización bayesiana para la búsqueda de configuraciones óptimas de hiperparámetros. Finalmente, se introduce el uso de corner plots, una herramienta visual clave para analizar y comprender las interacciones entre los hiperparámetros, proporcionando así un marco integral para la selección y evaluación de configuraciones óptimas.

La selección adecuada de hiperparámetros es un paso esencial para garantizar el desempeño eficiente y preciso de los modelos computacionales. Los hiperparámetros controlan aspectos fundamentales del comportamiento de los modelos durante el entrenamiento, como la velocidad de convergencia, la regularización y la complejidad de la arquitectura. Como se describió en las secciones anteriores, en este trabajo se propone una combinación de métodos, que incluyen el enfoque clásico de mínimos cuadrados, el cual no depende de la selección de hiperparámetros, y dos enfoques basados en inteligencia artificial: el gradiente descendente y la red hiperdinámica. Debido a que estos últimos dependen críticamente de la optimización de sus hiperparámetros, este proceso adquiere una relevancia particular para garantizar su efectivi-

dad.

A lo largo de este capítulo, se describirán las herramientas avanzadas que se utilizarán para la selección de hiperparámetros, estableciendo una base teórica sólida que permitirá comprender su impacto en el rendimiento y facilitar su aplicación en la implementación y evaluación de los modelos.

## 5.1. Importancia de los Hiperparámetros

Como se discutió en las secciones anteriores, los hiperparámetros son valores predefinidos que controlan aspectos clave del comportamiento de un modelo computacional durante su entrenamiento, influyendo directamente en su rendimiento final. A diferencia de los parámetros internos del modelo, que son ajustados automáticamente mediante algoritmos como el gradiente descendente, los hiperparámetros deben definirse antes de iniciar el entrenamiento y, por lo general, no se optimizan directamente durante este proceso.

Los hiperparámetros desempeñan un papel fundamental en el problema de ajuste de frentes de onda, dado que la complejidad inherente del entorno óptico requiere configuraciones cuidadosamente seleccionadas para lograr un ajuste eficiente y preciso. Un mal ajuste de hiperparámetros puede resultar en diversos problemas, tales como:

- **Convergencia lenta:** Si la tasa de aprendizaje ( $\eta$ ) es demasiado baja, el modelo tardará muchas iteraciones en acercarse a una solución óptima.
- **Divergencia del modelo:** Si  $\eta$  es demasiado alta, el modelo puede oscilar alrededor del óptimo sin llegar a converger.
- **Sobreajuste:** Hiperparámetros como la regularización  $L_2$  ( $\lambda$ ) influyen en la capacidad del modelo para generalizar. Si la regularización es insuficiente, el modelo puede ajustarse demasiado a los datos de entrenamiento, perdiendo capacidad predictiva en datos nuevos.
- **Subajuste:** Ocurre cuando el modelo no tiene suficiente capacidad para capturar las relaciones subyacentes en los datos, lo que suele deberse a una regularización excesiva o a una arquitectura demasiado simple.

A continuación, se describen los hiperparámetros específicos de los dos modelos basados en inteligencia artificial propuestos en este trabajo: el gradiente descendente con optimizador Adam y la red hiperdinámica.

### 5.1.1. Hiperparámetros del Gradiente Descendente

En el modelo de gradiente descendente con optimizador Adam, los principales hiperparámetros a considerar son los siguientes:

- **Tasa de aprendizaje (LR):** Controla el tamaño del paso que el modelo da en cada iteración durante el proceso de optimización. Una tasa de aprendizaje demasiado baja puede resultar en una convergencia lenta, mientras que una tasa demasiado alta puede causar inestabilidad y divergencia.
- **Coefficientes de momento ( $\beta_1$  y  $\beta_2$ ):** Estos parámetros regulan la acumulación de los gradientes y sus cuadrados a lo largo del tiempo, influyendo directamente en la velocidad y estabilidad de la convergencia.
- **Término de estabilidad numérica ( $\epsilon$ ):** Este valor pequeño se suma al denominador en el cálculo de las actualizaciones para evitar divisiones por cero, siendo crucial en problemas con gradientes muy pequeños.
- **Regularización  $L_2$  (L2 Reg):** Este término penaliza los pesos grandes en el modelo para reducir el riesgo de sobreajuste, equilibrando la capacidad del modelo para generalizar.
- **Número de iteraciones (Iters):** Determina cuántas actualizaciones realizará el optimizador. Un número demasiado bajo puede llevar a una solución incompleta, mientras que un número excesivo puede aumentar el costo computacional sin mejoras significativas.
- **Inicialización del kernel (Init):** Define cómo se inicializan los valores iniciales de los pesos en el modelo, lo que puede influir en la estabilidad y velocidad de la convergencia.

La actualización de los parámetros del modelo usando Adam se realiza mediante la siguiente ecuación:

$$\theta^{(t+1)} = \theta^{(t)} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}, \quad (5.1)$$

donde  $\hat{m}_t$  y  $\hat{v}_t$  son los momentos corregidos calculados durante el entrenamiento.

Estos hiperparámetros permiten controlar tanto la estabilidad como la eficiencia del entrenamiento, siendo su selección crítica para obtener buenos resultados.

### 5.1.2. Hiperparámetros de la Red Hiperdinámica

En el caso de la red hiperdinámica, los hiperparámetros juegan un papel aún más relevante debido a la complejidad inherente de la arquitectura. Los principales hiperparámetros incluyen:

- **Tasa de aprendizaje (LR):** Controla la magnitud de las actualizaciones durante el entrenamiento. Una tasa de aprendizaje demasiado baja puede resultar en una convergencia lenta, mientras que una tasa demasiado alta puede causar inestabilidad.
- **Regularización  $L_2$  (L2 Reg):** Penaliza los valores grandes de los pesos para reducir el riesgo de sobreajuste, equilibrando la capacidad del modelo para generalizar.
- **Tasa de abandono (DR):** Ayuda a prevenir el sobreajuste desactivando aleatoriamente un porcentaje de unidades durante el entrenamiento.
- **Tamaño del lote (UA):** Define cuántos ejemplos se usan para calcular cada actualización de los pesos. Tamaños pequeños permiten mayor precisión en las actualizaciones, pero incrementan el tiempo de entrenamiento.
- **Número de capas adicionales (NL):** Incrementar el número de capas permite modelar relaciones más complejas, pero también incrementa el costo computacional y la dificultad para entrenar el modelo.
- **Número de neuronas en la red madre ( $U_m$ ):** Controla la capacidad de la red madre para procesar representaciones complejas.
- **Número de neuronas en la red auxiliar ( $U_a$ ):** Influye directamente en la capacidad de la red auxiliar para aproximar las soluciones requeridas.
- **Función de activación (Act):** Determina cómo se transforman las entradas en cada capa. Funciones como ReLU y tanh son comunes y afectan tanto la convergencia como la capacidad de la red para modelar relaciones no lineales.
- **Inicialización del kernel (Init):** Define cómo se inicializan los valores de los pesos antes del entrenamiento, afectando la estabilidad y velocidad de convergencia.

La red hiperdinámica utiliza las entradas  $Z$  y  $W$ , produce los coeficientes  $A_{\text{gen}}$  de manera indirecta mediante la Ecuación (4.17). Este proceso está condicionado por los hiperparámetros que controlan la arquitectura y el en-

trenamiento de la red auxiliar, los cuales afectan directamente la capacidad del modelo para generar soluciones precisas. Además, la red principal, que utiliza los coeficientes  $A_{\text{gen}}$  junto con las matrices  $Z$  y  $W$  para calcular las mediciones  $W_{\text{pred}}$  según la Ecuación (4.18), también depende de hiperparámetros que influyen en su rendimiento. Por lo tanto, la selección adecuada de los hiperparámetros en ambas redes es esencial para garantizar el desempeño óptimo del modelo completo.

Dado que tanto el gradiente descendiente como la red hiperdinámica dependen críticamente de una correcta selección de estos hiperparámetros, es esencial emplear técnicas avanzadas para explorar el espacio de configuraciones de manera eficiente, minimizando el costo computacional y maximizando el rendimiento. En los próximos apartados se describen estas herramientas, comenzando con la definición del espacio de búsqueda.

## 5.2. Definición del Espacio de Búsqueda

Matemáticamente, el espacio de búsqueda se puede definir como un conjunto  $\mathcal{H} \subseteq \mathbb{R}^d$ , donde  $d$  es el número de hiperparámetros a optimizar, y cada hiperparámetro  $h_i \in \mathcal{H}$ , tiene un rango de valores permitido  $[h_i^{\min}, h_i^{\max}]$ . Así, el espacio de búsqueda completo se expresa como:

$$\mathcal{H} = [h_1^{\min}, h_1^{\max}] \times [h_2^{\min}, h_2^{\max}] \times \dots \times [h_d^{\min}, h_d^{\max}]. \quad (5.2)$$

La correcta delimitación de este espacio es fundamental para garantizar que las configuraciones evaluadas sean representativas, relevantes y computacionalmente factibles. En este trabajo, la selección de los rangos para cada hiperparámetro se fundamenta en dos aspectos principales: el impacto del hiperparámetro en el modelo y la necesidad de balancear la amplitud del espacio de búsqueda con los recursos computacionales disponibles.

La definición del espacio de búsqueda debe cumplir con los siguientes principios generales:

- **Relevancia:** Los valores incluidos en el espacio de búsqueda deben cubrir configuraciones potencialmente óptimas para los modelos, evitando rangos demasiado amplios que incluyan configuraciones improbables o ineficaces.
- **Equilibrio entre exploración y explotación:** Es necesario permitir suficiente exploración de diferentes configuraciones, pero sin sacrificar la posibilidad de explotar regiones que presentan mejores resultados.

- **Interacción entre hiperparámetros:** La selección de los rangos debe considerar cómo interactúan los hiperparámetros entre sí, por ejemplo, cómo la tasa de aprendizaje interactúa con el tamaño del lote o cómo la regularización afecta la profundidad de la red.

### 5.2.1. Selección del Espacio de Búsqueda para el Gradiente Descendente

En el modelo de gradiente descendente con optimizador Adam, los rangos de los hiperparámetros se seleccionarán considerando su impacto directo en el rendimiento del modelo. Algunos criterios clave incluyen:

- La **tasa de aprendizaje (LR)** debe equilibrar la velocidad de convergencia con la estabilidad del entrenamiento. Los rangos serán elegidos para garantizar que el modelo pueda converger eficientemente sin riesgo de oscilaciones.
- Los **coeficientes de momento** ( $\beta_1$  y  $\beta_2$ ) se definirán en rangos que permitan una acumulación efectiva de gradientes y suavicen las actualizaciones, asegurando una optimización estable.
- La **regularización  $L_2$  (L2 Reg)** estará orientada a encontrar un balance entre sobreajuste y subajuste, evitando que los pesos crezcan excesivamente mientras se preserva la capacidad de aprendizaje del modelo.
- Otros hiperparámetros, como el **término de estabilidad numérica** ( $\epsilon$ ) y la **inicialización del kernel (Init)**, serán seleccionados para prevenir problemas numéricos o inestabilidad en las primeras etapas del entrenamiento.

### 5.2.2. Selección del Espacio de Búsqueda para la Red Hiperdinámica

Debido a la mayor complejidad de la red hiperdinámica, la selección de los rangos para sus hiperparámetros requiere un enfoque más amplio, considerando tanto la estructura del modelo como los aspectos de regularización. Algunos puntos clave son:

- La **tasa de aprendizaje (LR)** debe permitir a la red explorar configuraciones con suficiente precisión para encontrar un ajuste adecuado en un espacio de solución complejo.

- Los parámetros estructurales, como el **número de capas adicionales (NL)** y el **número de neuronas en la red madre ( $U_m$ ) y auxiliar ( $U_a$ )**, serán definidos para garantizar que el modelo tenga suficiente capacidad para aprender relaciones complejas sin incurrir en un costo computacional prohibitivo.
- La **regularización  $L_2$  (L2 Reg)** y la **tasa de abandono (DR)** se elegirán para prevenir el sobreajuste, ajustando los niveles de penalización y de desactivación de neuronas.
- El **tamaño del lote (UA)** será seleccionado con base en el balance entre precisión y eficiencia, considerando la naturaleza de los datos y los recursos disponibles.
- La **función de activación (Act)** y la **inicialización del kernel (Init)** se definirán para garantizar que la red comience en un estado estable y pueda aprender relaciones no lineales de manera efectiva.

Aunque la selección de los rangos puede beneficiarse de criterios matemáticos y experimentales, en muchos casos, depende también de la experiencia del desarrollador, quien ajusta estos valores según el contexto del problema y las limitaciones prácticas. Este enfoque flexible permite construir un espacio de búsqueda que maximice las posibilidades de encontrar configuraciones óptimas sin comprometer la eficiencia computacional.

### 5.3. Ajuste de Hiperparámetros mediante Optimización Bayesiana

Como se definió en la sección anterior, el espacio de búsqueda  $\mathcal{H}$  delimita los posibles valores de los hiperparámetros para los modelos de gradiente descendente y red hiperdinámica. Sin embargo, encontrar la configuración óptima dentro de este espacio representa un desafío significativo debido a la alta dimensionalidad y las interacciones entre los hiperparámetros. La optimización bayesiana surge como una herramienta eficaz para abordar este problema, permitiendo una exploración eficiente del espacio de búsqueda mediante un modelo probabilístico que guía el proceso de selección (Xu et al., 2024; Tani and Veelken, 2024).

El objetivo principal de la optimización bayesiana es minimizar la función de pérdida  $f(\mathbf{x})$ , que depende de un conjunto de hiperparámetros  $\mathbf{x} \in \mathcal{H}$ . Para ello, se utiliza un proceso gaussiano ( $\mathcal{GP}$ ) como modelo probabilístico de  $f(\mathbf{x})$ , definido por una función de media  $\mu(\mathbf{x})$  y una función de covarianza

$\sigma^2(\mathbf{x})$ :

$$f(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), \sigma^2(\mathbf{x})). \quad (5.3)$$

Este modelo captura tanto la predicción del valor esperado de  $f(\mathbf{x})$  en un punto dado  $\mathbf{x}$ , como la incertidumbre asociada a esa predicción, representada por  $\sigma^2(\mathbf{x})$ . A medida que se evalúan nuevos conjuntos de hiperparámetros, el modelo probabilístico se actualiza, reduciendo la incertidumbre en regiones previamente exploradas del espacio  $\mathcal{H}$  y mejorando las predicciones en puntos no evaluados.

Para determinar el siguiente conjunto de hiperparámetros a evaluar, se utiliza una función de adquisición. Una de las más utilizadas es el *Expected Improvement (EI)*, que calcula la mejora esperada en la función objetivo con respecto al mejor valor observado hasta el momento  $f(\mathbf{x}^+)$ :

$$EI(\mathbf{x}) = \mathbb{E} [\text{máx}(0, f(\mathbf{x}^+) - f(\mathbf{x}))]. \quad (5.4)$$

El siguiente punto  $\mathbf{x}_{n+1}$  en el espacio de búsqueda se selecciona maximizando  $EI(\mathbf{x})$ :

$$\mathbf{x}_{n+1} = \underset{\mathbf{x} \in \mathcal{H}}{\text{argmax}} EI(\mathbf{x}). \quad (5.5)$$

Este criterio equilibra la exploración de nuevas regiones del espacio, donde  $\sigma^2(\mathbf{x})$  es alta, con la explotación de áreas prometedoras que ya han mostrado buenos resultados, donde  $\mu(\mathbf{x})$  sugiere mejoras potenciales.

La ventaja principal de la optimización bayesiana radica en su capacidad para reducir significativamente el número de evaluaciones necesarias al priorizar configuraciones prometedoras (Xu et al., 2024). Esto es particularmente relevante en este trabajo, ya que cada evaluación de  $f(\mathbf{x})$  implica el entrenamiento completo de un modelo, lo cual es computacionalmente costoso. En problemas como el ajuste de hiperparámetros de la red hiperdinámica, donde la dimensionalidad del espacio de búsqueda es alta, la optimización bayesiana permite explorar  $\mathcal{H}$  de manera eficiente sin recurrir a enfoques exhaustivos como la búsqueda en cuadrícula.

La conexión entre la optimización bayesiana y la definición del espacio de búsqueda es clave para garantizar un proceso eficiente. Los límites  $[h_i^{\min}, h_i^{\max}]$  definidos previamente para cada hiperparámetro en  $\mathcal{H}$  no solo restringen el

rango de valores posibles, sino que también guían el modelo probabilístico hacia configuraciones prácticas y relevantes. Este balance entre exploración y explotación maximiza las probabilidades de encontrar configuraciones óptimas para ambos modelos propuestos, optimizando recursos y tiempo computacional.

A pesar de sus ventajas, la optimización bayesiana presenta un costo adicional asociado a la construcción y actualización iterativa del modelo probabilístico. Sin embargo, este costo se ve compensado por la reducción en el número de evaluaciones necesarias para encontrar hiperparámetros óptimos, especialmente en problemas complejos y computacionalmente intensivos como los abordados en este trabajo. En última instancia, este enfoque equilibra de manera efectiva la precisión del modelo con la eficiencia computacional, destacándose como una herramienta adecuada para el ajuste de hiperparámetros en los modelos propuestos. Con el modelo optimizado, el siguiente paso es evaluar su rendimiento mediante métricas que permitan comparar objetivamente los diferentes enfoques analizados.

## 5.4. Métricas de Evaluación

En este trabajo, las métricas de evaluación se seleccionaron con el objetivo de analizar el desempeño de los modelos desde diferentes perspectivas, siendo el Error Absoluto Medio (MAE) la métrica principal que define la función objetivo optimizada por los modelos. Las demás métricas se utilizan para complementar el análisis y proporcionar una visión integral del rendimiento.

La función objetivo de los modelos propuestos, basada en el MAE, se expresa como:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (5.6)$$

donde  $y_i$  representa los valores reales,  $\hat{y}_i$  son las predicciones del modelo, y  $n$  es el número total de datos. Esta métrica mide la magnitud promedio de los errores cometidos, sin amplificar el impacto de errores grandes, lo que la convierte en una elección adecuada para garantizar una convergencia estable durante el proceso de optimización.

Además del MAE, se utilizan las siguientes métricas complementarias para evaluar otros aspectos del rendimiento:

- **Coefficiente de determinación ( $R^2$ ):** Esta métrica evalúa la capacidad del modelo para explicar la variabilidad en los datos. Se define como:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (5.7)$$

donde  $\bar{y}$  es la media de los valores reales. Valores más cercanos a 1 indican un modelo con alta capacidad explicativa.

- **Error Cuadrático Medio (MSE):** Penaliza con mayor severidad los errores grandes, proporcionando información sobre la sensibilidad del modelo a dichas desviaciones. Se calcula como:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (5.8)$$

- **Raíz del Error Cuadrático Medio (RMS):** Ofrece una interpretación directa del error promedio en la escala de los datos, siendo útil para comparar modelos en diferentes contextos. Se define como:

$$RMS = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (5.9)$$

Estas métricas permiten analizar el rendimiento de los modelos desde distintas perspectivas: el  $R^2$  proporciona una medida de la capacidad explicativa; el MAE cuantifica la precisión promedio; el MSE evalúa la sensibilidad a errores grandes; y el RMS presenta una visión directa del error en términos de la escala de los datos.

La selección del MAE como función objetivo responde a la necesidad de optimizar una métrica que sea robusta frente a errores atípicos, especialmente en problemas con alta variabilidad en los datos. Por su parte, las métricas adicionales permiten identificar fortalezas y limitaciones de cada modelo, ofreciendo un análisis complementario y detallado de su desempeño.

Estas métricas se aplican para evaluar los tres enfoques presentados: el método de mínimos cuadrados, el gradiente descendente y la configuración de red hiperdinámica. Cada métrica ofrece una perspectiva específica del rendimiento, permitiendo analizar tanto la precisión del modelo como su capacidad explicativa y sensibilidad a errores.

En particular, el MAE, como función objetivo optimizada por los modelos, permite una evaluación directa y robusta de la precisión, mientras que métricas como el  $R^2$ , el MSE y el RMS complementan este análisis al proporcionar información adicional sobre la calidad del ajuste y la escala de los errores.

Dado que el rendimiento de los modelos basados en inteligencia artificial depende de la correcta selección de los hiperparámetros, es esencial explorar no solo los valores óptimos de estos parámetros, sino también sus relaciones e interacciones. En este sentido, los *corner plots* ofrecen una herramienta visual clave para interpretar estas correlaciones y su impacto en las métricas de evaluación. La siguiente sección estará dedicada a describir y ejemplificar el uso de *corner plots* en el análisis de los resultados obtenidos mediante la optimización bayesiana.

## 5.5. Visualización y Análisis con Corner Plots

Los *corner plots* son una herramienta visual esencial para analizar las interacciones entre hiperparámetros en problemas de optimización compleja (Tani and Veelken, 2024). Estos gráficos presentan la distribución unidimensional de cada hiperparámetro, así como las correlaciones bidimensionales entre pares de ellos, proporcionando una visión detallada de cómo los hiperparámetros influyen en el desempeño del modelo y se relacionan entre sí.

En el contexto de la optimización bayesiana, los *corner plots* facilitan la interpretación de los resultados al identificar patrones importantes, como:

- **Correlaciones significativas:** Detectar pares de hiperparámetros cuya interacción afecta significativamente el desempeño del modelo.
- **Regiones óptimas:** Visualizar los rangos específicos en los que los hiperparámetros producen configuraciones con un desempeño superior.
- **Sensibilidad del modelo:** Evaluar qué hiperparámetros tienen mayor impacto en la función objetivo, ayudando a priorizar su ajuste.

Estos gráficos se generan utilizando los resultados obtenidos durante el proceso de optimización bayesiana. Cada punto en el espacio visualizado representa una configuración de hiperparámetros evaluada, mientras que las densidades y correlaciones reflejan la información recopilada por el modelo probabilístico.

El análisis visual proporcionado por los *corner plots* no solo complementa

el enfoque numérico, sino que también permite validar la robustez del modelo frente a cambios en los hiperparámetros. Esta validación es crucial para garantizar que las configuraciones seleccionadas sean aplicables a problemas reales y no estén sobreajustadas a un conjunto específico de datos.

Con base en los patrones observados en los *corner plots*, se identificarán las configuraciones óptimas para cada modelo, las cuales serán aplicadas en el siguiente capítulo. Este análisis asegura que los modelos operen con parámetros ajustados para maximizar su precisión y eficiencia en el problema de ajuste de frentes de onda.

En este capítulo, se describieron herramientas avanzadas para la selección de hiperparámetros, desde la definición del espacio de búsqueda hasta la optimización bayesiana y el análisis mediante *corner plots*. Estas técnicas establecen un marco sólido para ajustar los modelos y comprender las interacciones clave entre los hiperparámetros. Dado que ya se ha establecido una configuración óptima mediante estos procedimientos, es posible ahora aplicar dicho conocimiento al procesamiento de datos y al ajuste efectivo del frente de onda, tal como se desarrolla en el capítulo siguiente.

## Capítulo 6

# Generación y Procesamiento de Datos

En este capítulo se detallan los procedimientos para la generación y procesamiento de datos necesarios en el entrenamiento y evaluación de los modelos propuestos. Primero, se describe el procesamiento de imágenes para extraer muestras reales del sistema óptico a partir de mediciones. Luego, se aborda el problema de los puntos faltantes en los datos mediante técnicas de interpolación y ajuste. Finalmente, se explica cómo, mediante la introducción de ruido controlado, se generan muestras sintéticas que simulan condiciones realistas, permitiendo evaluar la robustez de los modelos frente a perturbaciones. Estas etapas establecen la base práctica para la implementación de los métodos presentados en capítulos anteriores.

Todos los scripts y herramientas utilizados para la generación, procesamiento y simulación de datos, así como para la construcción de los modelos y la evaluación de su desempeño, se encuentran disponibles en el repositorio Zenodo asociado a este trabajo. El código puede consultarse en ([Jara, 2025](#)). A partir de este punto, cualquier mención a rutinas de procesamiento, interpolación, entrenamiento o evaluación corresponde directamente a implementaciones contenidas en dicho repositorio. Cada archivo ha sido nombrado e identificado según la sección correspondiente de esta tesis, con el fin de facilitar su consulta y reproducibilidad.

El procesamiento adecuado de los datos es un paso crítico en problemas computacionales complejos como el ajuste de frentes de onda, donde las mediciones ópticas deben transformarse en información cuantitativa y estructurada. En particular, los datos provenientes de sensores como el Shack-Hartmann

requieren técnicas específicas para extraer las posiciones de los spots y traducirlas en desplazamientos que describan las aberraciones presentes en el sistema óptico. Este paso inicial sienta las bases para la construcción de las matrices necesarias, permitiendo un análisis y ajuste efectivos en las etapas posteriores.

## 6.1. Procesamiento de Imágenes

El procesamiento de imágenes del sensor Shack-Hartmann es esencial para extraer información precisa sobre los puntos desplazados y generar las matrices  $Z$  y  $W$ , las cuales se utilizan para el ajuste de los coeficientes  $A$ . Este proceso incluye la limpieza inicial de las imágenes, la segmentación de los spots mediante umbralización de intensidades, y el cálculo de los centroides de los puntos desplazados.

Para entrenar y evaluar los modelos de ajuste de frente de onda, se generaron muestras sintéticas a partir de mediciones reales obtenidas con un sensor Shack-Hartmann bajo diversas condiciones controladas. Las mediciones fueron adquiridas utilizando un sensor Shack-Hartmann WFS150-7AR acoplado a una cámara monocromática. Este sensor cuenta con una matriz de microlentes dispuesta en una cuadrícula de  $39 \times 31$  (1209 microlentes), donde cada microlente tiene un diámetro de 0.15 mm y una distancia focal de 5.2 mm.

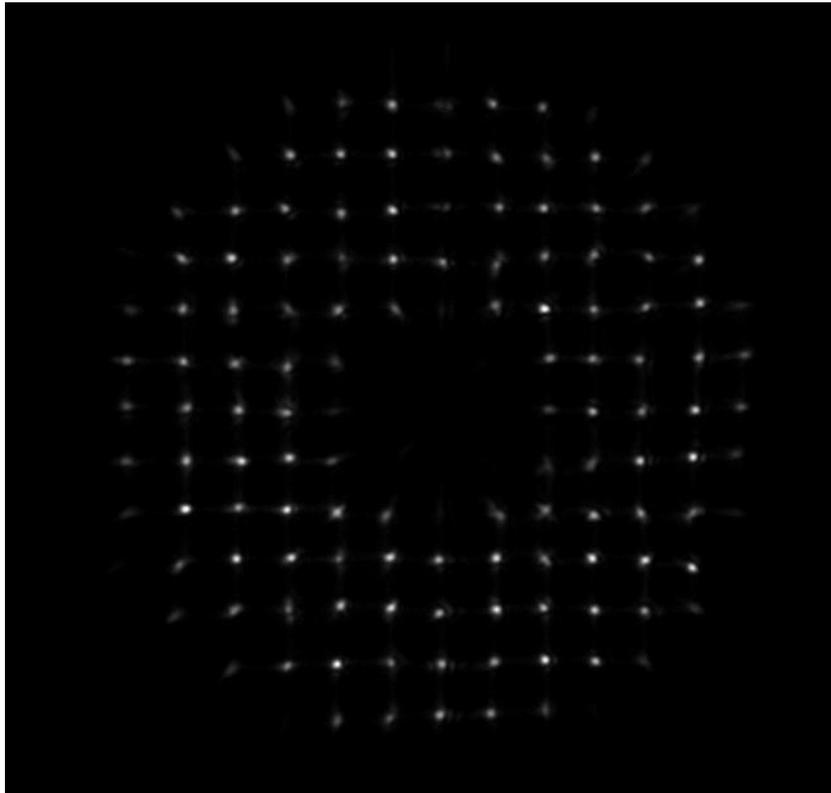
El sistema óptico utilizado para la adquisición de datos reales consistió en un telescopio de 1.5 m de apertura del Observatorio Astronómico Nacional de San Pedro Mártir (OAN-SPM), con una relación focal de  $f/13.5$ , una distancia focal de 19500 mm y un colimador con una distancia focal de 30 mm. Estas condiciones también se emplearon en la generación de datos simulados para garantizar que todas las simulaciones heredaran las mismas características ópticas del sistema real. A partir de ahora, nos referiremos como muestra base a la muestra real obtenida del sensor.

El primer paso del algoritmo es la limpieza de las imágenes, donde se elimina ruido de fondo utilizando un filtro mediano. Esto asegura que los artefactos presentes en las capturas no afecten la detección precisa de los puntos desplazados generados por la matriz de microlentes. Posteriormente, las intensidades de las imágenes se normalizan, facilitando la segmentación.

**Umbralización de intensidades.** Una vez limpiada la imagen, se aplica un umbral de intensidad que separa los píxeles relevantes (píxeles calientes) del fondo. Este proceso identifica las regiones de interés asociadas a los puntos

desplazados. A partir de estas regiones, se calculan sus centroides utilizando coordenadas cartesianas normalizadas.

La imagen en la Figura 6.1 presenta un caso real capturado por el sensor Shack-Hartmann después del preprocesamiento inicial. En ella, se distinguen las regiones iluminadas que reflejan las intensidades captadas en el detector, las cuales serán procesadas en las siguientes etapas.



**Figura 6.1:** Imagen real tomada del sensor Shack-Hartmann. Las regiones iluminadas representan las intensidades detectadas por el sensor, utilizadas para calcular las desviaciones respecto a las posiciones ideales.

Los desplazamientos detectados en la imagen se transforman a coordenadas cartesianas normalizadas dentro de un círculo unitario, donde los valores de  $x$  y  $y$  están restringidos al intervalo  $[-1, 1]$ . Estas coordenadas están definidas de manera que:

$$\sqrt{x^2 + y^2} \leq 1, \quad (6.1)$$

representando una normalización de las posiciones medidas en el plano del detector. Este sistema de referencia permite comparar directamente los desplazamientos con las posiciones ideales de los spots en condiciones óptimas.

A partir de las regiones de interés definidas, se calcula el centroide de cada punto desplazado mediante las siguientes ecuaciones:

$$P_x = \frac{\sum_i x_i I(x_i, y_i)}{\sum_i I(x_i, y_i)}, \quad (6.2)$$

$$P_y = \frac{\sum_i y_i I(x_i, y_i)}{\sum_i I(x_i, y_i)}, \quad (6.3)$$

donde  $P_x$  y  $P_y$  son las coordenadas del centroide del punto desplazado,  $x_i$  y  $y_i$  son las coordenadas de los píxeles que componen el punto, e  $I(x_i, y_i)$  es la intensidad de cada píxel. Estas posiciones medidas ( $P_x, P_y$ ) son comparadas con las posiciones ideales de los spots ( $L_x, L_y$ ), y las desviaciones se calculan como:

$$\Delta x = L_x - P_x, \quad (6.4)$$

$$\Delta y = L_y - P_y. \quad (6.5)$$

Estas desviaciones reflejan las pendientes locales del frente de onda. Una vez calculadas, se organizan en matrices estructuradas que sirven como entradas para las etapas posteriores del análisis.

El uso de coordenadas normalizadas y técnicas de segmentación asegura que los datos procesados reflejen de manera precisa las aberraciones presentes en el sistema óptico. Este procesamiento automatizado es crucial para la construcción de las matrices necesarias y el posterior ajuste de los coeficientes  $A$ .

## 6.2. Interpolación y Ajuste de Puntos Faltantes

En escenarios experimentales, es común que algunas imágenes capturadas por el sensor Shack-Hartmann presenten puntos faltantes en la matriz de spots, debido a desalineaciones mecánicas, turbulencias atmosféricas o imperfecciones ópticas. Estas condiciones pueden provocar que ciertos spots no

sean detectados correctamente o estén completamente ausentes. Dado que los métodos de inteligencia artificial utilizados en este trabajo requieren matrices de entrada completas y consistentes, es necesario aplicar un proceso de interpolación que permita reconstruir o ajustar las posiciones faltantes antes de que los datos puedan ser procesados por el modelo.

Para abordar este problema y garantizar la coherencia entre los datos de entrada y el conjunto de entrenamiento, se implementa un método de ajuste basado en el desplazamiento promedio vectorial de los puntos más cercanos.

El método comienza identificando, para cada punto faltante en la matriz de medición, los cuatro puntos más cercanos en una matriz de referencia que contiene las posiciones ideales (spots). Estos puntos cercanos se seleccionan utilizando la métrica de distancia euclidiana, definida como:

$$d(p_i, s_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad (6.6)$$

donde  $p_i = (x_i, y_i)$  representa la posición del punto desplazado, y  $s_j = (x_j, y_j)$  corresponde a las posiciones ideales de los spots.

Una vez identificados los cuatro spots más cercanos, se calcula el desplazamiento promedio como el promedio vectorial de las diferencias entre las posiciones desplazadas y las correspondientes posiciones ideales en la matriz de referencia:

$$\vec{d}_{\text{promedio}} = \frac{1}{4} \sum_{j=1}^4 (\vec{p}_j - \vec{s}_j), \quad (6.7)$$

donde  $\vec{p}_j$  y  $\vec{s}_j$  representan las posiciones correspondientes en las matrices de medición y referencia, respectivamente. Este vector de desplazamiento promedio captura cómo el punto desplazado debería ajustarse para alinearse con la configuración ideal.

El ajuste del punto faltante se realiza corrigiendo su posición inicial con el desplazamiento promedio calculado, según la ecuación:

$$p_{\text{ajustado}} = p_i - \vec{d}_{\text{promedio}}. \quad (6.8)$$

Este proceso asegura que los puntos ajustados mantengan la coherencia con la distribución utilizada durante el entrenamiento del modelo, lo que permite operar con un número consistente de puntos en todas las mediciones.

**Ventajas del Método de Ajuste** El método de ajuste basado en desplazamientos promedio presenta varias ventajas clave:

- **Consistencia en el número de puntos:** Asegura que la cantidad de puntos coincida con el conjunto de entrenamiento, independientemente de las variaciones en las mediciones experimentales.
- **Adaptabilidad a nuevas mediciones:** Permite procesar mediciones con diferentes cantidades o distribuciones de puntos sin necesidad de reentrenar el modelo.
- **Preservación de la estructura espacial:** Utilizando los cuatro puntos más cercanos, el método conserva la geometría original del frente de onda y asegura una representación precisa de los datos.

Este método es particularmente útil en áreas como los bordes de la matriz de microlentes, donde es más probable que ocurran variaciones en el número de puntos detectados debido a imperfecciones ópticas o perturbaciones externas. Al ajustar los puntos faltantes con este enfoque, se garantiza una correspondencia precisa entre los datos medidos y los utilizados en el entrenamiento del modelo, fortaleciendo la robustez y la precisión de los métodos propuestos.

### 6.2.1. Ejemplo de Interpolación y Ajuste de Puntos Faltantes

Para validar la efectividad del método de interpolación y ajuste, se realizó un experimento en el que se eliminaron cinco puntos de forma aleatoria en zonas de interés dentro de la matriz de medición. Posteriormente, se aplicó el método de ajuste basado en desplazamientos promedio para restaurar las posiciones de los puntos faltantes.

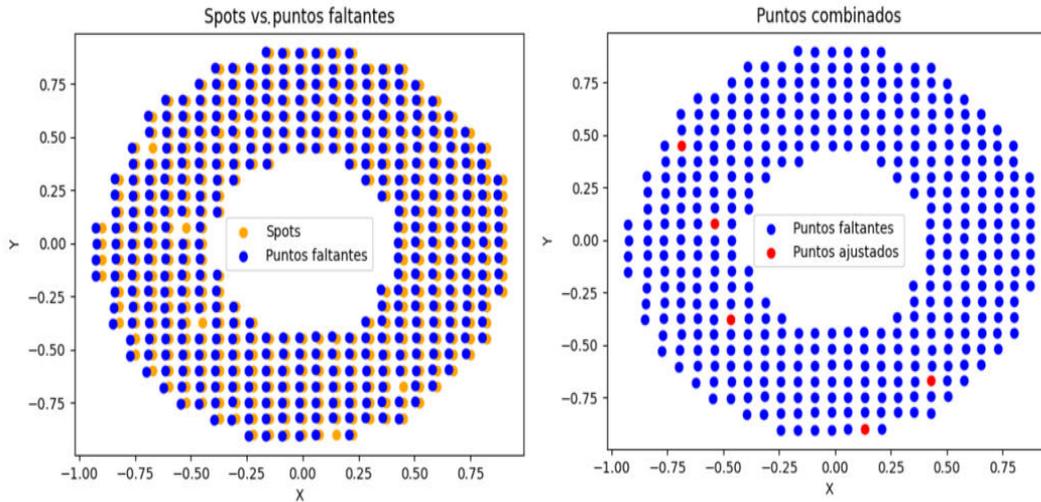
En la Tabla 6.1 se presentan las posiciones originales  $(x, y)$ , las posiciones interpoladas  $(x_{\text{interp}}, y_{\text{interp}})$  y las diferencias  $(\Delta x, \Delta y)$  obtenidas después de aplicar el método. Los resultados muestran que las diferencias son pequeñas, indicando un ajuste preciso.

**Tabla 6.1:** Resultados de interpolación para puntos eliminados aleatoriamente.

$x$	$y$	$x_{\text{interp}}$	$y_{\text{interp}}$	$\Delta x$	$\Delta y$
0.1327	-0.9005	0.1322	-0.8967	-0.0005	0.0038
0.4312	-0.6721	0.4300	-0.6703	-0.0012	0.0018
-0.4660	-0.3746	-0.4655	-0.3750	0.0005	-0.0004
-0.5391	0.0775	-0.5392	0.0772	-0.0001	-0.0003
-0.6848	0.4502	-0.6881	0.4519	-0.0032	0.0016

Adicionalmente, el valor de *Root Mean Square* (RMS) calculado para las diferencias fue de **0.00257**, lo que confirma la efectividad del método y su capacidad para realizar ajustes precisos con desviaciones mínimas.

En la Figura 6.2, se muestran dos gráficos que ilustran el caso de prueba. A la izquierda, se visualizan los spots ideales y los puntos faltantes identificados en el plano del detector. A la derecha, se presentan los puntos combinados después de aplicar el ajuste, destacando en rojo los puntos ajustados. Estos resultados muestran cómo el método logra restaurar las posiciones de los puntos faltantes respetando la estructura original.



**Figura 6.2:** Visualización del proceso de interpolación. Los puntos azules representan los puntos desplazados que podrían considerarse faltantes en el sistema, mientras que los puntos rojos muestran los resultados después del ajuste por interpolación vectorial.

Los resultados obtenidos destacan la robustez del método de interpolación

y ajuste, especialmente en escenarios donde las mediciones experimentales contienen datos incompletos. Este enfoque asegura que las matrices generadas sean consistentes con los requisitos del modelo de entrenamiento y permite mantener la precisión en las etapas posteriores de análisis.

### 6.2.2. Evaluación del Método con Diferentes Porcentajes de Puntos Faltantes

El método de ajuste basado en el desplazamiento promedio vectorial de los cuatro puntos más cercanos muestra un rendimiento sólido en escenarios con un bajo porcentaje de puntos faltantes. En sistemas con una distribución uniforme de aproximadamente 400 puntos, se observó que el método mantiene resultados precisos hasta un 15% de puntos faltantes. Sin embargo, a medida que este porcentaje aumenta, el error promedio en el ajuste crece significativamente debido a la reducción en la disponibilidad de puntos vecinos cercanos para calcular desplazamientos precisos.

En la Tabla 6.2, se presenta un análisis del rendimiento del método en función del porcentaje de puntos faltantes. El error RMS promedio y el error máximo son métricas utilizadas para evaluar la precisión del ajuste. Como se observa, el método opera de manera confiable hasta aproximadamente un 15% de puntos faltantes, con un error RMS promedio de 0.0052. Más allá de este porcentaje, las inconsistencias comienzan a manifestarse, alcanzando un error máximo de 0.0500 cuando el 30% de los puntos están ausentes.

**Tabla 6.2:** Rendimiento del método en función del porcentaje de puntos faltantes

<b>% de puntos faltantes</b>	<b>Error RMS (promedio)</b>	<b>Error máximo (desplazamiento)</b>	<b>Condición del sistema</b>
5%	0.0025	0.0050	Estable
10%	0.0038	0.0080	Estable
15%	0.0052	0.0120	Límite aceptable
20%	0.0075	0.0200	Inconsistencias observables
25%	0.0100	0.0300	Degradación significativa
30%	0.0150	0.0500	Fallo en algunos puntos

Este análisis destaca la efectividad del método bajo condiciones experimentales con un bajo número de puntos faltantes y demuestra sus limitaciones

en escenarios más desafiantes. De este modo, la interpolación y el ajuste de puntos faltantes no solo mejoran la robustez del sistema ante fallos o condiciones adversas, sino que también garantizan que los datos cumplan con los requisitos estructurales esperados por los modelos de aprendizaje profundo utilizados posteriormente.

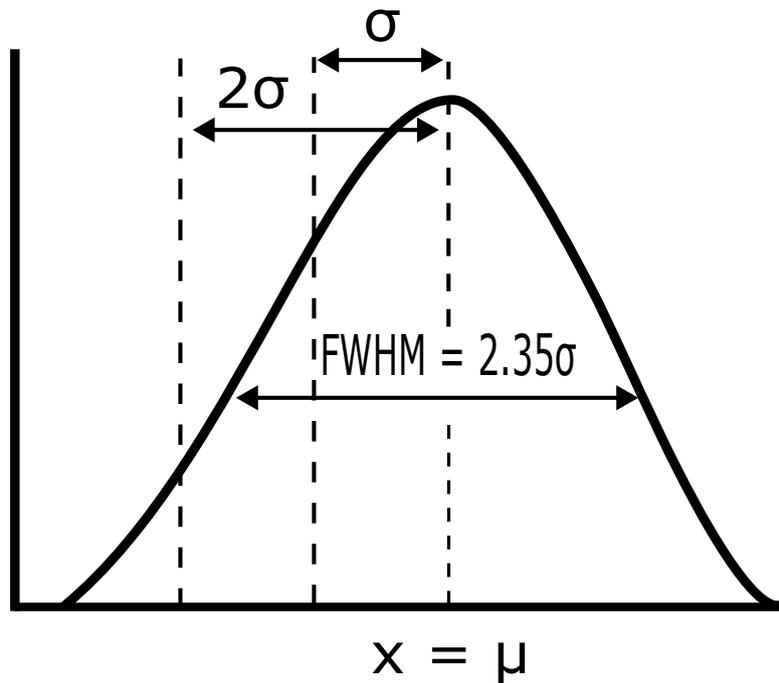
### 6.3. Introducción de Ruido en las Medidas

Para generar las muestras de entrenamiento utilizadas en este estudio, se simula la influencia de la turbulencia atmosférica sobre las mediciones del sensor Shack-Hartmann mediante la adición de ruido controlado en las posiciones de los spots. Este procedimiento permite reproducir variaciones espaciales asociadas al *seeing* astronómico, un fenómeno que describe el desenfoco causado por la atmósfera terrestre y que provoca desplazamientos de los spots en el plano del detector. Como los modelos de reconstrucción de frente de onda requieren conjuntos de datos suficientemente diversos y representativos de condiciones reales, la introducción de esta perturbación es esencial tanto para la generación de datos sintéticos como para entrenar modelos robustos ante variaciones inducidas por la atmósfera.

El *seeing* se mide en segundos de arco (arcsec) y está relacionado directamente con la función de dispersión de puntos (PSF). La anchura total a media altura (FWHM, por sus siglas en inglés) se utiliza para caracterizar el ancho de una distribución gaussiana a la mitad de su máxima intensidad y se calcula mediante:

$$FWHM \approx 2.355\sigma, \quad (6.9)$$

donde  $\sigma$  es la desviación estándar de la distribución. Un valor de *seeing* más bajo indica una mayor concentración de luz en un área pequeña, generando una imagen más nítida. Por el contrario, un *seeing* elevado implica una mayor dispersión de luz, resultando en imágenes más borrosas. La Figura 6.3 ilustra esta relación entre el *seeing* y la distribución gaussiana.



**Figura 6.3:** Relación entre el FWHM y la desviación estándar  $\sigma$  en una distribución gaussiana. El FWHM caracteriza el ancho de la distribución a media altura, siendo un parámetro clave en la descripción del seeing atmosférico.

El procedimiento de simulación incluye los pasos descritos a continuación:

1. **Generación aleatoria de seeing:** Se genera un valor de seeing aleatorio dentro del rango de 0.8 a 2 arcsec, cubriendo condiciones desde atmósferas favorables hasta desfavorables.
2. **Conversión del seeing angular a coordenadas lineales:** El valor de seeing angular se convierte a términos lineales (en mm) utilizando la siguiente ecuación:

$$\text{Seeing Lineal} = \text{Seeing Angular} \times \left( \frac{\text{Distancia Focal del Telescopio}}{206265} \right). \quad (6.10)$$

3. **Cálculo de la amplificación de la pupila:** Se calcula la amplificación de la pupila como el cociente entre la distancia focal del colimador y la distancia focal del telescopio:

$$\text{Amplificación de la Pupila} = \frac{\text{Distancia Focal del Colimador}}{\text{Distancia Focal del Telescopio}}. \quad (6.11)$$

4. **Generación de ruido gaussiano:** Se generan valores aleatorios mediante una distribución normal (ruido gaussiano) con una desviación estándar  $\sigma$  derivada del valor de seeing lineal y el factor de amplificación de la pupila. Este ruido representa desplazamientos aleatorios en las coordenadas  $x$  y  $y$  de los spots.
5. **Aplicación de ruido a los spots:** El ruido generado se suma a las coordenadas originales de los spots, simulando el efecto de las aberraciones atmosféricas en el sistema.
6. **Cálculo de métricas de calidad:** Para cada simulación, se calcula la raíz cuadrática media (RMS) de los desplazamientos y el valor de seeing correspondiente, lo que cuantifica la magnitud de las aberraciones en cada caso:

$$RMS = \sqrt{\frac{\sum(\Delta x^2 + \Delta y^2)}{N}}, \quad (6.12)$$

donde  $\Delta x$  y  $\Delta y$  son los desplazamientos en las coordenadas  $x$  y  $y$ , respectivamente, y  $N$  es el número total de puntos.

Este enfoque permite reproducir condiciones atmosféricas realistas y evaluar cómo las aberraciones inducidas por el ruido afectan las mediciones. Además, las métricas de calidad calculadas proporcionan una referencia cuantitativa para analizar la robustez y precisión de los métodos de ajuste implementados. A partir de este proceso, se generaron 3500 muestras sintéticas.

### 6.3.1. Rango de Respuesta Lineal del Sensor

Para confirmar que las muestras sintéticas generadas se encuentran dentro del rango de respuesta lineal (Endo et al., 2018), analizamos el desplazamiento máximo de los spots a través de las 3500 muestras generadas y los comparamos con las limitaciones físicas del sistema. Como se describe en la Sección 6.1, el sensor Shack-Hartmann contiene una matriz de microlentes con una cuadrícula de  $39 \times 31$ . El desplazamiento máximo permitido para que un spot permanezca en el rango lineal del sensor es de 0.075 mm (la mitad del diámetro de la microlente).

Para cada spot perturbado, calculamos la distancia euclidiana entre su posición y el centroide ideal más cercano (la ubicación esperada en ausencia de ruido) y registramos el desplazamiento más grande en cada simulación. Los resultados estadísticos obtenidos se resumen en la Tabla 6.3.

El desplazamiento más grande registrado en cualquier muestra fue  $3.1997 \times 10^{-2}$  mm, lo cual se mantiene muy por debajo del umbral de 0.075 mm, confir-

**Tabla 6.3:** Resumen de las estadísticas del desplazamiento máximo de los spots

Métrica	Valor
Total de muestras	3500
Desplazamiento máximo (mm)	$3.1997 \times 10^{-2}$
Desplazamiento mínimo (mm)	$3.1193 \times 10^{-2}$
Distancia media (mm)	$3.1634 \times 10^{-2}$
Desviación estándar (mm)	$9.0147 \times 10^{-5}$
Muestras que excedieron el límite	0

mando que todas las aberraciones simuladas permanecieron dentro del rango de respuesta lineal del sensor SHWFS. Al controlar cuidadosamente los parámetros de ruido y asegurarnos de que ningún desplazamiento simulado excediera el subaperturado de la microlente, verificamos que las reconstrucciones del frente de onda no se vieron afectadas por no linealidades, preservando la validez de la aproximación lineal requerida para un ajuste preciso del frente de onda.

### 6.3.2. Análisis Estadístico de las Aberraciones

Una vez generadas las muestras y verificado que representan de manera realista las condiciones experimentales, transformamos los datos desde coordenadas cartesianas, donde las desviaciones están expresadas como desplazamientos, hacia su representación matricial. Como se detalla en la Sección 4.4, esto permite la construcción de las matrices  $Z$  (Ecuación (4.4)) y  $W$  (Ecuación (4.5)).

Para garantizar una evaluación robusta de los métodos de reconstrucción, es fundamental realizar una caracterización estadística detallada de las aberraciones del frente de onda utilizadas en este estudio. En particular, es necesario cuantificar tanto la magnitud general de las aberraciones como su descomposición en componentes  $x$  y  $y$ . Esta caracterización es crucial, ya que las aberraciones definen el espacio de entrada para los modelos de reconstrucción y determinan su efectividad bajo diferentes escenarios de distorsión del frente de onda.

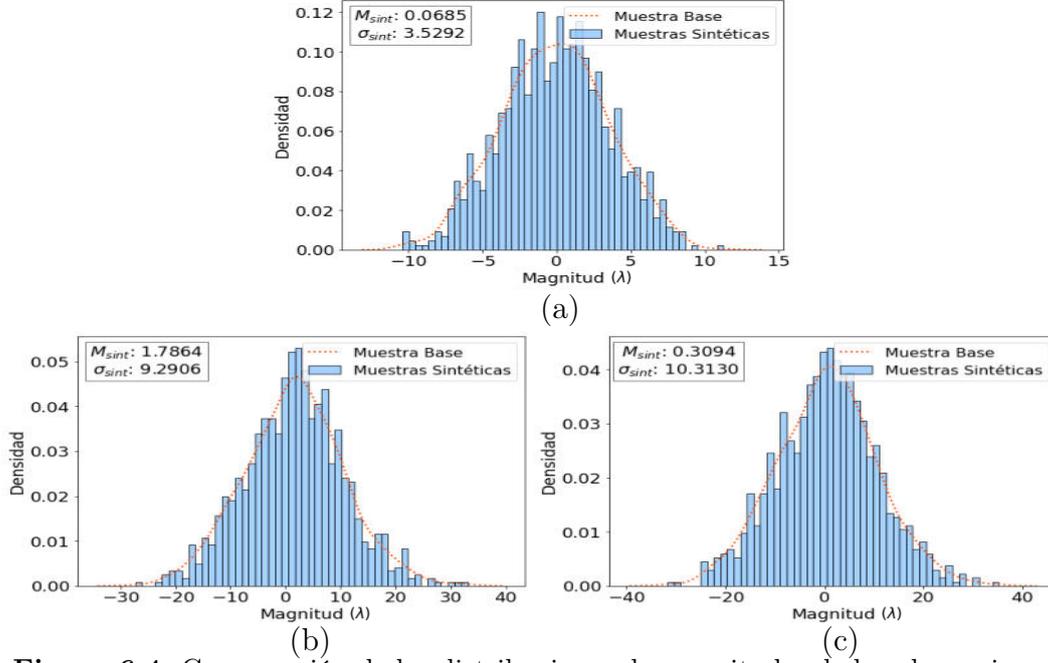
La Tabla 6.4 presenta un resumen de las magnitudes de las aberraciones en el conjunto de datos simulado, incluyendo la media, desviación estándar y el rango de valores a lo largo de las 3500 muestras sintéticas. Dado que las muestras sintéticas fueron generadas a partir de la muestra base y la metodología

de ruido gaussiano explicada en la Sección 6.3, estas siguen una distribución similar a la de la muestra base. Sin embargo, presentan una pequeña variación debido al ruido añadido, lo que permite evaluar cómo las perturbaciones controladas afectan el desempeño de los modelos de reconstrucción.

**Tabla 6.4:** Resumen de las estadísticas de la magnitud de las aberraciones para las muestras base y sintéticas

Métrica	Muestra Base ( $\lambda$ )	Muestras Sintéticas ( $\lambda$ )
Mediana	$5.9683 \times 10^{-1}$	$5.9650 \times 10^{-1}$
$\sigma$	7.6221	7.6219

Además, para evaluar la consistencia entre los datos reales y sintéticos, la Figura 6.4 compara las distribuciones de probabilidad de tres aberraciones seleccionadas de ambos conjuntos de datos. Nos centramos en tres casos representativos que abarcan órdenes bajos, medios y altos:  $Z_4$  (Astigmatismo Y de bajo orden),  $Z_{14}$  (Tetrafoil Y de orden medio) y  $Z_{24}$  (Tetrafoil Secundario X de alto orden). Los histogramas muestran que las muestras sintéticas replican efectivamente las características estadísticas del conjunto de datos real, preservando la distribución de las distorsiones del frente de onda mientras introducen variaciones controladas. Este ajuste cercano indica que el conjunto de datos generado es adecuado para entrenar y evaluar los modelos de reconstrucción sin introducir sesgos artificiales.



**Figura 6.4:** Comparación de las distribuciones de magnitudes de las aberraciones entre la muestra real (Base) y las muestras sintéticas para aberraciones seleccionadas: (a)  $Z_4$  Astigmatismo  $Y$ , (b)  $Z_{14}$  Tetrafoil  $Y$ , y (c)  $Z_{24}$  Tetrafoil Secundario  $X$ . Cada panel muestra el histograma de las muestras sintéticas (barras azules) y la línea punteada correspondiente a la muestra base, lo que ilustra la estrecha coincidencia estadística entre ambos conjuntos de datos. El eje  $x$  representa la magnitud de cada coeficiente de Zernike en unidades de la longitud de onda ( $\lambda$ ), mientras que el eje  $y$  indica la frecuencia de ocurrencia. La mediana ( $M_{sint}$ ) y la desviación estándar ( $\sigma_{sint}$ ) de las muestras sintéticas se muestran dentro de cada panel.

Para un análisis más detallado, el Apéndice A.2 presenta tablas estadísticas que resumen los datos de las aberraciones: (i) las magnitudes generales de las aberraciones en los conjuntos de datos base y sintéticos (ver Tabla A.1), (ii) los componentes  $x$  y  $y$  de las aberraciones en el conjunto de datos base (ver Tabla A.2), y (iii) los componentes  $x$  y  $y$  de las aberraciones en el conjunto de datos sintético (ver Tabla A.3). Estas tablas describen de manera integral la distribución estadística del espacio de entrada para los modelos de reconstrucción, reforzando la validez del conjunto de datos para la evaluación del rendimiento realizada en este estudio.

Una vez establecido el proceso para generar muestras sintéticas que replican condiciones experimentales realistas y verificada la consistencia de los datos para el ajuste de frentes de onda, se cuenta ya con los insumos necesarios

para entrenar y configurar los modelos. En el capítulo siguiente se exploran precisamente las estrategias de entrenamiento y los criterios empleados para evaluar su desempeño.

# Capítulo 7

## Entrenamiento y Configuración de Modelos

Dado que la implementación de los modelos se apoya directamente en los conceptos teóricos desarrollados en los capítulos anteriores, este capítulo presenta una descripción concreta y orientada a la práctica de las configuraciones utilizadas para el ajuste de frentes de onda. Se comienza con la implementación del método de mínimos cuadrados como referencia inicial, seguido por el entrenamiento del modelo de gradiente descendente optimizado con Adam, donde se destacan el ajuste de hiperparámetros y el proceso de entrenamiento. Finalmente, se describe la configuración y entrenamiento de la red hiperdinámica, integrando los hiperparámetros óptimos obtenidos mediante optimización bayesiana. Este conjunto de metodologías sienta las bases para evaluar el rendimiento de los modelos en escenarios prácticos.

El enfoque de este capítulo es plenamente práctico y se centra en mostrar cómo se entrenan y configuran los modelos para lograr un ajuste preciso del frente de onda. Se detallan las particularidades de cada implementación y se aplica la optimización bayesiana, previamente explicada de forma teórica (véase Sección 5.3), como herramienta clave para seleccionar los hiperparámetros óptimos en cada caso.

Dada la complejidad inherente de los datos utilizados y el costo computacional de los procesos de entrenamiento, se requirió un entorno computacional de alto rendimiento. Este entorno fue esencial para manejar las matrices de alta dimensionalidad generadas en las etapas previas (véase Capítulo 6) y realizar las iteraciones intensivas propias de los modelos de aprendizaje automático.

El entrenamiento y la evaluación de los modelos se realizaron en un servidor

del Laboratorio de Ciencia de Datos del Centro de Investigación y Desarrollo de Tecnología Digital (CITEDI) del Instituto Politécnico Nacional (IPN), con las siguientes especificaciones técnicas: dos procesadores Intel Xeon Gold 5218 de 32 núcleos cada uno, funcionando a 2.30 GHz, 512 GiB de memoria RAM y 8 GPUs NVIDIA TITAN RTX, cada una con 24 GiB de memoria dedicada. El servidor operó bajo el sistema operativo Ubuntu 22.04.4 LTS, con soporte para CUDA versión 12.4. Este entorno computacional permitió realizar entrenamientos complejos y optimizaciones bayesianas, así como manejar grandes volúmenes de datos, garantizando una evaluación eficiente y uniforme de los modelos propuestos.

## 7.1. Entrenamiento Mínimos Cuadrados

El método de mínimos cuadrados es ampliamente conocido por su simplicidad y eficacia al resolver sistemas lineales sobredeterminados, y su descripción teórica ya ha sido detallada en la Sección 4.5. Como se explicó, el método calcula los coeficientes  $A$  resolviendo el sistema descrito en la Ecuación (4.6), minimizando el error cuadrático medio entre los datos medidos y las predicciones del modelo.

Dado que este enfoque no requiere un proceso de entrenamiento, su implementación en esta etapa se centró en la aplicación directa del método utilizando las matrices  $Z$  y  $W$  generadas durante las simulaciones. El cálculo de  $A$  se realizó mediante la pseudoinversa de  $Z$ , permitiendo obtener los coeficientes óptimos sin necesidad de ajustar hiperparámetros ni realizar iteraciones.

A pesar de su carácter estático, la implementación del método en el entorno computacional descrito previamente permitió evaluar su desempeño de manera consistente y compararlo con los modelos basados en aprendizaje automático. Este método sirvió como referencia fundamental, estableciendo un punto de partida para medir las mejoras introducidas por técnicas más avanzadas, como el gradiente descendente con optimizador Adam y la red hiperdinámica.

## 7.2. Gradiente Descendente para Una Sola Muestra

El entrenamiento del modelo de gradiente descendente para la muestra base se centró en optimizar la velocidad de convergencia. Este escenario busca

evaluar el desempeño del modelo en condiciones controladas y con datos limitados, priorizando configuraciones que minimicen rápidamente la función de pérdida sin comprometer la precisión.

El procedimiento llevado a cabo incluyó las siguientes etapas clave:

1. **Definición del objetivo:** Reducir el Error Absoluto Medio (MAE) en la predicción de los coeficientes  $A$ , garantizando estabilidad durante el entrenamiento y minimizando la cantidad de iteraciones necesarias para alcanzar la convergencia.
2. **Selección del espacio de búsqueda:** Se definió un espacio de búsqueda que abarcara configuraciones potencialmente óptimas basándose en la teoría previa y resultados experimentales. Este espacio se limitó estratégicamente para equilibrar la exploración y la explotación, priorizando rangos donde el modelo pudiera demostrar estabilidad y una rápida convergencia.
3. **Evaluación iterativa mediante optimización bayesiana:** El modelo fue evaluado iterativamente utilizando optimización bayesiana, que seleccionó automáticamente los siguientes hiperparámetros a evaluar en cada paso. Este enfoque permitió maximizar la eficiencia en la búsqueda de configuraciones óptimas.
4. **Análisis del desempeño mediante *corner plots*:** Los resultados de la optimización bayesiana fueron analizados utilizando *corner plots*, los cuales permitieron visualizar las correlaciones entre hiperparámetros. Con base en las medianas obtenidas para cada hiperparámetro, se seleccionaron configuraciones con la mejor correlación y desempeño global.

A partir de este análisis, se determinaron los hiperparámetros óptimos para el caso de una muestra, asegurando que el modelo alcanzara una solución precisa con un consumo mínimo de recursos computacionales.

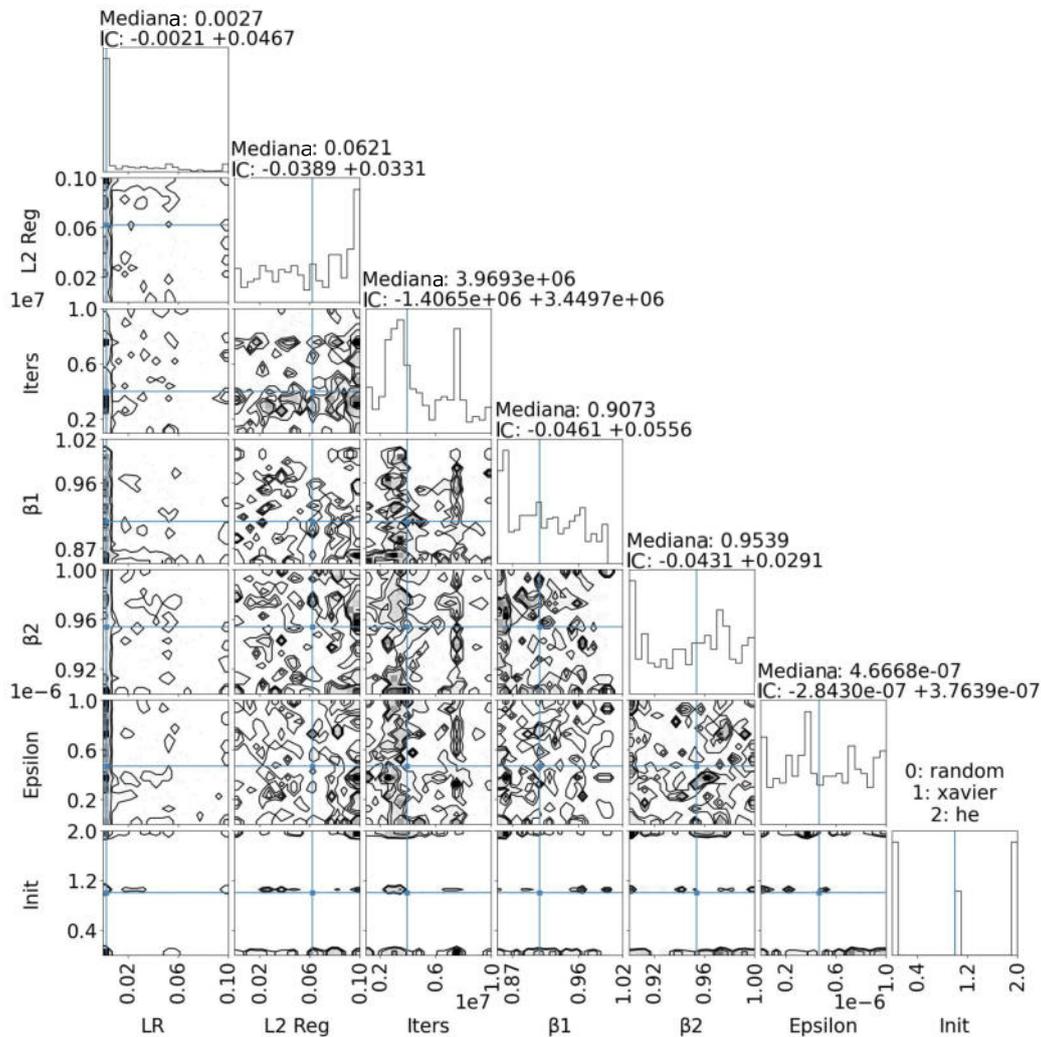
Los hiperparámetros evaluados y sus respectivos rangos se presentan en la Tabla 7.1, los cuales incluyen parámetros clave del optimizador Adam.

**Tabla 7.1:** Espacio de búsqueda de hiperparámetros para la optimización bayesiana en el modelo de gradiente descendente.

Hiperparámetro	Rango
Tasa de aprendizaje ( $\alpha$ )	$[1 \times 10^{-15}, 1 \times 10^{-1}]$
Regularización L2	$[1 \times 10^{-8}, 1 \times 10^{-1}]$
Número de iteraciones	$[1 \times 10^6, 1 \times 10^7]$
$\beta_1$ (Optimizador Adam)	$[0.85, 0.99]$
$\beta_2$ (Optimizador Adam)	$[0.90, 0.99]$
$\epsilon$ (Optimizador Adam)	$[1 \times 10^{-10}, 1 \times 10^{-6}]$
Método de inicialización	{Aleatorio, Xavier, He}

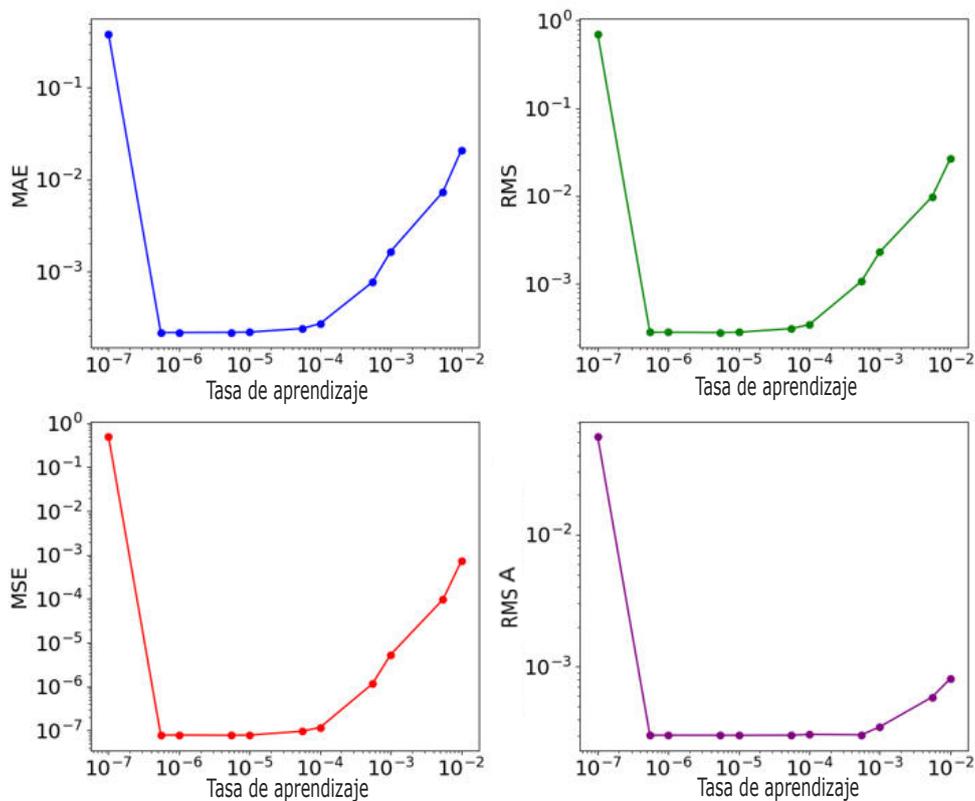
Una vez seleccionados los hiperparámetros iniciales, se procedió a realizar la optimización bayesiana, la cual tuvo una duración aproximada de 2 horas. Este proceso permitió identificar los hiperparámetros con mayor impacto en la estabilidad y convergencia del modelo. Durante las 500 evaluaciones realizadas, se generó un diagrama de esquinas (ver Figura 7.1) que visualizó las correlaciones entre los hiperparámetros y su influencia en la función objetivo, destacando las regiones más relevantes del espacio de búsqueda.

A partir del análisis del diagrama de esquinas, se seleccionaron las medianas de los hiperparámetros  $\beta_1$ , regularización L2,  $\beta_2$  y  $\epsilon$ , dado que estos valores representaban las configuraciones más estables y con menor incertidumbre. Asimismo, se optó por el método de inicialización Xavier, que demostró ser adecuado para este tipo de modelo debido a su capacidad de mantener un equilibrio en la magnitud de los gradientes durante el entrenamiento (Glorot and Bengio, 2010).



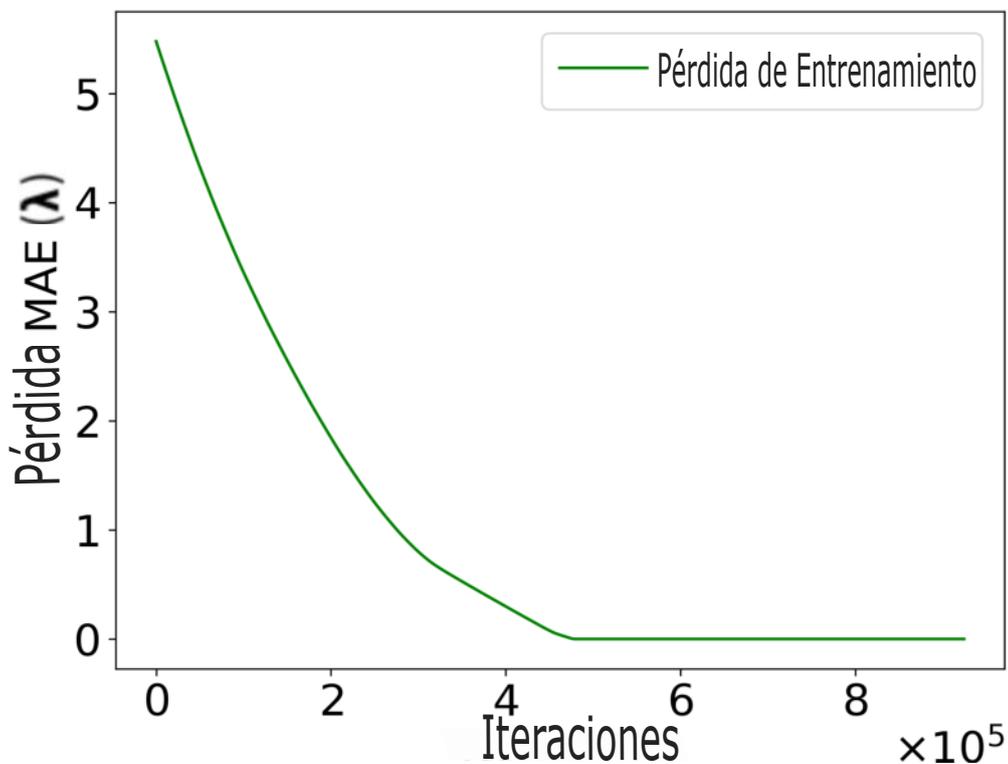
**Figura 7.1:** Resultados de la optimización bayesiana de los hiperparámetros para el modelo de gradiente descendente. La figura muestra las distribuciones marginales (en la diagonal) y las correlaciones por pares (fuera de la diagonal) de los hiperparámetros explorados durante el proceso de optimización. Los hiperparámetros y sus respectivos rangos de búsqueda se definen en la Tabla 7.1. El objetivo de la optimización fue minimizar la pérdida de entrenamiento, definida como el MAE entre los valores predichos y los valores objetivo de  $W$ , expresados en unidades de longitud de onda ( $\lambda$ ). La figura resalta las regiones del espacio de hiperparámetros con mayor densidad de evaluaciones, las cuales están asociadas con un mejor desempeño en la minimización de la función de pérdida.

Sin embargo, para la tasa de aprendizaje y el número de iteraciones, se consideró necesario un análisis más detallado. En el caso de la tasa de aprendizaje, aunque la optimización bayesiana permitió identificar que los valores óptimos se encontraban en rangos pequeños, no fue posible determinar con precisión el valor específico debido a la sensibilidad del modelo a este hiperparámetro. Por ello, se realizó una búsqueda en malla (*grid search*) para explorar valores en un rango reducido y determinar aquel que garantizara una convergencia rápida y estable. Este análisis permitió seleccionar  $1 \times 10^{-6}$  como el valor más adecuado, como se observa en la Figura 7.2.



**Figura 7.2:** Resultados de la búsqueda en malla para la tasa de aprendizaje en el modelo de gradiente descendente. Cada gráfico muestra la evolución de una métrica de error: MAE, RMS, MSE y RMS A. El eje  $x$  representa el valor de la tasa de aprendizaje explorada, mientras que el eje  $y$  muestra el valor correspondiente de la métrica evaluada. Se observa que valores muy bajos o muy altos de la tasa de aprendizaje generan errores elevados, mientras que en un intervalo intermedio las métricas se minimizan, lo que indica una mejor convergencia del modelo. Este análisis permitió seleccionar un valor óptimo de tasa de aprendizaje que garantiza un entrenamiento eficiente y estable.

Para el caso del número de iteraciones, se utilizó un mecanismo de detención temprana (*early stopping*), diseñado para finalizar el proceso de entrenamiento una vez que las mejoras en la función de pérdida (MAE) se volvían insignificantes. Este enfoque permitió ajustar dinámicamente la duración del entrenamiento, reduciendo el tiempo computacional sin comprometer la precisión. La Figura 7.3 muestra la evolución de la pérdida durante el proceso de entrenamiento, indicando una rápida convergencia seguida de estabilización. Este método garantizó que el modelo alcanzara su rendimiento óptimo sin necesidad de realizar un número fijo de iteraciones.



**Figura 7.3:** Evolución de la pérdida de entrenamiento (MAE) durante el proceso de optimización utilizando el método de gradiente descendente. El eje  $x$  indica el número de iteraciones (en unidades de  $10^5$ ), y el eje  $y$  muestra el MAE entre los valores predichos y los valores objetivo de  $W$ , expresados en unidades de longitud de onda ( $\lambda$ ). Esta curva ilustra la reducción progresiva del error de predicción a medida que los parámetros del modelo se actualizan durante el entrenamiento.

Los hiperparámetros finales seleccionados se resumen en la Tabla 7.2.

**Tabla 7.2:** Hiperparámetros seleccionados mediante optimización bayesiana y búsqueda en malla.

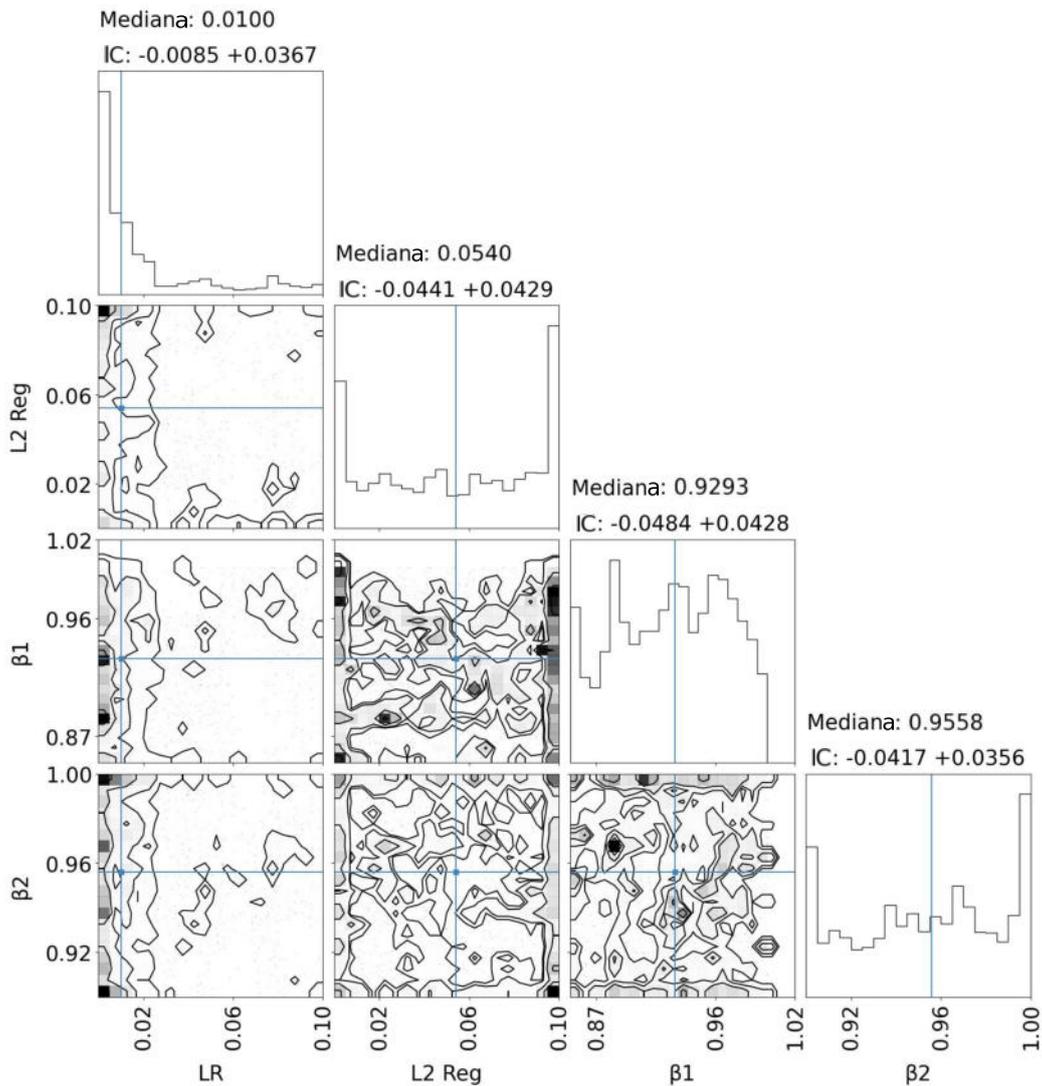
Hiperparámetro	Valor óptimo
Tasa de aprendizaje ( $\alpha$ )	$1 \times 10^{-6}$
Regularización L2	$6.2100 \times 10^{-2}$
Número de iteraciones	969478
$\beta_1$ (Optimizador Adam)	0.9073
$\beta_2$ (Optimizador Adam)	0.9539
$\epsilon$ (Optimizador Adam)	$4.6668 \times 10^{-7}$
Método de inicialización	Xavier

El proceso completo de entrenamiento concluyó después de 969,478 iteraciones, con una duración total de 1 minuto y 40 segundos. Este resultado demuestra la eficiencia alcanzada al combinar la selección óptima de hiperparámetros y técnicas avanzadas de optimización.

### 7.3. Gradiente Descendente para Múltiples Muestras

Con el objetivo de evaluar la capacidad de generalización y adaptabilidad del modelo de gradiente descendente, se llevó a cabo un entrenamiento utilizando un conjunto de 3000 muestras ( $Z$  y  $W$  matrices), extraídas del total de 3500 muestras sintéticas generadas, descritas en el Capítulo 6. A diferencia del caso de una sola muestra, este escenario se centra en validar el desempeño del modelo en un entorno más complejo, donde la estabilidad y la precisión deben mantenerse constantes en un volumen considerable de datos.

Para explorar el espacio de hiperparámetros, se utilizó nuevamente la optimización bayesiana, cuyos resultados generaron el diagrama de esquinas mostrado en la Figura 7.4. Este análisis permitió identificar diferentes zonas de densidad para los hiperparámetros evaluados, destacando aquellos valores con mayor influencia en la estabilidad y rendimiento del modelo. Este proceso de optimización fue computacionalmente intensivo, con una duración aproximada de 24 horas, debido a la complejidad del modelo y la cantidad de muestras utilizadas.



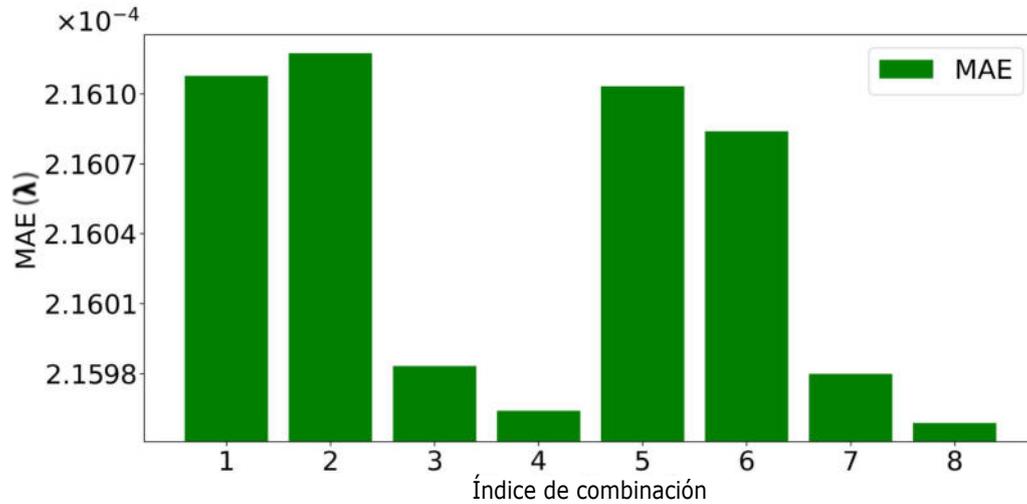
**Figura 7.4:** Resultados de la optimización bayesiana de los hiperparámetros utilizando el conjunto de datos sintético de 3000 muestras. La figura muestra las distribuciones marginales (en la diagonal) y las correlaciones por pares (fuera de la diagonal) de los hiperparámetros explorados durante el proceso de optimización del modelo de gradiente descendente. El objetivo de la optimización fue minimizar la pérdida de entrenamiento, definida como el MAE entre los valores predichos y los valores objetivo de  $W$ , expresados en unidades de longitud de onda ( $\lambda$ ). La figura resalta las regiones del espacio de hiperparámetros con mayor densidad de evaluaciones, las cuales se asocian a un mejor desempeño del modelo al minimizar la función de pérdida.

A partir de este análisis, se identificaron combinaciones críticas para los hiperparámetros, las cuales se presentan en la Tabla 7.3. Estas zonas fueron posteriormente exploradas mediante una búsqueda en malla (*grid search*) con el fin de refinar las configuraciones y seleccionar los valores óptimos para cada hiperparámetro.

**Tabla 7.3:** Zonas críticas identificadas mediante optimización bayesiana para múltiples muestras para el modelo de gradiente descendente.

Hiperparámetro	Valores Críticos
Regularización L2	0.01, 0.09
$\beta_1$ (Optimizador Adam)	0.88, 0.98
$\beta_2$ (Optimizador Adam)	0.91, 0.99
Tasa de aprendizaje ( $\alpha$ )	$1 \times 10^{-6}$

La búsqueda en malla permitió explorar las combinaciones de hiperparámetros identificadas como críticas en la optimización bayesiana. Los resultados obtenidos se presentan en la Figura 7.5, que muestra el MAE asociado a cada combinación. Asimismo, la Tabla 7.4 detalla los valores específicos evaluados para cada conjunto de hiperparámetros. De esta manera, se seleccionó la combinación que minimizaba el MAE, priorizando un equilibrio adecuado entre precisión y estabilidad del modelo.



**Figura 7.5:** Resultados del MAE para cada combinación de hiperparámetros evaluada durante la búsqueda en malla del modelo de gradiente descendente. El eje  $x$  representa el índice de cada combinación probada, mientras que el eje  $y$  muestra el valor correspondiente de MAE, expresado en unidades de longitud de onda. Cada combinación de hiperparámetros se detalla en la Tabla 7.4. Esta figura permite comparar el desempeño de las distintas configuraciones y seleccionar la combinación óptima que minimiza el error de predicción.

**Tabla 7.4:** Combinaciones de hiperparámetros evaluadas y sus respectivos valores de MAE.

Índice	$\alpha$	Regularización L2	$\beta_1$	$\beta_2$	MAE ( $\times 10^{-4}$ )
1	$1.0 \times 10^{-6}$	0.01	0.88	0.91	2.1611
2	$1.0 \times 10^{-6}$	0.01	0.88	0.99	2.1612
3	$1.0 \times 10^{-6}$	0.01	0.98	0.91	2.1598
4	$1.0 \times 10^{-6}$	0.01	0.98	0.99	2.1596
5	$1.0 \times 10^{-6}$	0.09	0.88	0.91	2.1610
6	$1.0 \times 10^{-6}$	0.09	0.88	0.99	2.1608
7	$1.0 \times 10^{-6}$	0.09	0.98	0.91	2.1598
8	$1.0 \times 10^{-6}$	0.09	0.98	0.99	2.1595

Finalmente, los hiperparámetros óptimos seleccionados para el modelo entrenado con múltiples muestras se resumen en la Tabla 7.5. Esta configura-

ción garantiza la estabilidad del modelo y su capacidad para generalizar en escenarios más complejos.

**Tabla 7.5:** Hiperparámetros óptimos seleccionados para el modelo de gradiente descendente con múltiples muestras.

Hiperparámetro	Valor Óptimo
Tasa de aprendizaje ( $\alpha$ )	$1 \times 10^{-6}$
Regularización L2	0.09
$\beta_1$ (Optimizador Adam)	0.98
$\beta_2$ (Optimizador Adam)	0.99

El análisis de desempeño con este conjunto de muestras más amplio demuestra la adaptabilidad y capacidad de generalización del modelo, consolidando su aplicabilidad para resolver problemas más complejos en un entorno realista.

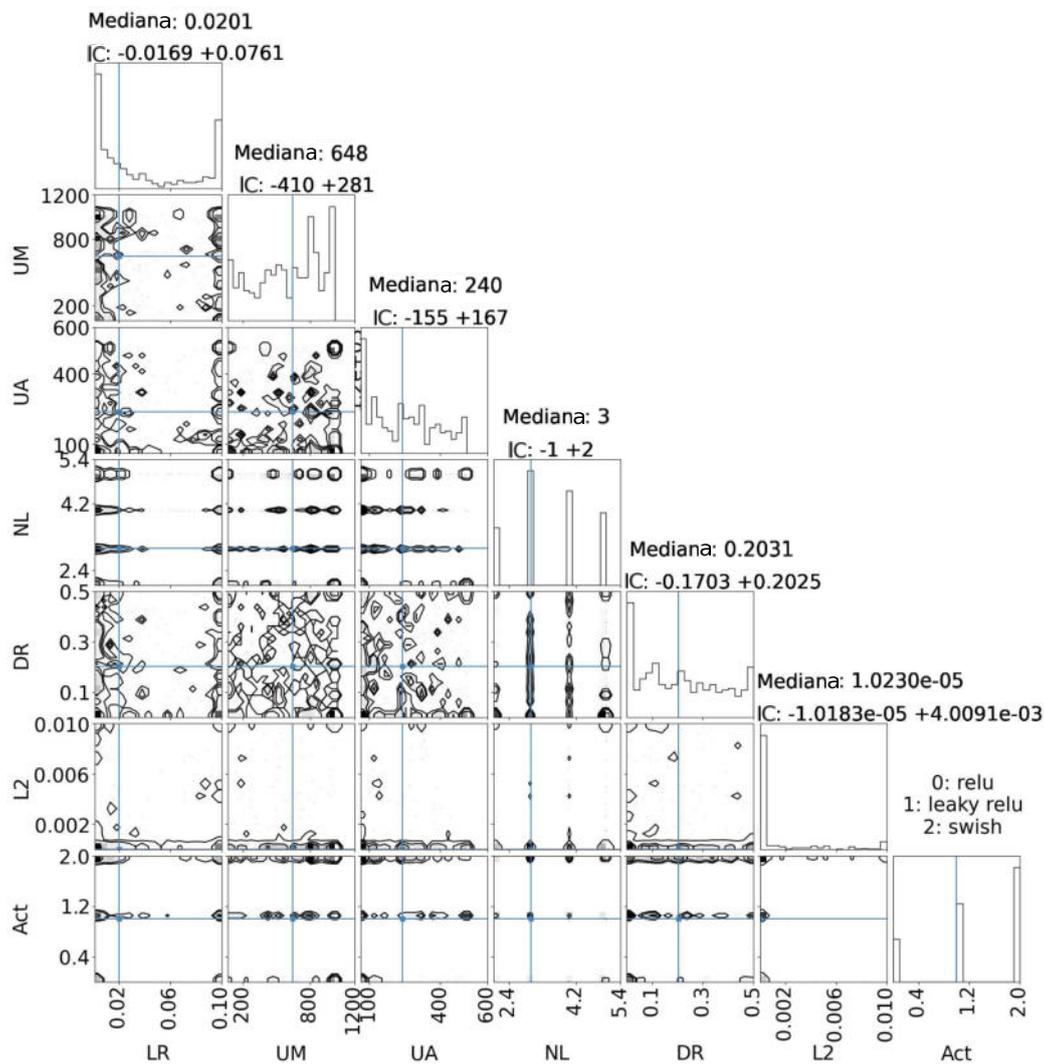
## 7.4. Entrenamiento del Modelo de Red Hiperdinámica

El entrenamiento del modelo de red hiperdinámica se diseñó con el objetivo principal de garantizar su capacidad de generalizar ante cualquier entrada, evitando caer en el riesgo de sobreajuste que podría llevar al modelo a memorizar respuestas específicas en lugar de aprender patrones generales. Para ese propósito, se seleccionaron las mismas 3000 muestras ( $Z$  y  $W$ ) utilizadas para entrenar el modelo de gradiente descendente en múltiples casos, descrito en la Sección 7.3, buscando identificar configuraciones que promovieran un aprendizaje robusto, estable y adaptable en escenarios variados y complejos. Este enfoque permite que el modelo mantenga un rendimiento consistente al enfrentarse a datos no vistos, maximizando su utilidad en aplicaciones prácticas.

Para determinar la configuración óptima de los hiperparámetros, se realizó una optimización bayesiana que exploró el espacio definido en la Tabla 7.6. Este proceso abarcó 500 evaluaciones ( $n\_calls$ ) con una duración aproximada de 49 horas. Durante este análisis, se identificaron correlaciones críticas entre los hiperparámetros, como se observa en la Figura 7.6.

**Tabla 7.6:** Espacio de búsqueda de hiperparámetros en el entrenamiento de la red hiperdinámica.

Hiperparámetro	Rango
Tasa de aprendizaje	$[1 \times 10^{-5}, 1 \times 10^{-2}]$
Regularización L2	$[1 \times 10^{-6}, 1 \times 10^{-1}]$
Tasa de dropout	$[0.1, 0.5]$
Tamaño de batch	$[16, 64]$
Número de capas adicionales	$[2, 6]$
Neuronas en red madre ( $U_m$ )	$[200, 1200]$
Neuronas en red auxiliar ( $U_a$ )	$[100, 600]$
Función de activación	{ReLU, Leaky ReLU, Tanh}
Inicialización del kernel	{Glorot Normal, He Normal}



**Figura 7.6:** Resultados de la optimización bayesiana de los hiperparámetros del modelo de red hiperdinámica. La figura muestra las distribuciones marginales (en la diagonal) y las correlaciones por pares (fuera de la diagonal) de los hiperparámetros explorados durante el proceso de optimización. Los hiperparámetros y sus respectivos rangos de búsqueda se definen en la Tabla 7.6. El objetivo de la optimización fue minimizar la pérdida de entrenamiento, definida como el MAE entre los valores predichos y los valores objetivo de  $W$ , expresados en unidades de longitud de onda ( $\lambda$ ). Este diagrama de esquinas revela interdependencias entre los hiperparámetros y ayuda a identificar la configuración óptima, la cual se resume en la Tabla 7.7.

A partir del análisis del diagrama de esquinas, se seleccionaron los valores óptimos para cada hiperparámetro, como se muestra en la Tabla 7.7. Estos valores garantizaron una configuración estable y eficiente para el entrenamiento del modelo.

**Tabla 7.7:** Hiperparámetros óptimos seleccionados para el entrenamiento de la red hiperdinámica.

Hiperparámetro	Valor Óptimo
Tasa de aprendizaje	$1 \times 10^{-5}$
Regularización L2	$1.0230 \times 10^{-2}$
Tasa de dropout	0.2031
Tamaño de batch	32
Número de capas adicionales	3
Neuronas en red madre ( $U_m$ )	648
Neuronas en red auxiliar ( $U_a$ )	240
Función de activación	Leaky ReLU
Inicialización del kernel	Glorot Normal

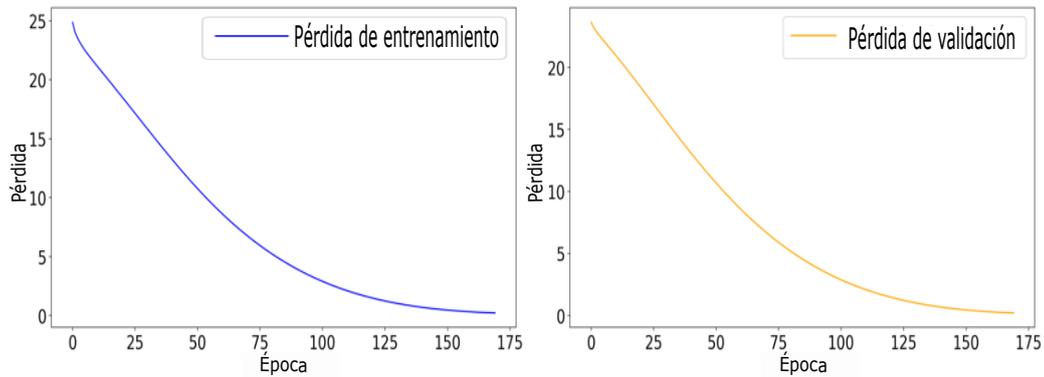
## Entrenamiento y Validación

El modelo fue entrenado durante 200 épocas utilizando un mecanismo de detención temprana para detener el proceso una vez que la pérdida de validación dejó de mejorar. La evolución de las métricas de pérdida y MAE se muestran en las Figuras 7.7 y 7.8, respectivamente.

La Figura 7.7 muestra la evolución de la pérdida tanto en el conjunto de entrenamiento como en el de validación:

- Ambas curvas presentan un descenso constante hasta estabilizarse alrededor de las 150 épocas, indicando que el modelo alcanza un punto de convergencia.
- La pérdida de validación es ligeramente superior a la de entrenamiento, un comportamiento esperado que sugiere que el modelo generaliza bien.
- El comportamiento paralelo entre ambas curvas descarta la presencia de sobreajuste, ya que no hay una separación significativa hacia el final del entrenamiento.

En la Figura 7.8, que detalla la evolución del MAE:

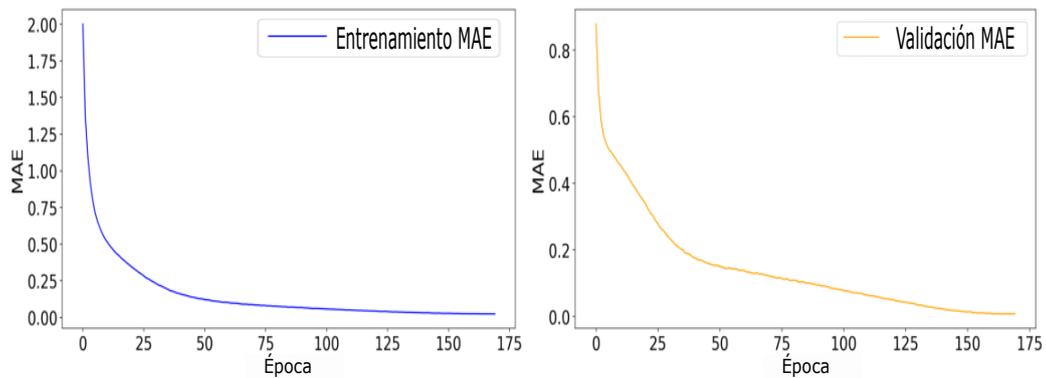


**Figura 7.7:** Evolución de la pérdida de entrenamiento y validación para el modelo de red hiperdinámica. El panel izquierdo muestra la pérdida de entrenamiento, y el panel derecho muestra la pérdida de validación, ambas medidas utilizando el MAE entre los valores predichos y los valores objetivo. El modelo fue entrenado con datos de entrada y salida normalizados, por lo que los valores de pérdida son adimensionales. Estas curvas ilustran cómo el modelo generaliza a datos no vistos y permiten evaluar el comportamiento de convergencia durante el entrenamiento.

- Las curvas de MAE también muestran un descenso significativo durante las primeras épocas, seguido de una estabilización en torno a las 150 épocas.
- Los valores de MAE en entrenamiento y validación son cercanos, lo cual refuerza la capacidad del modelo para generalizar y evitar la memorización de patrones específicos.
- La estabilidad de estas métricas hacia el final del proceso demuestra que el modelo puede adaptarse a nuevos datos sin degradar su rendimiento.

Estos resultados evidencian que la red hiperdinámica tiene una alta capacidad de generalización, manejando de manera efectiva escenarios complejos sin caer en el sobreajuste.

Con los modelos adecuadamente configurados y asegurando que están optimizados a su máxima capacidad, el siguiente capítulo se enfocará en su evaluación, analizando comparativamente su precisión, eficiencia, tiempo computacional en la predicción y adaptabilidad frente a los desafíos planteados.



**Figura 7.8:** Evolución del MAE durante el entrenamiento y la validación del modelo de red hiperdinámica. El panel izquierdo muestra el MAE en el conjunto de entrenamiento, y el panel derecho muestra el MAE en el conjunto de validación, ambos graficados a lo largo de las épocas de entrenamiento. El MAE se calcula entre los valores predichos y los valores objetivo. Dado que el modelo opera con datos de entrada y salida normalizados, el error es adimensional. Estas curvas permiten analizar el comportamiento de aprendizaje del modelo y su capacidad de generalización.

## Capítulo 8

# Evaluación del Desempeño y Discusión

En este capítulo se analizan los resultados obtenidos para los tres métodos evaluados en el ajuste de aberraciones del frente de onda. Se presentan métricas cuantitativas y visualizaciones que permiten comparar la precisión, estabilidad, capacidad de generalización y tiempo computacional de cada enfoque. Además, se evalúa su comportamiento en presencia de valores atípicos, destacando las implicaciones prácticas de los hallazgos.

En este capítulo se presentan los resultados del desempeño de los tres métodos utilizados para el ajuste de aberraciones del frente de onda: mínimos cuadrados, gradiente descendente y la red hiperdinámica. Primero, se exponen los resultados correspondientes a la muestra base, obtenida en el Capítulo 6, incluyendo el resumen de las métricas de error y el análisis en presencia de valores atípicos. Posteriormente, se realiza el análisis en el conjunto de pruebas, que consiste en 500 muestras seleccionadas del total de las 3500 muestras sintéticas generadas en el mismo capítulo. Estas corresponden a las 500 muestras restantes que no fueron utilizadas durante el proceso de entrenamiento en los casos del modelo de gradiente descendente y de la red hiperdinámica, descritos en las Secciones 7.3 y 7.4, respectivamente. Finalmente, se discuten las implicaciones prácticas de los hallazgos, con un enfoque en la interpretación profunda de los resultados.

## 8.1. Evaluación Comparativa de Desempeño

Esta sección presenta un análisis detallado del desempeño de los tres métodos evaluados para el ajuste de aberraciones del frente de onda: mínimos cuadrados, gradiente descendente y red hiperdinámica. El análisis se basa en la **muestra base**, obtenida en el Capítulo 6, lo cual permite establecer una primera comparación controlada y directa entre los enfoques propuestos. La evaluación se realizó utilizando las métricas definidas previamente en la Sección 5.4, incluyendo el coeficiente de determinación ( $R^2$ ), el error absoluto medio (MAE), el error cuadrático medio (MSE) y los valores RMS tanto para las salidas  $W$  como para los coeficientes  $A$ . Además, se incluye el tiempo computacional requerido por cada modelo durante el proceso de predicción, aspecto clave en aplicaciones de inferencia en tiempo real.

### 8.1.1. Resumen de Resultados

La Tabla 8.1 resume las métricas de error obtenidas por cada uno de los métodos al trabajar sobre la muestra base. Esta comparación cuantitativa permite evaluar tanto la precisión de las predicciones como la fidelidad en el ajuste de los coeficientes del frente de onda.

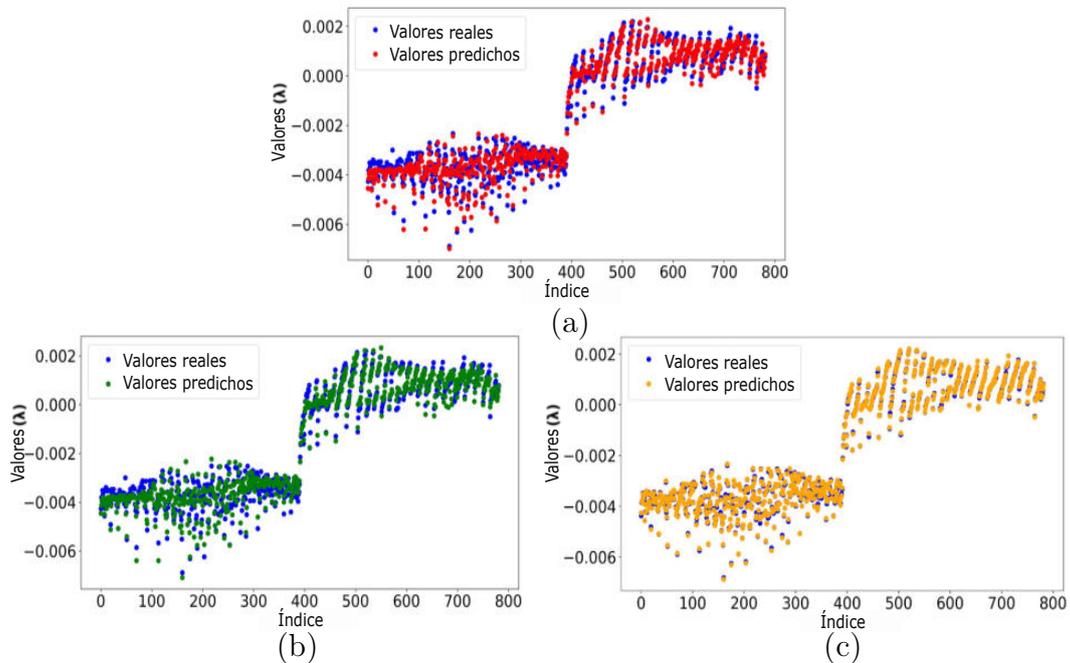
**Tabla 8.1:** Métricas de error y tiempo computacional para el ajuste de  $W$  utilizando la muestra base con los tres métodos evaluados.

Métrica	Mínimos Cuadrados	Gradiente Descendente	Red Hiperdinámica
Tiempo Computacional	0.4250 ms	52.2384 s	0.0273 ms
$R^2$ (%)	98.62	98.57	99.95
MAE ( $\lambda$ )	$2.2135 \times 10^{-4}$	$2.1787 \times 10^{-4}$	$4.2706 \times 10^{-5}$
MSE ( $\lambda^2$ )	$7.5606 \times 10^{-8}$	$7.8606 \times 10^{-8}$	$2.8795 \times 10^{-9}$
RMS de $W$ ( $\lambda$ )	$2.7497 \times 10^{-4}$	$2.8037 \times 10^{-4}$	$5.3661 \times 10^{-5}$
RMS de $A$ (adimensional)	$1.8183 \times 10^{-3}$	$3.0309 \times 10^{-3}$	$1.6824 \times 10^{-1}$

Los resultados muestran que la red hiperdinámica supera significativamente a los otros métodos en todas las métricas de error relacionadas con  $W$ , lo que evidencia una mayor capacidad para modelar con precisión las aberraciones del frente de onda. Si bien el valor de RMS en  $A$  es más elevado, esto puede atribuirse a la naturaleza generativa de su arquitectura, que permite mayor flexibilidad en la predicción de los coeficientes. Este incremento en la

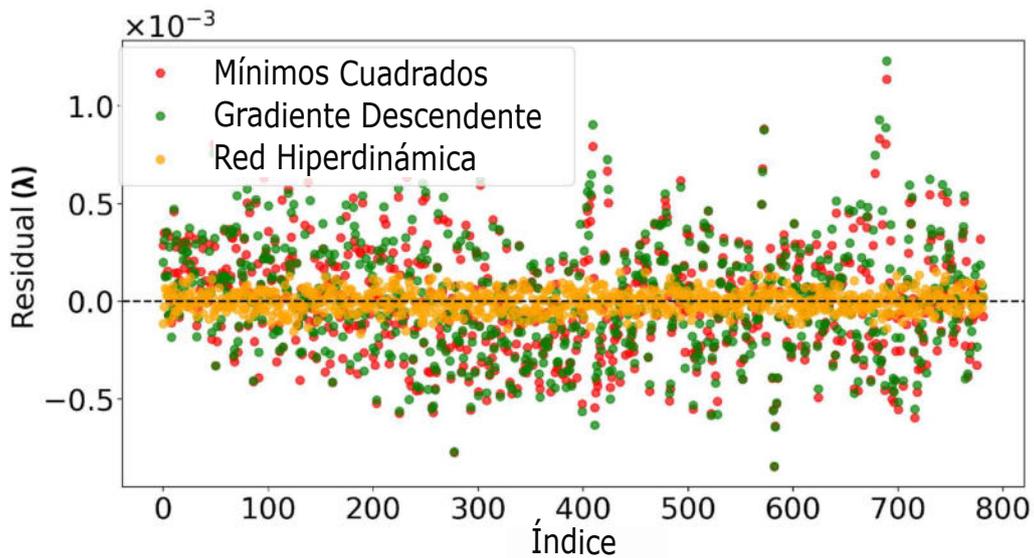
magnitud sugiere que los coeficientes de los polinomios de Zernike varían en un orden más amplio, lo cual refleja una mayor sensibilidad del modelo a las características locales del frente de onda. A pesar de ello, su rendimiento general demuestra una capacidad superior para aprender representaciones complejas. Además, su tiempo computacional es considerablemente menor que el del gradiente descendente, lo que la convierte en una opción especialmente eficiente para aplicaciones en tiempo real.

La Figura 8.1 ofrece una comparación visual de los valores predichos y reales de  $W$  para cada método, lo que permite apreciar directamente la calidad del ajuste realizado.



**Figura 8.1:** Comparación entre los valores predichos y reales de  $W$  para la muestra base, utilizando tres métodos diferentes: (a) mínimos cuadrados, (b) gradiente descendente y (c) red hiperdinámica. El eje  $x$  corresponde al índice de cada punto de datos en la rejilla del sensor Shack-Hartmann, y el eje  $y$  muestra los valores de desplazamiento de  $W$ , expresados en unidades de longitud de onda ( $\lambda$ ). Esta comparación ilustra la precisión de cada método para reproducir la forma original del frente de onda en este caso específico.

De la inspección visual, se observa que la red hiperdinámica presenta una alineación mucho más precisa con los valores reales, con desviaciones mínimas a lo largo de toda la muestra. Los métodos de mínimos cuadrados y gradiente descendente también muestran un buen ajuste, aunque con errores residuales más notorios, particularmente en zonas donde el comportamiento del frente de onda se vuelve más complejo. Esta diferencia se aprecia con mayor claridad en la Figura 8.2, donde se muestra la distribución de los residuales para cada método. En dicha figura, se observa que los residuales generados por la red hiperdinámica están más concentrados alrededor del cero y distribuidos de manera simétrica, lo que indica un ajuste más preciso y sin sesgos sistemáticos. En contraste, los otros métodos presentan una dispersión mayor, lo que sugiere la presencia de errores más amplios en determinadas regiones del dominio.



**Figura 8.2:** Distribución de los residuales para los tres métodos: mínimos cuadrados, gradiente descendente y red hiperdinámica. Los residuales se calculan como la diferencia entre los valores predichos y los valores reales de  $W$  para cada punto en la muestra base. El eje  $x$  indica el índice de cada punto de medición en el sensor Shack-Hartmann, y el eje  $y$  representa los residuales en unidades de longitud de onda ( $\lambda$ ). Este análisis proporciona información sobre la precisión de cada método y permite identificar posibles errores sistemáticos o anomalías.

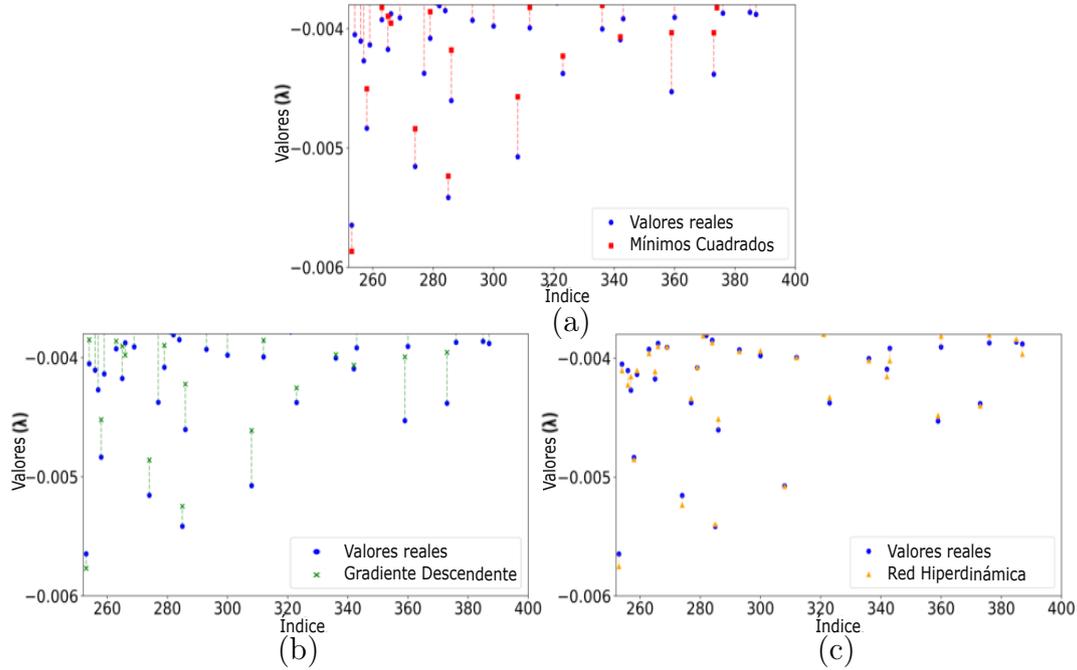
La combinación de métricas cuantitativas y análisis visual respalda la superioridad de la red hiperdinámica en esta etapa del estudio. Su precisión, eficiencia y capacidad de generalización la posicionan como una herramienta

robusta para el ajuste del frente de onda, especialmente en contextos donde se requiere alto rendimiento computacional. Sin embargo, también se destaca la utilidad del método de mínimos cuadrados como referencia rápida, y del gradiente descendente como un enfoque intermedio entre simplicidad y capacidad de adaptación.

### **8.1.2. Análisis en Presencia de Valores Atípicos**

Con el fin de evaluar la robustez de los métodos frente a condiciones adversas, se examinó una región del conjunto de datos caracterizada por la presencia de valores atípicos, principalmente localizados en los bordes. Este análisis permite observar cómo se comportan los modelos cuando enfrentan desviaciones significativas que podrían comprometer la precisión de las predicciones.

La Figura 8.3 presenta una comparación detallada de las predicciones generadas por los tres métodos en dicha región. Esta visualización permite identificar diferencias clave en la capacidad de cada enfoque para seguir el comportamiento real.



**Figura 8.3:** Comparación ampliada entre los valores predichos y reales de  $W$  en una región de interés de la muestra base, enfocándose en un área con valores atípicos. Los tres métodos mostrados son: (a) mínimos cuadrados, (b) gradiente descendente y (c) red hiperdinámica. El eje  $x$  muestra el índice de los puntos de datos afectados en el sensor Shack-Hartmann, y el eje  $y$  representa los valores de desplazamiento de  $W$ , en unidades de longitud de onda ( $\lambda$ ). Este acercamiento permite observar la robustez de cada método ante la presencia de anomalías en los datos.

En esta región, la red hiperdinámica destaca por su capacidad para mantener una alta fidelidad con respecto a los valores reales, incluso ante la presencia de anomalías. A diferencia de los métodos clásicos, que muestran desviaciones notables en sus predicciones, la arquitectura generativa de la red hiperdinámica le permite adaptarse mejor a cambios inesperados, reforzando su utilidad en aplicaciones donde la estabilidad frente a condiciones cambiantes es crucial.

### 8.1.3. Análisis en el Conjunto de Pruebas

Para validar la capacidad de generalización de los modelos, se realizó una evaluación independiente sobre un conjunto de pruebas compuesto por 500 muestras sintéticas. Estas muestras representan una porción del total de 3500

generadas en el Capítulo 6 y corresponden específicamente a aquellas que no fueron utilizadas durante el proceso de entrenamiento de los modelos de gradiente descendente y red hiperdinámica (descritos en las Secciones 7.3 y 7.4, respectivamente). Esta separación garantiza una evaluación justa y objetiva del desempeño de los modelos frente a datos completamente no vistos.

La Tabla 8.2 presenta un resumen de las métricas de error obtenidas por cada uno de los tres métodos. Se incluyen el tiempo computacional requerido por cada modelo, el error absoluto medio (MAE), el error cuadrático medio (MSE), así como los valores RMS de las matrices  $W$  y  $A$ .

**Tabla 8.2:** Métricas de error para los tres métodos en el conjunto de pruebas (500 muestras no vistas).

Métrica	Mínimos Cuadrados	Gradiente Descendente	Red Hiperdinámica
Tiempo Computacional	211.7550 ms	9.1315 h	188.7908 ms
MAE ( $\lambda$ )	$2.1909 \times 10^{-4}$	$2.1586 \times 10^{-4}$	$2.9559 \times 10^{-5}$
MSE ( $\lambda^2$ )	$7.3810 \times 10^{-8}$	$7.6464 \times 10^{-8}$	$1.4613 \times 10^{-9}$
RMS de $W$ ( $\lambda$ )	$2.7167 \times 10^{-4}$	$2.7651 \times 10^{-4}$	$2.9559 \times 10^{-5}$
RMS de $A$ (adimensional)	$1.8199 \times 10^{-3}$	$1.8224 \times 10^{-3}$	$1.6800 \times 10^{-1}$

Los resultados muestran con claridad que la red hiperdinámica mantiene un rendimiento sobresaliente incluso ante datos no vistos, alcanzando las métricas más bajas en MAE, MSE y RMS de  $W$ . Esto indica que su arquitectura generativa no solo permite un aprendizaje más profundo de las relaciones presentes en los datos, sino también una mejor capacidad de generalización en contextos complejos. A pesar de presentar un valor más elevado en el RMS de  $A$ , este comportamiento se encuentra alineado con su diseño estructural, que permite mayor variabilidad en los coeficientes generados para adaptarse a distintos escenarios.

En conclusión, este análisis confirma que la red hiperdinámica no solo supera a los enfoques clásicos en precisión, sino que también ofrece tiempos computacionales competitivos y una mayor robustez frente a datos no utilizados durante el entrenamiento. Estos resultados posicionan a este modelo como una herramienta altamente eficaz para tareas avanzadas de ajuste de aberraciones del frente de onda.

## Capítulo 9

# Conclusiones Generales

Esta tesis evaluó y comparó tres métodos para el ajuste de aberraciones del frente de onda utilizando datos provenientes de sensores Shack-Hartmann: el método de mínimos cuadrados, el gradiente descendente y la red hiperdinámica. A lo largo de este estudio, se analizaron sus desempeños en la reconstrucción de los valores de  $W$ , tanto en una única muestra como en un conjunto de pruebas de 500 muestras, considerando también su comportamiento en presencia de valores atípicos.

El método de mínimos cuadrados demostró un desempeño estable, con un coeficiente de determinación ( $R^2$ ) del 98.62% y un error absoluto medio (MAE) de  $2.2135 \times 10^{-4} \lambda$  en una única muestra. Sin embargo, su dependencia de la pseudo-inversa limita su robustez ante casos donde faltan puntos o existen datos superpuestos, lo que podría afectar su aplicación en sistemas ópticos reales. El gradiente descendente mostró resultados similares a mínimos cuadrados, alcanzando un  $R^2$  del 98.57% y un MAE de  $2.1787 \times 10^{-4} \lambda$  en una única muestra. Este método elimina la necesidad de la inversión de matrices, lo que lo convierte en una opción computacionalmente más eficiente para aplicaciones con recursos limitados o datos no estructurados. La red hiperdinámica sobresalió en todos los escenarios, alcanzando un  $R^2$  del 99.95% y un MAE de  $4.2706 \times 10^{-5} \lambda$  en una única muestra. En el conjunto de pruebas de 500 muestras, logró un MAE de  $2.9559 \times 10^{-5} \lambda$ , superando ampliamente a los otros dos métodos. Además, demostró ser más robusta ante valores atípicos y evidenció un comportamiento consistente en el análisis de residuales, donde estos estuvieron concentrados alrededor de cero.

Un punto distintivo de la red hiperdinámica es su capacidad para generar coeficientes que mejoran la cuantificación de aberraciones, como se refleja

en los valores RMS de  $A$ . Aunque estos valores son notablemente mayores comparados con otros métodos ( $1.6800 \times 10^{-1}$ ), indican una mayor capacidad para caracterizar errores del frente de onda de manera precisa, complementando su alta precisión en las métricas de error. Los resultados de esta tesis tienen importantes implicaciones para el desarrollo de sistemas ópticos de alta precisión. Mientras que mínimos cuadrados y el gradiente descendente son métodos confiables, la red hiperdinámica destaca como una herramienta superior debido a su capacidad para generalizar en conjuntos de prueba diversos y proporcionar coeficientes que optimizan la calibración de aberraciones. Estas propiedades la convierten en una opción prometedora para aplicaciones como telescopios adaptativos, sistemas de visión computarizada y otros dispositivos que dependen de ajustes precisos del frente de onda.

Más allá de su precisión predictiva, la red hiperdinámica también demostró una mayor eficiencia computacional. En el caso de una sola muestra (ver Tabla 8.1), alcanzó un tiempo de cómputo de 0.0273 ms, significativamente menor que los 0.4250 ms requeridos por el método de mínimos cuadrados. De manera similar, en el conjunto de pruebas (ver Tabla 8.2), la red hiperdinámica necesitó 188.7908 ms, en comparación con los 211.7550 ms del método de mínimos cuadrados. Este resultado refuerza la ventaja de las redes hiperdinámicas, ya que ofrecen una mayor precisión, lo que las hace adecuadas para aplicaciones en tiempo real donde el ajuste rápido del frente de onda es esencial.

Aunque la red hiperdinámica demostró ser el método más efectivo, presenta un costo computacional considerable debido a su complejidad intrínseca. Por lo tanto, futuros trabajos podrían enfocarse en optimizar su implementación mediante estrategias avanzadas, como la optimización bayesiana y la paralelización en hardware especializado. Además, sería valioso explorar la aplicación de la red hiperdinámica en otros contextos, como la calibración en tiempo real o su combinación con metodologías híbridas que integren las fortalezas del gradiente descendente y mínimos cuadrados. Este enfoque podría abrir nuevas posibilidades para sistemas ópticos integrados con inteligencia artificial, permitiendo ajustes más rápidos y precisos en entornos dinámicos.

En conclusión, esta tesis ha demostrado que la red hiperdinámica representa un avance significativo en el ajuste de aberraciones del frente de onda. Su precisión, robustez y capacidad para manejar condiciones complejas la posicionan como una metodología de elección para aplicaciones ópticas modernas. Aunque existen áreas que requieren mejoras, los resultados obtenidos sientan las bases para futuras investigaciones y desarrollos en la integración de inteligencia artificial en sistemas ópticos de alta precisión.

# Bibliografía

- Barakat, R. and Newsam, G. (1980). Numerically stable iterative method for the inversion of wave-front aberrations from measured point-spread-function data. *Journal of the Optical Society of America*, 70(10):1255–1263. [10.1364/JOSA.70.001255](#).
- Born, M. and Wolf, E. (1999). *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, 7th edition.
- Chai, X., Lin, X., Chen, H., Wei, Q., and Yu, Y. (2024). Zernike polynomials fitting of arbitrary shape wavefront. In Wang, H., Cheng, F., Dang, C. H., Danner, A. J., and Kemao, Q., editors, *International Conference on Optical and Photonic Engineering (icOPEN 2023)*, volume 13069 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 1306912. [10.1117/12.3023145](#).
- Chauhan, V. K., Zhou, J., Lu, P., Molaei, S., and Clifton, D. A. (2024). A brief review of hypernetworks in deep learning. *Artificial Intelligence Review*, 57(9):250. [10.1007/s10462-024-10862-8](#).
- Duvall, J., Joly, M., Duraisamy, K., and Sarkar, S. (2024). Design-variable hypernetworks for flowfield emulation and shape optimization of compressor airfoils. *AIAA Journal*, 62(2):741–757. [10.2514/1.J063156](#).
- Endo, Y., Amako, Y., and Fukuchi, T. (2018). Improving techniques for shack-hartmann wavefront sensing. [Available online](#).
- Fowler, J. and Landman, R. (2023a). Tempestas ex machina: A review of machine learning methods for wavefront control. [arXiv:2309.00730](#).
- Fowler, J. and Landman, R. (2023b). Tempestas ex machina: A review of machine learning methods for wavefront control. *arXiv preprint*, arXiv:2309.00730. [arXiv:2309.00730](#).
- Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training

- deep feedforward neural networks. *Journal of Machine Learning Research - Proceedings Track*, 9:249–256. [proceedings.mlr.press/v9/glorot10a](https://proceedings.mlr.press/v9/glorot10a).
- Goel, N. and Ganotra, D. (2024). Sensorless wavefront sensing: Techniques, applications, and challenges. *Atmospheric and Oceanic Optics*, 37(1):103–117. [10.1134/S1024856023700148](https://doi.org/10.1134/S1024856023700148).
- Guo, Y., Zhong, L., Min, L., Wang, J., Wu, Y., Chen, K., Wei, K., and Rao, C. (2022). Adaptive optics based on machine learning: a review. *Opto-Electronic Advances*, 5(7):200082. [10.29026/oea.2022.200082](https://doi.org/10.29026/oea.2022.200082).
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 2 edition.
- Hecht, E. (2002). *Optics*. Addison-Wesley, 4th edition.
- Herrmann, J. (1980). Least-squares wave front errors of minimum norm. *Journal of the Optical Society of America*, 70(1):28–35. [10.1364/JOSA.70.000028](https://doi.org/10.1364/JOSA.70.000028).
- Jara, M. (2025). Integración de técnicas de inteligencia artificial en la calibración de aberraciones en frente de onda mediante el método shack-hartmann. [10.5281/zenodo.15087610](https://doi.org/10.5281/zenodo.15087610).
- Karttunen, H., Pekka, K., Oja, H., Poutanen, M., and Donner, K. (2017). *Fundamental Astronomy*. Springer-Verlag Berlin Heidelberg, 6 edition.
- Keskar, N. S. and Socher, R. (2017). Improving generalization performance by switching from adam to sgd. *arXiv preprint*, arXiv:1712.07628. [arXiv:1712.07628](https://arxiv.org/abs/1712.07628).
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint*, arXiv:1412.6980. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- Lin, X., Rivenson, Y., Yardimci, N. T., Veli, M., Luo, Y., Jarrahi, M., and Ozcan, A. (2018). All-optical machine learning using diffractive deep neural networks. *Science*, 361(6406):1004–1008. [10.1126/science.aat8084](https://doi.org/10.1126/science.aat8084).
- Liu, X., Morris, T., Saunter, C., de Cos Juez, F. J., González-Gutiérrez, C., and Bardou, L. (2020). Wavefront prediction using artificial neural networks for open-loop adaptive optics. *Monthly Notices of the Royal Astronomical Society*, 496(1):456–464. [10.1093/mnras/staa1558](https://doi.org/10.1093/mnras/staa1558).
- Malacara, D. (2007). *Optical Shop Testing*. John Wiley & Sons, 3rd edition.
- Noel, Z. A., Kemnetz, M. R., Gordeyev, S., and Rennie, R. M. (2023). Robust,

- regularized wavefront reconstruction of shack-hartmann data in discontinuous flow. *Proc. SPIE*, 12693:1269318. [10.1117/12.2677884](https://doi.org/10.1117/12.2677884).
- Noll, R. J. (1976). Zernike polynomials and atmospheric turbulence. *Journal of the Optical Society of America*, 66(3):207–211. <https://doi.org/10.1364/JOSA.66.000207>.
- Novak, J., Novak, P., and Miks, A. (2007). Application of shack-hartmann wavefront sensor for testing optical systems. *Proc. SPIE*, 6609:660915. [10.1117/12.739665](https://doi.org/10.1117/12.739665).
- Pedrotti, F. L. and Pedrotti, L. S. (2006). *Introduction to Optics*. Pearson, 3rd edition.
- Pérez, S., Buendía, A., González, C., Rodríguez, J., Iglesias, S., Fernández, J., and Cos, F. J. D. (2024). Enhancing open-loop wavefront prediction in adaptive optics through 2d-lstm neural network implementation. *Photonics*, 11(3):240. [10.3390/photonics11030240](https://doi.org/10.3390/photonics11030240).
- Rayces, J. (1964a). Exact relation between wave aberration and ray aberration. *Optica Acta: International Journal of Optics*, 11(2):85–88. [10.1080/713817854](https://doi.org/10.1080/713817854).
- Rayces, J. L. (1964b). Exact relation between wave aberration and ray aberration. *Optica Acta: International Journal of Optics*, 11(2):85–88. <https://doi.org/10.1080/713817854>.
- Reyna, M. A. J., Peña, C. A. G., Vázquez, J. H., Armenta, J. J. T., and Tijerina, E. G. P. (2025). Calibration of wavefront aberrations using supervised learning and deep learning for a shack–hartmann wavefront sensor. *Optical Engineering*, 64(5):053101. [10.1117/1.OE.64.5.053101](https://doi.org/10.1117/1.OE.64.5.053101).
- Shack, R. V. (1971). Production and use of a lenticular hartmann screen. *Spring Meeting of Optical Society of America*, 656. <https://cir.nii.ac.jp/crid/1573387450670986240>.
- Smith, W. J. (2007). *Modern Optical Engineering: The Design of Optical Systems*. McGraw-Hill, 4th edition.
- Southwell, W. (1980). Wave-front estimation from wave-front slope measurements. *Journal of the Optical Society of America*, 70(8):998–1006. [10.1364/JOSA.70.000998](https://doi.org/10.1364/JOSA.70.000998).
- Tani, L. and Veelken, C. (2024). Comparison of bayesian and particle swarm algorithms for hyperparameter optimisation in machine learning

- applications in high energy physics. *Computer Physics Communications*, 294:108955. [10.1016/j.cpc.2023.108955](https://doi.org/10.1016/j.cpc.2023.108955).
- Tyson, R. K. (2015). *Principles of Adaptive Optics*. CRC Press.
- Wetzstein, G., Ikoma, H., Metzler, C., and Peng, Y. (2020). Deep optics: Learning cameras and optical computing systems. In *2020 54th Asilomar Conference on Signals, Systems, and Computers*, pages 1313–1315. IEEE. [10.1109/IEEECONF51394.2020.9443575](https://doi.org/10.1109/IEEECONF51394.2020.9443575).
- Xu, Y., Sarah, R., Habib, A., Liu, Y., and Khoda, B. (2024). Constraint based bayesian optimization of bioink precursor: a machine learning framework. *Biofabrication*, 16(4):045031. [10.1088/1758-5090/ad716e](https://doi.org/10.1088/1758-5090/ad716e).
- Zernike, F. (1934). Beugungstheorie des schneidenverfahrens und seiner verbesserten form, der phasenkontrastmethode. *Physica*, 1(7–12):689–704. [https://doi.org/10.1016/S0031-8914\(34\)80259-5](https://doi.org/10.1016/S0031-8914(34)80259-5).

# Apéndice A

## Apéndice

### A.1. Artículo científico resultante de esta investigación

En este apéndice se presenta el artículo titulado *Calibration of wavefront aberrations using supervised learning and deep learning for a Shack–Hartmann wavefront sensor*, elaborado a partir de los resultados obtenidos durante el desarrollo de esta tesis. El artículo fue redactado por Mario Alberto Jara Reyna, Carlos Alberto Guerrero Peña, Joel Herrera Vázquez, Juan José Tapia Armenta y Eduardo Gerardo Pérez Tijerina, y fue publicado en la revista *Optical Engineering* en su volumen 64, número 5, correspondiente al año 2025 (Reyna et al., 2025).

#### A.1.1. Introducción

La turbulencia atmosférica distorsiona el frente de onda de la luz que llega a los sistemas ópticos, afectando la calidad de imagen. Para corregir estas aberraciones, es necesario estimar con precisión el frente de onda, tarea en la que los sensores Shack–Hartmann resultan fundamentales gracias a su capacidad para medir desplazamientos locales mediante una matriz de microlentes. El propósito de este trabajo fue explorar y comparar diferentes enfoques para realizar dicha tarea, partiendo de datos obtenidos y generando un conjunto de muestras sintéticas para evaluar el desempeño de los métodos.

En el artículo se presentan tres metodologías: el ajuste mediante mínimos cuadrados como referencia clásica; una implementación basada en gradiente descendiente utilizando el optimizador Adam; y un enfoque alternativo

a través de redes hiperdinámicas, diseñadas para adaptarse mejor a distintas configuraciones de entrada. Se discuten aspectos como la generación de muestras sintéticas a partir de mediciones reales, el procesamiento de los desplazamientos de spots, y la construcción de las matrices necesarias para el ajuste del frente de onda mediante polinomios de Zernike, y la evaluación de desempeño mediante distintas métricas de error.

El contenido del artículo representa una síntesis de los avances más relevantes alcanzados durante esta investigación y se incluye a continuación en su versión final publicada.

#### **A.1.2. Manuscrito publicado en la revista *Optical Engineering***

# Calibration of wavefront aberrations using supervised learning and deep learning for a Shack–Hartmann wavefront sensor

Mario Alberto Jara Reyna <sup>a</sup>, Carlos Alberto Guerrero Peña <sup>a,\*</sup>,  
Joel Herrera Vázquez <sup>b</sup>, Juan José Tapia Armenta <sup>c</sup>, and  
Eduardo Gerardo Pérez Tijerina <sup>a</sup>

<sup>a</sup>Universidad Autónoma de Nuevo León, Facultad de Ciencias Físico Matemáticas,  
San Nicolás de los Garza, Nuevo León, México

<sup>b</sup>Universidad Nacional Autónoma de México, Instituto de Astronomía, Ensenada, Baja California, México

<sup>c</sup>Instituto Politécnico Nacional – CITEDI, Tijuana, Baja California, México

---

**ABSTRACT.** We evaluate wavefront fitting methods in single-sample and multisample scenarios, focusing on least squares (LSQ), gradient descent, and dynamic hypernetwork approaches. LSQ is a well-established method, whereas gradient descent offers a low-cost, adaptable solution. By contrast, the dynamic hypernetwork achieves the highest precision, with a mean absolute error of  $2.9559 \times 10^{-5}\lambda$ , significantly lower than the  $2.1909 \times 10^{-4}\lambda$  obtained by LSQ. In addition, it reduces computation time to 0.0273 ms per sample, outperforming LSQ's 0.4250 ms, and exhibits superior robustness in the presence of outliers. On the full testing dataset, the dynamic hypernetwork maintains consistent performance, demonstrating its ability to generalize effectively under varying levels of atmospheric turbulence. These results demonstrate that AI-based methods—particularly dynamic hypernetworks—not only address LSQ's limitations but also enable fast and accurate wavefront fitting, making them promising tools for real-time optical applications.

© 2025 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.OE.64.5.053101](https://doi.org/10.1117/1.OE.64.5.053101)]

**Keywords:** wavefront fitting; Shack–Hartmann wavefront sensor; supervised learning; deep learning; gradient descent; dynamic hypernetwork

Paper 20241217G received Dec. 24, 2024; revised Apr. 5, 2025; accepted Apr. 8, 2025; published May 5, 2025.

---

## 1 Introduction

The acquisition of the wavefront obtained from an optical system is essential to evaluate its performance and verify aspects such as alignment, manufacturing errors, or compliance with design specifications.<sup>1</sup> Aberrations in the wavefront can lead to significant image degradations, which are especially critical in applications requiring high accuracy, such as astronomy and advanced imaging technologies.<sup>2</sup>

Traditional methods, such as the Shack–Hartmann wavefront sensor (SHWFS), have been widely employed for wavefront fitting.<sup>3</sup> The SHWFS is designed to measure local wavefront gradients by analyzing the displacement of spots generated by a microlens array, offering a reliable method for evaluating the wavefront shape.<sup>4</sup> However, traditional methods such as least squares (LSQ) fitting can be sensitive to outliers, potentially affecting precision. The reliance of LSQ on the pseudoinverse makes it vulnerable to decreased precision under these conditions, which can result in the reconstruction of nonexistent aberrations. These issues are particularly

---

\*Address all correspondence to Carlos Alberto Guerrero Peña, [carlos.guerrerop@uanl.edu.mx](mailto:carlos.guerrerop@uanl.edu.mx)

problematic in scenarios where incomplete or ambiguous measurements hinder accurate wavefront fitting.<sup>5</sup>

In adaptive optics systems, particularly in astronomy and laser communications, accurate wavefront reconstruction is crucial for maintaining image resolution and system performance. Incomplete or ambiguous measurements—caused by atmospheric turbulence, optical misalignments, or sensor noise—can lead to significant errors in real-time corrections, ultimately degrading the system's ability to compensate for distortions.<sup>6,7</sup> Traditional LSQ methods lack adaptability as they treat each measurement independently, disregarding the temporal evolution of wavefront changes.<sup>8</sup> By contrast, artificial intelligence (AI)-based methods can exploit historical wavefront variations to enhance adaptability, improve stability, and provide predictive capabilities for future corrections.<sup>7,9</sup> This predictive advantage is particularly valuable in scenarios where rapid wavefront updates are required, such as in high-speed astronomical observations or real-time laser beam shaping.<sup>9,10</sup> By leveraging AI-driven techniques, we can develop more robust reconstructors that continuously refine their estimates, overcoming the limitations of static, frame-by-frame reconstructions in classical methods. In addition, the efficiency of these approaches can be assessed not only by their accuracy but also by their prediction speed, which is critical for real-time applications.<sup>10</sup>

Recent advances, such as the application of deep learning in optical computation, have demonstrated the immense potential of AI-based approaches, particularly in addressing complex, high-dimensional problems commonly encountered in optical systems.<sup>11,12</sup>

This study focuses on integrating supervised learning techniques, including gradient descent,<sup>13</sup> and advanced deep learning approaches, such as dynamic hypernetworks,<sup>14</sup> into the wavefront aberration calibration process using the SHWFS method. Gradient descent enables continuous optimization of Zernike polynomials, enhancing calibration efficiency. Meanwhile, dynamic hypernetworks provide a flexible and adaptive framework, exploring alternative methodologies for dynamic parameter estimation, departing from traditional LSQ approaches. Their implementation, performance, and potential advantages over traditional techniques are discussed.

## 2 Theory and Analysis

Wavefronts in optical systems describe the phase of light as it propagates. Ideal planar or spherical wavefronts can become aberrated due to optical imperfections, causing distortions that degrade image quality.<sup>15</sup> The SHWFS measures transverse aberrations by detecting spot displacements from their ideal positions on a charge-coupled device (CCD). Each microlens in the array focuses light into subareas, providing local wavefront slope information.<sup>16</sup> The wavefront  $W$  represents the phase of light across the aperture of the optical system. Its relationship with the transverse aberrations can be mathematically expressed through its partial derivatives in the  $x$  and  $y$  directions. To clarify, Eq. (1) describes the wavefront slope in terms of measured spot displacements, whereas Eq. (2) relates the slope to the transverse aberrations

$$\frac{\partial W}{\partial x} = \frac{L_x - P_x}{f_m}, \quad \frac{\partial W}{\partial y} = \frac{L_y - P_y}{f_m}, \quad (1)$$

$$\frac{\partial W}{\partial x} = \frac{-A_T^x}{f_m}, \quad \frac{\partial W}{\partial y} = \frac{-A_T^y}{f_m}, \quad (2)$$

where  $L_x, L_y$  and  $P_x, P_y$  represent the unaberrated and aberrated spot coordinates, respectively,  $f_m$  is the focal length of the microlens array, and  $A_T^x, A_T^y$  are the transverse aberrations.

Wavefront aberrations are commonly modeled using Zernike polynomials,<sup>17</sup> which decompose the wavefront into components such as coma or astigmatism. This study employs 37 Zernike terms, balancing precision and computational feasibility.<sup>18</sup> Following the numbering scheme proposed by Noll<sup>17</sup> (see Table 11 in the Appendix), the wavefront  $W$  is expressed as

$$W(x_j, y_j) = \sum_{i=2}^N A_i Z_i(x_j, y_j), \quad (3)$$

where  $A_i$  are coefficients to be determined, and  $Z_i(x_j, y_j)$  are the Zernike polynomials at pupil coordinates. The first coefficient ( $A_0$ ) is omitted as it represents the piston term, which is

conventionally set to 0. Coefficients  $A_i$  provide insight into the system's aberrations, essential for wavefront fitting. Although  $A$  and  $W$  are mathematically vectors, they are treated as matrices throughout this study to simplify their handling and notation.

### 2.1 Equation System for Wavefront Fitting

Wavefront fitting translates SHWFS measurements into a system of linear equations

$$Z * A = W, \quad (4)$$

where  $Z$  is a matrix of Zernike polynomial derivatives,  $A$  contains polynomial coefficients, and  $W$  represents measured wavefront slopes. Fitting minimizes the difference between  $Z * A$  and  $W$ , effectively describing the aberrations.<sup>16,18</sup>

### 2.2 Least Squares and Pseudoinverse

The LSQ method<sup>19</sup> solves Eq. (4) by minimizing the squared differences between  $W$  and  $Z * A$ , yielding

$$A = (Z^T Z)^{-1} Z^T W. \quad (5)$$

It remains a robust method widely used for wavefront fitting.

### 2.3 Gradient Descent with Adam

Gradient descent is a supervised learning method that optimizes  $A$  iteratively by minimizing a loss function, such as the mean absolute error (MAE)

$$A_{\text{new}} = A_{\text{previous}} - \alpha \nabla J(A), \quad (6)$$

where  $\alpha$  is the learning rate and  $\nabla J(A)$  is the loss function gradient.

To enhance convergence, the Adam optimizer adaptively adjusts learning rates using gradient moments<sup>20</sup>

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla L(A^{(t)}), \quad (7)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla L(A^{(t)}))^2, \quad (8)$$

$$A^{(t+1)} = A^{(t)} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}, \quad (9)$$

where  $m_t$  is the exponentially weighted average of gradients (first-moment estimate),  $v_t$  is the exponentially weighted average of squared gradients (second-moment estimate),  $\beta_1, \beta_2$  are decay rate hyperparameters for  $m_t$  and  $v_t$ , respectively,  $\nabla L(A^{(t)})$  is the gradient of the loss function with respect to  $A$  at iteration  $t$ ,  $\hat{m}_t$  and  $\hat{v}_t$  are bias-corrected versions of  $m_t$  and  $v_t$ ,  $\eta$  is the learning rate, and  $\epsilon$  is a small constant added for numerical stability.

The Adam optimizer is particularly relevant for wavefront fitting as it efficiently handles challenges such as noisy data and varying gradient scales, ensuring faster and more stable convergence. This is demonstrated in Sec. 3.5.2, where the convergence curve highlights Adam's ability to stabilize the learning process, reducing execution time and improving learning behavior. By adaptively adjusting learning rates for each parameter, Adam improves efficiency and is well suited for scenarios where precise adjustments of  $A$  are critical to accurately model wavefront aberrations.

### 2.4 Dynamic Hypernetwork

The dynamic hypernetwork is an advanced deep learning model that generates coefficients  $A$  dynamically based on the wavefront measurements  $W$  and matrix  $Z$ . Unlike traditional methods, it adapts to varying input conditions, improving precision and generalization.<sup>14,21</sup> The network learns a function  $H(Z, W; \theta)$  that directly maps inputs to Zernike polynomial coefficients without explicitly solving the equation system

$$A = H(Z, W; \theta), \quad (10)$$

where  $H$  is the hypernetwork parameterized by  $\theta$ . This approach avoids traditional computations, offering efficiency in handling noisy or incomplete data.

The performance of the dynamic hypernetwork is discussed in Sec. 3.5.4, where we analyze the evolution of the training and validation loss, measured as the MAE between predicted and target wavefront values. The consistent reduction in loss across epochs highlights the model's ability to generalize effectively to unseen data.

In addition to its robust performance, the dynamic hypernetwork offers several practical advantages. It continuously adjusts to measurement variations, making it suitable for dynamic optical systems. By minimizing the MAE, the model improves the accuracy of the estimated Zernike coefficients used to reconstruct the wavefront. Furthermore, it bypasses explicit equation solving by directly learning the solutions, improving efficiency in high-complexity scenarios. These characteristics make the dynamic hypernetwork a powerful tool for addressing challenges in wavefront fitting, particularly in systems with noisy or incomplete data.

## 2.5 Hyperparameter Optimization Using Bayesian Approach

Bayesian optimization, a probabilistic model-based approach, was utilized to efficiently estimate the optimal hyperparameters for both the gradient descent model and the dynamic hypernetwork. This method minimizes the computational burden of exhaustive searches by evaluating configurations based on the expected improvement criterion

$$EI(\mathbf{x}) = \mathbb{E}[\max(0, f(\mathbf{x}^+) - f(\mathbf{x}))], \quad (11)$$

where  $f(\mathbf{x})$  represents the loss function and  $f(\mathbf{x}^+)$  denotes the best observed value so far.

Bayesian optimization has proven effective in enhancing search efficiency across high-dimensional hyperparameter spaces, as demonstrated by Xu et al.<sup>22</sup> and Tani and Veelken.<sup>23</sup>

## 2.6 Adjustment Method Based on Average Displacements

Wavefront measurements in real-world scenarios often differ from the ideal training set due to optical misalignments or system variations. To address this, we implemented an adjustment method that aligns missing points by calculating the vector average of the four nearest neighbors in the ideal set

$$\vec{d}_{\text{avg}} = \frac{1}{4} \sum_{j=1}^4 (\vec{p}_{\text{real}}^j - \vec{p}_{\text{ideal}}^j), \quad (12)$$

where  $\vec{d}_{\text{avg}}$  is the average displacement vector. Each missing point  $p_i$  is adjusted as

$$p_{\text{adjusted}} = p_i - \vec{d}_{\text{avg}}. \quad (13)$$

This method ensures consistency with the training set, particularly near wavefront boundaries, where perturbations are more common.

## 2.7 Evaluation Metrics

To analyze model performance, the following metrics are employed.

Correlation coefficient, which indicates the proportion of variability in the data explained by the model, is calculated as

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (14)$$

where  $y_i$  are the actual values,  $\hat{y}_i$  are the predictions, and  $\bar{y}$  is the mean of the actual values.

MAE measures the average magnitude of errors, defined as

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (15)$$

whereas mean squared error (MSE) penalizes large errors, calculated as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (16)$$

Finally, root mean square (RMS) error provides a direct interpretation in terms of the data scale, defined as

$$\text{RMS} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (17)$$

These metrics were selected to evaluate explanatory capability ( $R^2$ ), average precision (MAE), sensitivity to large errors (MSE), and error scale (RMS), providing a comprehensive analysis of model performance.

The theoretical principles outlined in Sec. 2 provide the foundation for understanding wavefront aberrations and their representation using Zernike polynomials. Building on this framework, Sec. 3 introduces the methodologies employed to solve the wavefront fitting problem, exploring both classical and machine learning approaches to achieve precise aberration correction.

### 3 Datasets and Network Training

#### 3.1 Input Data

To train and evaluate the wavefront fitting models, simulated data were generated based on real measurements obtained from a Shack–Hartmann sensor under various controlled conditions. The measurements were acquired using a Shack–Hartmann wavefront sensor (WFS150-7AR) coupled to a monochromatic camera. The sensor includes a microlens array, arranged in a  $39 \times 31$  grid (1209 microlenses), with each microlens having a diameter of 0.15 mm and a focal length of 5.2 mm. The optical system consisted of a 1.5 m aperture telescope from the Observatorio Astronómico Nacional de San Pedro Mártir (OAN-SPM), with a focal ratio of  $f/13.5$ , a telescope focal length of 19,500 mm, and a collimator with a focal length of 30 mm. These parameters define the characteristics of the dataset, ensuring that all simulated data inherit the same optical conditions as the real system.

Spot deviations were processed using an image processing algorithm to clean the data, identify spot centroids, and align them with theoretical positions. The resulting information was then used to construct matrices for fitting the coefficients  $A$ , which represent the weights of the Zernike polynomials. All the datasets used throughout this paper are available in a Zenodo repository.<sup>24</sup>

When measurements for some points were missing, simulating potential system failures, the interpolation method based on average displacements, as described in Sec. 2.6, was applied to ensure consistency with the training set.

#### 3.2 Simulating Data by Adding Noise

To introduce variability while preserving physical consistency, Gaussian noise was applied to the measured displacements, considering a seeing range of 0.8 to 2.0 arcsec. The noise was scaled based on the telescope focal length and collimator focal length, following the standard relation

$$\text{sigma} = \frac{\text{seeing(arcsec)} \times f_{\text{telescope}}}{206265}, \quad (18)$$

where  $\text{sigma}$  represents the standard deviation of the seeing-induced perturbations. The noise values were further adjusted using the pupil amplification factor to maintain consistency with the optical system. The final spot positions were computed by adding these perturbations to the real measurements, ensuring that the dataset accurately reflects realistic wavefront deviations.

To confirm that the Shack–Hartmann sensor operated within its linear response range,<sup>25</sup> we analyzed the maximum displacement of spots across the 3500 generated samples and compared them with the physical constraints of the system. As described in Sec. 3.1, the Shack–Hartmann sensor contains a microlens array with a  $39 \times 31$  grid. The maximum permissible displacement for a spot to remain in the linear range of the sensor is 0.075 mm (half of the microlens diameter).

For each perturbed spot, we computed the Euclidean distance between its position and the closest ideal centroid (the expected location in the absence of noise) and recorded the largest displacement in each simulation. The statistical results obtained are summarized in Table 1.

The largest recorded displacement in any sample was  $3.1997 \times 10^{-2}$  mm, which remains well below the 0.075 mm threshold, confirming that all simulated aberrations remained within the linear response range of the SHWFS. By carefully controlling the noise parameters and ensuring that no simulated displacement exceeded the microlens subaperture, we verified that the

**Table 1** Summary of maximum spot displacement statistics.

Metric	Value
Total samples	3500
Max distance (mm)	$3.1997 \times 10^{-2}$
Min distance (mm)	$3.1193 \times 10^{-2}$
Mean distance (mm)	$3.1634 \times 10^{-2}$
Standard deviation (mm)	$9.0147 \times 10^{-5}$
Exceeded samples	0

wavefront reconstructions were not affected by nonlinearities, preserving the validity of the linear approximation required for accurate wavefront fitting. The complete implementation is available in the script `data_noise.py` within the Zenodo repository.<sup>24</sup>

### 3.3 Generation of $Z$ and $W$ Matrices

Section 2.1 introduced the  $Z$  and  $W$  matrices, essential for aberration adjustment methods. This section explains in detail how these matrices are generated.

- Matrix  $Z$ : This matrix contains the partial derivatives of the Zernike polynomials with respect to  $x$  and  $y$  at each sensor sampling point  $(P_{xi}, P_{yi})$ . Each column corresponds to a coefficient  $A_n$ , and rows are filled with the derivatives for each function, first with respect to  $x$  and then  $y$ . The  $Z$  matrix is represented as follows:

$$Z = \begin{bmatrix} \frac{\partial Z(A_0, P_{x0}, P_{y0})}{\partial x} & \dots & \frac{\partial Z(A_n, P_{x0}, P_{y0})}{\partial x} \\ \vdots & \ddots & \vdots \\ \frac{\partial Z(A_0, P_{xn}, P_{yn})}{\partial x} & \dots & \frac{\partial Z(A_n, P_{xn}, P_{yn})}{\partial x} \\ \frac{\partial Z(A_0, P_{x0}, P_{y0})}{\partial y} & \dots & \frac{\partial Z(A_n, P_{x0}, P_{y0})}{\partial y} \\ \vdots & \ddots & \vdots \\ \frac{\partial Z(A_0, P_{xn}, P_{yn})}{\partial y} & \dots & \frac{\partial Z(A_n, P_{xn}, P_{yn})}{\partial y} \end{bmatrix}. \quad (19)$$

- Matrix  $W$ : The  $W$  matrix is constructed from the wavefront slope measurements at each microlens point, taking partial derivatives of the wavefront in the  $x$  and  $y$  directions. These derivatives represent the local slopes and are organized in  $W$  as follows:

$$W = \begin{bmatrix} \frac{\partial W(P_{x0}, P_{y0})}{\partial x} \\ \vdots \\ \frac{\partial W(P_{xn}, P_{yn})}{\partial x} \\ \frac{\partial W(P_{x0}, P_{y0})}{\partial y} \\ \vdots \\ \frac{\partial W(P_{xn}, P_{yn})}{\partial y} \end{bmatrix}. \quad (20)$$

These values are essential for calculating the coefficients  $A_n$  as they represent the local wavefront inclinations.

From this process, 3500  $Z$  and  $W$  matrices were obtained, forming the foundation for evaluating the models described in Secs. 2.2–2.4.

### 3.4 Statistical Analysis of Aberrations

A detailed statistical characterization of the wavefront aberrations used in this study is essential to ensure a robust evaluation of the reconstruction methods. In particular, it is necessary to quantify

both the overall aberration magnitude and its decomposition into  $x$ - and  $y$ -components. This characterization is crucial as the aberrations define the input space for the reconstruction models and determine their effectiveness under different wavefront distortion scenarios.

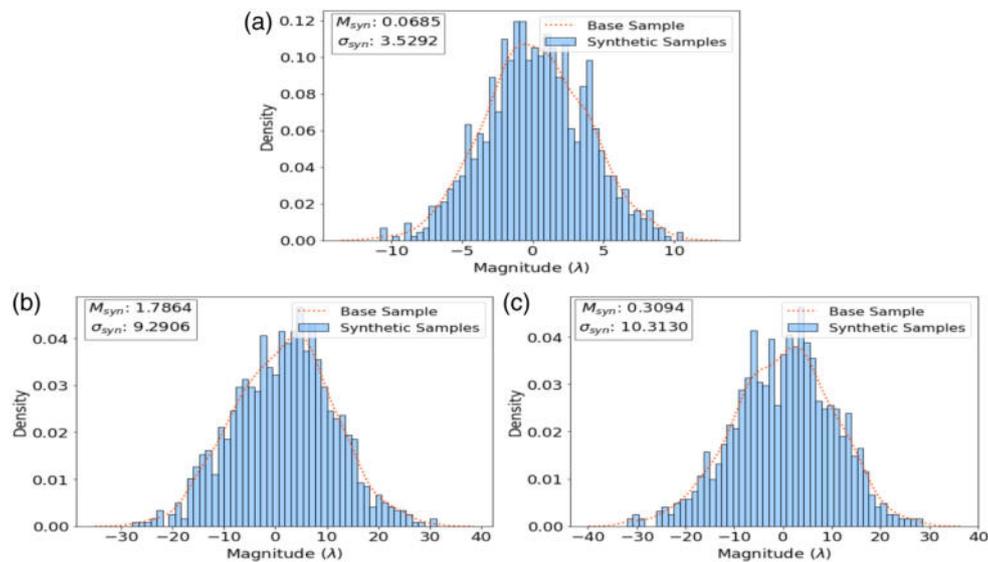
Table 2 presents a summary of the aberration magnitudes in the simulated dataset, including the mean, standard deviation, and range of values across the 3500 synthetic samples. The dataset was generated by combining Zernike polynomials under controlled conditions, ensuring a diverse distribution of wavefront distortions. The synthetic dataset preserves the statistical properties of the real sample while incorporating controlled perturbations.

To assess the consistency between real and synthetic data, Fig. 1 compares the probability distributions of three selected aberrations from both datasets. Because displaying all 37 aberrations would lead to excessive visual complexity, we focus on three representative cases that span low, mid, and high orders:  $Z_4$  (low-order astigmatism  $Y$ ),  $Z_{14}$  (mid-order tetrafoil  $Y$ ), and  $Z_{24}$  (high-order secondary tetrafoil  $X$ ). The histograms demonstrate that the synthetic samples effectively replicate the statistical characteristics of the real dataset, preserving the distribution of wavefront distortions while introducing controlled variations. This close agreement indicates that the generated dataset is suitable for training and evaluating the reconstruction models without introducing artificial biases.

For a more detailed analysis, the Appendix presents statistical tables summarizing the aberration data: (1) the overall aberration magnitudes in both the base and synthetic datasets (see Table 12), (2) the  $x$ - and  $y$ -components of the aberrations in the base dataset (see Table 13), and (3) the  $x$ - and  $y$ -components of the aberrations in the synthetic dataset (see Table 14).

**Table 2** Summary of aberration magnitude statistics for base and synthetic samples.

Metric	Base sample ( $\lambda$ )	Synthetic samples ( $\lambda$ )
Median	$5.9683 \times 10^{-1}$	$5.9650 \times 10^{-1}$
$\sigma$	7.6221	7.6219



**Fig. 1** Comparison of aberration magnitude distributions between the real (base) sample and the synthetic sample for selected aberrations: (a)  $Z_4$  astigmatism  $Y$ , (b)  $Z_{14}$  tetrafoil  $Y$ , and (c)  $Z_{24}$  secondary tetrafoil  $X$ . Each panel shows the histogram of synthetic samples (blue bars) and the dotted line from the base sample, illustrating the close statistical alignment between both datasets. The  $x$ -axis represents the magnitude of each Zernike coefficient in units of the wavelength ( $\lambda$ ), whereas the  $y$ -axis indicates the frequency of occurrence. The median ( $M_{syn}$ ) and standard deviation ( $\sigma_{syn}$ ) of the synthetic samples are displayed within each panel.

These tables comprehensively describe the statistical distribution of the input space for the reconstruction models, reinforcing the validity of the dataset for the performance evaluation conducted in this study.

### 3.5 Model Configuration and Training

This section describes the configuration parameters used in the proposed models and, where applicable, the training procedures. As the three approaches for wavefront aberration calibration differ in their implementation, a detailed description of each is provided. To ensure robust evaluation and implementation, the models were trained and tested in a high-performance computational environment optimized for intensive machine learning tasks.

The training and evaluation processes utilized a server with the following technical specifications: two 32-core Intel Xeon Gold 5218 processors at 2.30 GHz, 512 GB of RAM, and 8 NVIDIA TITAN RTX graphics processing units (GPUs), each with 24 GB of dedicated memory. The server runs under Ubuntu 22.04.4 LTS operating system and supports CUDA version 12.4. This computational environment allowed for intensive training, Bayesian optimization, and efficient handling of high-dimensional datasets. Input and output data were managed using storage units with capacities of up to 3.6 TB, enabling efficient handling of large data volumes during experiments.

#### 3.5.1 Coefficient adjustment with least squares

The LSQ approach directly computes the solution  $A$  by solving the system of equations using the pseudoinverse of the  $Z$  matrix, minimizing the quadratic error in wavefront reconstruction. As described in Sec. 2.2, this method does not require a training process, relying solely on the properties of the  $Z$  and  $W$  matrices derived from the simulations in Sec. 3.3.

#### 3.5.2 Training with gradient descent: single sample case

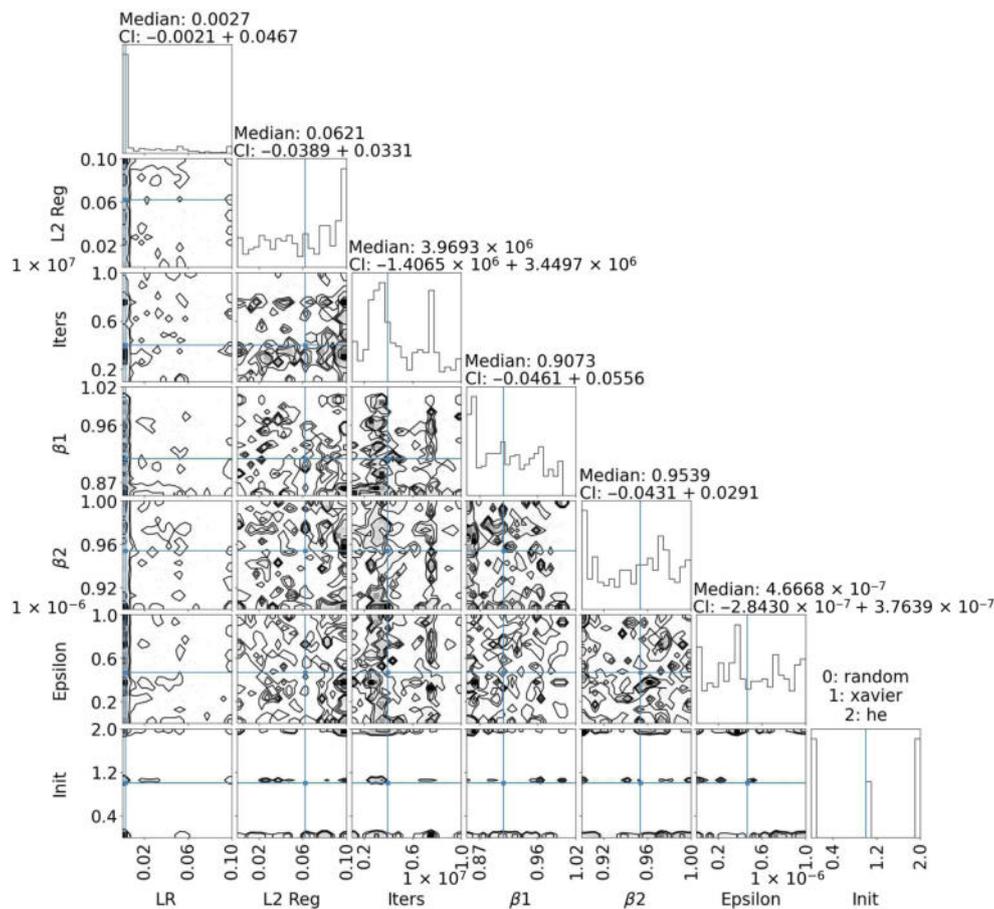
Bayesian optimization was applied to explore the hyperparameter space for the gradient descent model, ensuring optimal configurations for stability and precision. The search included the hyperparameters and their ranges, as listed in Table 3.

The optimization involved 500 function calls ( $n\_calls$ ), generating a corner plot that highlights the correlation between hyperparameters and their influence on the objective function defined as MAE between the predicted and target values of  $W$  (Fig. 2). This visualization was instrumental in identifying critical regions in the hyperparameter space, guiding the selection of optimal values.

By analyzing the corner plot, hyperparameters with strong correlations, such as  $\beta_1$  and L2 regularization, were selected based on their confidence intervals and impact on the objective function. Conversely, parameters such as  $\beta_2$  and  $\epsilon$ , which exhibited minimal influence, were set to their median values. The learning rate required additional refinement through grid search,

**Table 3** Hyperparameter search space for Bayesian optimization in the gradient descent model.

Hyperparameter	Range
Learning rate ( $\alpha$ )	$[1 \times 10^{-15}, 1 \times 10^{-1}]$
L2 regularization	$[1 \times 10^{-8}, 1 \times 10^{-1}]$
Number of iterations	$[1 \times 10^6, 1 \times 10^7]$
$\beta_1$ (Adam optimizer)	[0.85, 0.99]
$\beta_2$ (Adam optimizer)	[0.90, 0.99]
$\epsilon$ (Adam optimizer)	$[1 \times 10^{-10}, 1 \times 10^{-6}]$
Initialization method	{Random, Xavier, He}



**Fig. 2** Bayesian optimization results for the hyperparameters of the gradient descent model. The figure displays the marginal distributions (on the diagonal) and pairwise correlations (off-diagonal) of the hyperparameters explored during the optimization process. The hyperparameters and their corresponding search ranges are defined in Table 3. The objective of the optimization was to minimize the training loss, defined as the MAE between predicted and target values of  $W$ , expressed in units of wavelength ( $\lambda$ ). The figure highlights regions in the hyperparameter space with a higher density of evaluations, which are associated with better performance in minimizing the loss function.

ultimately selecting  $10^{-6}$  as the optimal value for achieving stable convergence. The final selected hyperparameters are summarized in Table 4.

As described in Sec. 2.3, the model training incorporated early stopping to terminate the process when improvements in the loss function ceased, ensuring computational efficiency by avoiding unnecessary iterations. The overall loss trend during training is shown in Fig. 3, illustrating rapid convergence and stabilization.

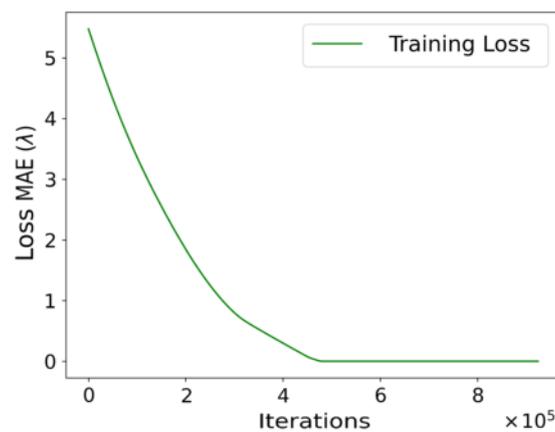
The training process completed 969,478 iterations before the early stopping activation, resulting in a runtime of 1 min and 40 s. This combination of optimized hyperparameters and adaptive stopping mechanisms ensured high accuracy while maintaining computational efficiency.

### 3.5.3 Training with gradient descent: multiple samples case

To validate the model's performance on a larger dataset, training was performed using 3000 samples ( $Z$  and  $W$  matrices). Bayesian optimization was employed using the same search ranges described in Sec. 3.5.2, as detailed in Table 3. These ranges were applied to explore the hyperparameter space for the extended dataset, identifying critical zones with high density for certain parameters, as summarized in Table 5.

**Table 4** Optimal hyperparameters selected from Bayesian optimization for the gradient descent model.

Hyperparameter	Optimal value
Learning rate ( $\alpha$ )	$1 \times 10^{-6}$
L2 regularization	$6.21 \times 10^{-2}$
Number of iterations	5,780,000
$\beta_1$ (Adam optimizer)	0.9073
$\beta_2$ (Adam optimizer)	0.9539
$\epsilon$ (Adam optimizer)	$4.6668 \times 10^{-7}$
Initialization method	Xavier

**Fig. 3** Evolution of the training loss (MAE) during the optimization process using the gradient descent method. The x-axis indicates the number of iterations (in units of  $10^5$ ), and the y-axis shows the MAE between the predicted and target values of  $W$ , expressed in units of wavelength ( $\lambda$ ). This curve illustrates the progressive reduction in prediction error as the model parameters are updated during training.

Following the results of the Bayesian optimization, a grid search was conducted to refine the hyperparameter selection. This search focused on combinations within the critical zones identified while keeping the learning rate fixed at  $1 \times 10^{-6}$ . The optimal configuration selected from the grid search is presented in Table 6.

The results of training with gradient descent across multiple samples demonstrated stability and consistent performance throughout the dataset. Specific performance metrics and detailed evaluations will be discussed in Sec. 4.3.

### 3.5.4 Dynamic hypernetwork training

The dynamic hypernetwork model, described in Sec. 2.4, was trained using Bayesian optimization on 3000 samples to determine the optimal hyperparameter configurations. The search space explored during the optimization is summarized in Table 7.

Bayesian optimization involved 500 function calls ( $n\_calls$ ) over approximately 49 hours, with the results highlighting correlations and distributions of critical hyperparameters, as shown in Fig. 4. The optimal configuration selected after this process is summarized in Table 8.

The model was trained for 200 epochs using early stopping to terminate the process once the validation loss ceased to improve. Similar to the Bayesian optimization process, the training utilized the same 3000 samples. The total training time was 1 min and 45 s.

The evolution of the training and validation metrics is shown in Fig. 5. The smooth convergence of the loss over epochs demonstrates the model's stability during training, with no

**Table 5** Critical zones identified by Bayesian optimization for multiple samples.

Hyperparameter	Critical values
L2 regularization	0.01, 0.09
$\beta_1$ (Adam optimizer)	0.88, 0.98
$\beta_2$ (Adam optimizer)	0.91, 0.99
Learning rate ( $\alpha$ )	$1 \times 10^{-6}$

**Table 6** Optimal hyperparameters selected after grid search.

Hyperparameter	Optimal value
L2 regularization	0.09
$\beta_1$ (Adam optimizer)	0.98
$\beta_2$ (Adam optimizer)	0.99
Learning rate ( $\alpha$ )	$1 \times 10^{-6}$

**Table 7** Search space for hyperparameters in dynamic hypernetwork training.

Hyperparameter	Range/options
Learning rate	$[1 \times 10^{-5}, 1 \times 10^{-2}]$
L2 regularization	$[1 \times 10^{-6}, 1 \times 10^{-1}]$
Dropout rate	[0.1, 0.5]
Batch size	[16, 64]
Number of additional layers	[2, 6]
Neurons in mother network ( $U_m$ )	[200, 1200]
Neurons in auxiliary network ( $U_a$ )	[100,600]
Activation function	{ReLU, Leaky ReLU, Tanh}
Kernel initialization	{Glorot Normal, He Normal}

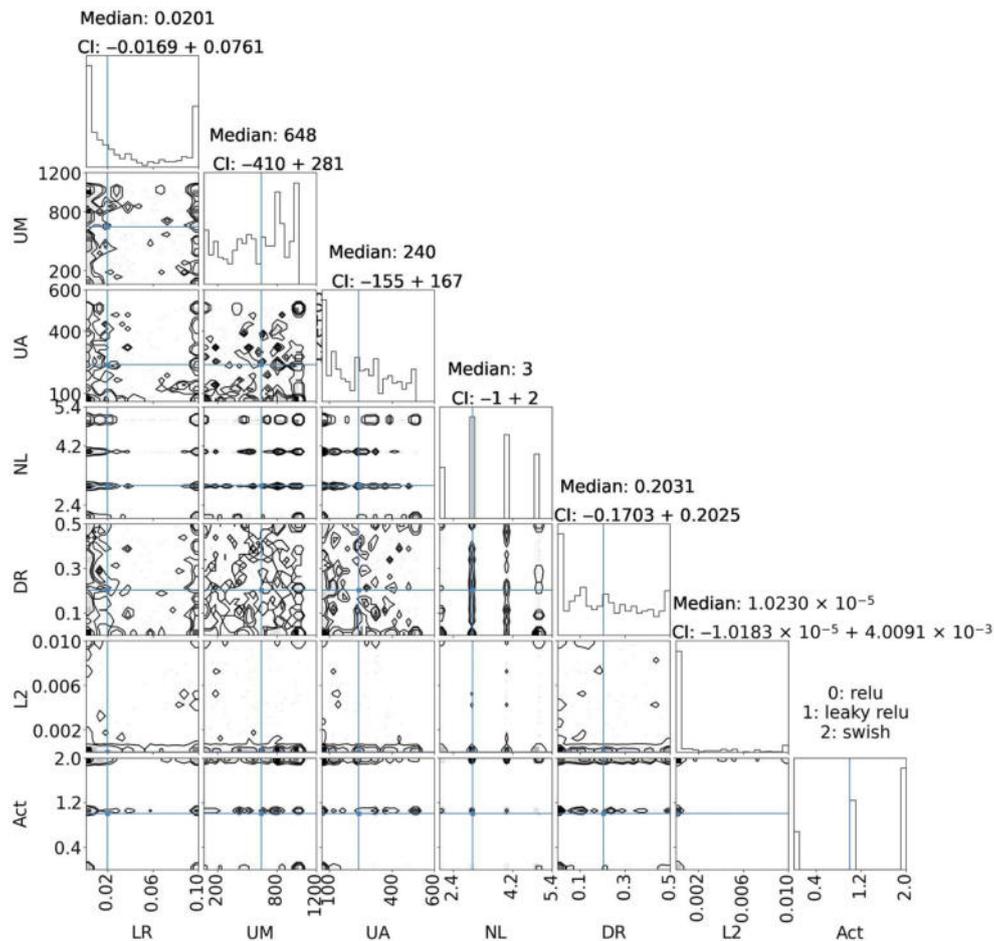
evidence of overfitting. These results confirm the dynamic hypernetwork's robustness and effectiveness in learning the target mappings.

Figure 6(a) shows the MAE on the training set, whereas Fig. 6(b) shows the MAE on the validation set. MAE was selected for its interpretability and effectiveness in tracking the model's convergence. Although other metrics were computed, MAE provides the most direct insight during training. The lack of overfitting and the stable performance across epochs demonstrate the robustness of the dynamic hypernetwork in learning the target mappings.

The methods described in this section are applied to simulated datasets designed to replicate realistic Shack–Hartmann sensor measurements. The results in Sec. 4 demonstrate the effectiveness of each method, offering a comparative analysis of their performance under various conditions, and highlighting their practical implications in wavefront fitting.

## 4 Performance Evaluation

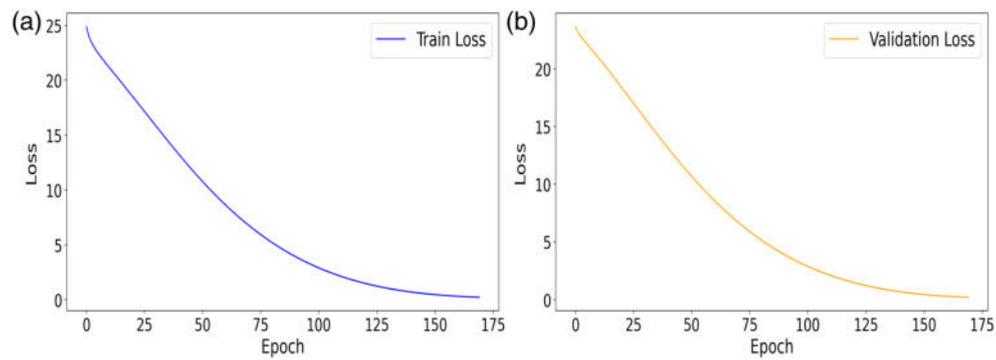
This section presents the performance of the three methods (least squares, gradient descent, and dynamic hypernetwork) through a series of metrics and visual comparisons. The results



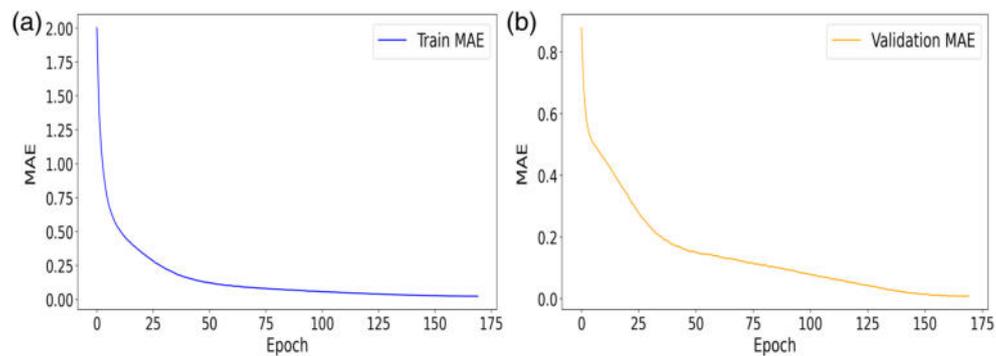
**Fig. 4** Bayesian optimization results for the hyperparameters of the dynamic hypernetwork model. The figure displays marginal distributions (on the diagonal) and pairwise correlations (off-diagonal) of the hyperparameters explored during the optimization process. The hyperparameters and their corresponding search ranges are defined in Table 7. The objective of the optimization was to minimize the training loss, defined as the MAE between predicted and target values of  $W$ , expressed in units of wavelength ( $\lambda$ ). This corner plot reveals interdependencies among hyperparameters and helps identify the optimal configuration summarized in Table 8.

**Table 8** Optimal hyperparameters selected for dynamic hypernetwork training.

Hyperparameter	Optimal value
Learning rate	$1 \times 10^{-5}$
L2 regularization	$1.0230 \times 10^{-2}$
Dropout rate	0.2031
Batch size	32
Number of additional layers	3
Neurons in mother network ( $U_m$ )	648
Neurons in auxiliary network ( $U_a$ )	240
Activation function	Leaky ReLU
Kernel initialization	Glorot normal



**Fig. 5** Evolution of the training and validation loss for the dynamic hypernetwork model. Panel (a) shows the training loss, and panel (b) shows the validation loss, both measured using the MAE between predicted and target values. The model was trained on normalized input and output data, making the loss values dimensionless. These curves illustrate how the model generalizes to unseen data and help assess the convergence behavior during training.



**Fig. 6** Evolution of the MAE during training and validation of the dynamic hypernetwork model. Panel (a) shows the MAE on the training set, and panel (b) shows the MAE on the validation set, both plotted over training epochs. The MAE is computed between predicted and target values. As the model operates on normalized input and output data, the error is dimensionless. These curves provide insight into the model's learning behavior and generalization performance.

are organized to highlight key aspects of the methods' behavior and their alignment with the data.

#### 4.1 Results Overview

The error metrics for the three methods are summarized in Table 9. These metrics provide a quantitative evaluation of the performance of each method. The table includes values for  $R^2$ , MAE, MSE, and the RMS values for both  $W$  and  $A$ , offering a comprehensive view of the quality of predictions and the coefficients used. In addition, the table reports the computational time required for each method, allowing for a comparison of both accuracy and efficiency.

The dynamic hypernetwork offers the fastest computation time and excels in predictive capability, enabling it to anticipate and correct wavefront changes more effectively in real-time applications. By contrast, LSQ, despite its computational efficiency, lacks adaptability and does not incorporate previous wavefront information for improved performance.

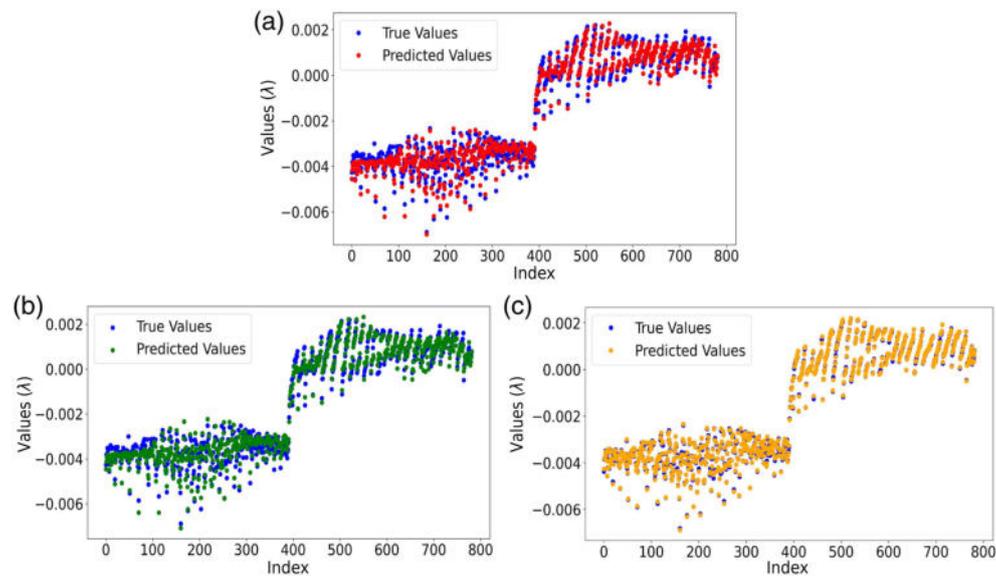
Figure 7 complements these metrics by providing a visual comparison of the predicted and true values of  $W$  across the dataset. The  $x$ -axis represents the index of each data point, whereas the  $y$ -axis shows the corresponding values. This plot allows for a direct observation of the alignment between predictions and actual data for each method.

#### 4.2 Analysis in the Presence of Outliers

To evaluate the robustness of each method, a region containing outliers at the edge of the dataset was analyzed. Figure 8 shows the prediction behavior within this region for (a) LSQ, (b) gradient

**Table 9** Error metrics for the adjustment of  $W$  using the three methods.

Metric	LSQ	Gradient descent	Dynamic hypernetwork
Computation time	0.4250 ms	52.2384 s	0.0273 ms
$R^2$ (%)	98.62	98.57	99.95
MAE ( $\lambda$ )	$2.2135 \times 10^{-4}$	$2.1787 \times 10^{-4}$	$4.2706 \times 10^{-5}$
MSE ( $\lambda^2$ )	$7.5606 \times 10^{-8}$	$7.8606 \times 10^{-8}$	$2.8795 \times 10^{-9}$
RMS of $W$ ( $\lambda$ )	$2.7497 \times 10^{-4}$	$2.8037 \times 10^{-4}$	$5.3661 \times 10^{-5}$
RMS of $A$ (dimensionless)	$1.8183 \times 10^{-3}$	$3.0309 \times 10^{-3}$	$1.6824 \times 10^{-1}$



**Fig. 7** Comparison of predicted and true values of  $W$  for the base sample using three different methods: (a) LSQ, (b) gradient descent, and (c) dynamic hypernetwork. The  $x$ -axis corresponds to the index of each data point across the Shack–Hartmann sensor grid, and the  $y$ -axis shows the displacement values of  $W$ , expressed in units of wavelength ( $\lambda$ ). This comparison illustrates the accuracy of each method in reproducing the original wavefront shape for this specific case.

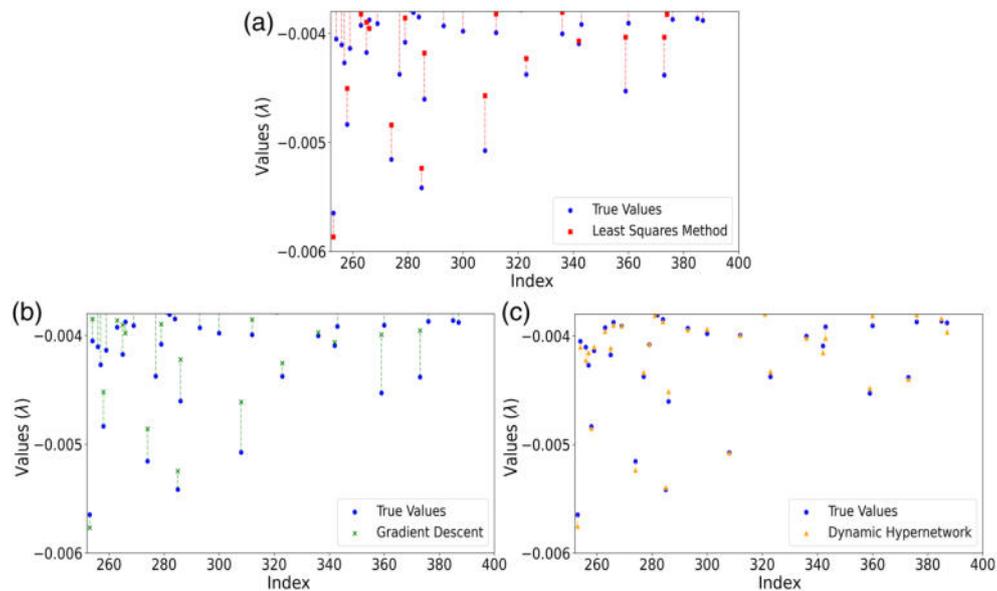
descent, and (c) dynamic hypernetwork. The  $x$ -axis represents the index of the data points, whereas the  $y$ -axis indicates their corresponding values of  $W$ . This localized analysis provides insights into how the methods respond to deviations in the data.

Finally, Fig. 9 illustrates the residual distribution for the three methods. The residuals, calculated as the difference between true and predicted values, are plotted against the index of each prediction. This figure provides an additional perspective on the precision of the methods by showing how closely predictions align with true values across the dataset.

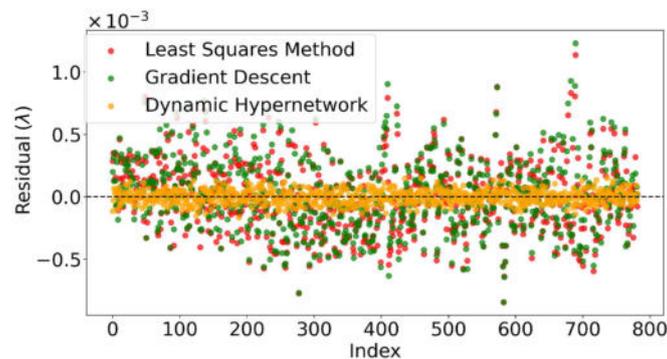
### 4.3 Testing Dataset Analysis

This section presents the evaluation of the three methods (LSQ, gradient descent, and dynamic hypernetwork) applied to a testing dataset of 500 samples. The results focus on key error metrics, as shown in Table 10.

Table 10 provides a comparative summary of the performance of the three methods, averaged over the 500 samples in the testing dataset. The metrics include computation time, MAE, MSE, RMS of  $W$ , and RMS of  $A$ . These metrics evaluate the methods' predictive performance and their ability to model the coefficients.



**Fig. 8** Zoomed-in comparison of predicted and true values of  $W$  in a region of interest from the base sample, focusing on an area with outliers. The three methods shown are as follows: (a) LSQ, (b) gradient descent, and (c) dynamic hypernetwork. The x-axis shows the index of the affected data points across the Shack–Hartmann sensor, and the y-axis represents the displacement values of  $W$  in units of wavelength ( $\lambda$ ). This close-up highlights each method's robustness in the presence of data anomalies.



**Fig. 9** Residual distribution for the three methods: LSQ, gradient descent, and dynamic hypernetwork. Residuals are computed as the difference between the predicted and true values of  $W$  for each point in the base sample. The x-axis denotes the index of each measurement point across the Shack–Hartmann sensor, and the y-axis represents the residuals in units of wavelength ( $\lambda$ ). This analysis provides insight into the precision of each method and highlights potential systematic errors or anomalies.

**Table 10** Error metrics for the three methods on the testing dataset.

Metric	LSQ	Gradient descent	Dynamic hypernetwork
Computation time	211.7550 ms	9.1315 h	188.7908 ms
MAE ( $\lambda$ )	$2.1909 \times 10^{-4}$	$2.1586 \times 10^{-4}$	$2.9559 \times 10^{-5}$
MSE ( $\lambda^2$ )	$7.3810 \times 10^{-8}$	$7.6464 \times 10^{-8}$	$1.4613 \times 10^{-9}$
RMS of $W$ ( $\lambda$ )	$2.7167 \times 10^{-4}$	$2.7651 \times 10^{-4}$	$2.9559 \times 10^{-5}$
RMS of $A$ (dimensionless)	$1.8199 \times 10^{-3}$	$1.8224 \times 10^{-3}$	$1.6800 \times 10^{-1}$

The dataset consisted of 500 samples, and the evaluation focused on how accurately each method predicted the  $W$  values using the provided  $Z$  matrices. The general analysis reveals that dynamic hypernetwork excels in terms of precision and stability.

## 5 Conclusions

This study evaluated three methods—LSQ, gradient descent, and dynamic hypernetwork—for wavefront fitting using Shack–Hartmann sensor data. The methods were assessed based on their performance in reconstructing  $W$  values, both for a single sample and a dataset of 500 testing samples while also analyzing their behavior in the presence of outliers.

The LSQ method achieved stable performance with an  $R^2$  of 98.62% and an MAE of  $2.2135 \times 10^{-4} \lambda$  on a single sample, as reflected in Table 9. However, its reliance on the pseudoinverse limits its robustness in cases involving missing or overlapping points. Gradient descent provided similar performance, achieving an  $R^2$  of 98.57% and a slightly lower MAE of  $2.1787 \times 10^{-4} \lambda$  while eliminating the need for matrix inversion. This makes it a computationally simpler alternative, particularly for resource-constrained applications.

The dynamic hypernetwork demonstrated superior results across all scenarios, achieving an  $R^2$  of 99.95% and an MAE of  $4.2706 \times 10^{-5} \lambda$  on the single sample. The evaluation of the testing dataset further confirmed its robustness, achieving an MAE of  $2.9559 \times 10^{-5} \lambda$ , which outperformed the other methods as shown in Table 10. In addition, the dynamic hypernetwork exhibited superior performance across all other metrics, including MSE (in  $\lambda^2$ ) and RMS of  $W$  (in  $\lambda$ ).

One key distinction of the dynamic hypernetwork lies in the RMS values of  $A$  ( $1.6800 \times 10^{-1}$ ), which are notably larger compared with the other methods. However, RMS of  $A$  is not a direct measure of predictive accuracy; it reflects the model's capacity to produce coefficients that enhance aberration quantification. This capacity translates into a significant impact on wavefront error characterization, complementing the superior accuracy indicated by metrics such as MAE, MSE, and RMS of  $W$ .

The residual analysis highlighted the dynamic hypernetwork's precision, with residuals concentrated closely around 0, indicating its consistent alignment with true values across the testing dataset. This robustness was also evident in its ability to manage outliers effectively, as shown in localized prediction evaluations.

Beyond its predictive accuracy, the dynamic hypernetwork also demonstrated improved computational efficiency. In the single-sample case (see Table 9), it achieved a computation time of 0.0273 ms, significantly lower than the 0.4250 ms required by LSQ. Similarly, for the test set (see Table 10), the dynamic hypernetwork required 188.7908 ms, compared with 211.7550 ms for LSQ. This result reinforces the advantage of dynamic hypernetworks as they provide both enhanced accuracy and reduced computation time, making them suitable for real-time applications where rapid wavefront fitting is essential.

In conclusion, the LSQ and gradient descent methods provide reliable options for wavefront fitting, with gradient descent offering additional flexibility in handling irregular data without matrix inversion. However, the dynamic hypernetwork emerges as the most effective method, excelling in precision, robustness, and computational efficiency. These attributes make the dynamic hypernetwork a promising tool for high-precision optical systems, particularly in adaptive optics, where stable, adaptive, and predictive wavefront fitting is essential. Future work will focus on refining the efficiency of these AI-based approaches to further enhance their applicability in real-time astronomical imaging and laser communication systems.

## 6 Appendix: Detailed Aberration Statistics

This appendix presents detailed statistical tables of the aberration magnitudes and their  $x$ - and  $y$ -components for both the base and synthetic datasets, providing a comprehensive numerical overview for wavefront reconstruction evaluation. Following Noll's standard numbering scheme, Tables 11–14 list the Zernike polynomials from coefficient 1 to 36, facilitating cross-referencing and ensuring consistency in the analysis.

**Table 11** List of Zernike polynomials (Noll's nomenclature) from coefficient 1 to 36.

Coefficient	Name	Polynomial
1	Tilt $x$	$\sqrt{4p} \cos(\varphi)$
2	Tilt $y$	$\sqrt{4p} \sin(\varphi)$
3	Power (focus)	$\sqrt{3}(2p^2 - 1)$
4	Astigmatism $y$	$\sqrt{6}p^2 \sin(2\varphi)$
5	Astigmatism $x$	$\sqrt{6}p^2 \cos(2\varphi)$
6	Coma $y$	$\sqrt{8}(3p^3 - 2p) \sin(\varphi)$
7	Coma $x$	$\sqrt{8}(3p^3 - 2p) \cos(\varphi)$
8	Trefoil $y$	$\sqrt{8}p^3 \sin(3\varphi)$
9	Trefoil $x$	$\sqrt{8}p^3 \cos(3\varphi)$
10	Primary spherical	$\sqrt{5}(6p^4 - 6p^2 + 1)$
11	Secondary astigmatism $x$	$\sqrt{10}(4p^4 - 3p^2) \cos(2\varphi)$
12	Secondary astigmatism $y$	$\sqrt{10}(4p^4 - 3p^2) \sin(2\varphi)$
13	Tetrafoil $x$	$\sqrt{10}p^4 \cos(4\varphi)$
14	Tetrafoil $y$	$\sqrt{10}p^4 \sin(4\varphi)$
15	Secondary coma $x$	$\sqrt{12}(10p^5 - 12p^3 + 3p) \cos(\varphi)$
16	Secondary coma $y$	$\sqrt{12}(10p^5 - 12p^3 + 3p) \sin(\varphi)$
17	Secondary trefoil $x$	$\sqrt{12}(5p^5 - 4p^3) \cos(3\varphi)$
18	Secondary trefoil $y$	$\sqrt{12}(5p^5 - 4p^3) \sin(3\varphi)$
19	Pentafoil $x$	$\sqrt{12}p^5 \cos(5\varphi)$
20	Pentafoil $y$	$\sqrt{12}p^5 \sin(5\varphi)$
21	Secondary spherical	$\sqrt{7}(20p^6 - 30p^4 + 12p^2 - 1)$
22	Tertiary astigmatism $y$	$\sqrt{14}(15p^6 - 20p^4 + 6p^2) \sin(2\varphi)$
23	Tertiary astigmatism $x$	$\sqrt{14}(15p^6 - 20p^4 + 6p^2) \cos(2\varphi)$
24	Secondary tetrafoil $y$	$\sqrt{14}(6p^6 - 5p^4) \sin(4\varphi)$
25	Secondary tetrafoil $x$	$\sqrt{14}(6p^6 - 5p^4) \cos(4\varphi)$
26	Hexafoil $y$	$\sqrt{14}p^6 \sin(6\varphi)$
27	Hexafoil $x$	$\sqrt{14}p^6 \cos(6\varphi)$
28	Tertiary coma $y$	$\sqrt{16}(35p^7 - 60p^5 + 30p^3 - 4p) \sin(\varphi)$
29	Tertiary coma $x$	$\sqrt{16}(35p^7 - 60p^5 + 30p^3 - 4p) \cos(\varphi)$
30	Tertiary trefoil $y$	$\sqrt{16}(21p^7 - 30p^5 + 10p^3) \sin(3\varphi)$
31	Tertiary trefoil $x$	$\sqrt{16}(21p^7 - 30p^5 + 10p^3) \cos(3\varphi)$
32	Secondary pentafoil $y$	$\sqrt{16}(7p^7 - 6p^5) \sin(5\varphi)$
33	Secondary pentafoil $x$	$\sqrt{16}(7p^7 - 6p^5) \cos(5\varphi)$
34	Septafoil $y$	$\sqrt{16}p^7 \sin(7\varphi)$
35	Septafoil $x$	$\sqrt{16}p^7 \cos(7\varphi)$
36	Tertiary spherical	$\sqrt{9}(70p^8 - 140p^6 + 90p^4 - 20p^2 + 1)$

**Table 12** Statistical summary of aberration coefficients in the base and synthetic samples.

Coefficient	$M_{\text{base}}(\lambda)$	$\sigma_{\text{base}}(\lambda)$	$M_{\text{syn}}(\lambda)$	$\sigma_{\text{syn}}(\lambda)$	$M_{\text{diff}}(\lambda)$	$\sigma_{\text{diff}}(\lambda)$
1	2.0000	$1.2480 \times 10^{-11}$	2.0000	$1.2580 \times 10^{-11}$	0.0000	$1.0000 \times 10^{-13}$
2	2.0000	$1.3100 \times 10^{-11}$	2.0000	$1.3200 \times 10^{-11}$	0.0000	$1.0000 \times 10^{-13}$
3	$9.6494 \times 10^{-2}$	4.9910	$9.6853 \times 10^{-2}$	4.9910	$3.5929 \times 10^{-4}$	$1.1530 \times 10^{-6}$
4	$6.8231 \times 10^{-2}$	3.5292	$6.8485 \times 10^{-2}$	3.5292	$2.5419 \times 10^{-4}$	$8.1500 \times 10^{-7}$
5	$6.8231 \times 10^{-2}$	3.5292	$6.8485 \times 10^{-2}$	3.5292	$2.5419 \times 10^{-4}$	$8.1500 \times 10^{-7}$
6	2.0173	5.7903	2.0171	5.7903	$1.3668 \times 10^{-4}$	$8.8240 \times 10^{-6}$
7	1.9510	5.8908	1.9515	5.8908	$4.5151 \times 10^{-4}$	$8.5000 \times 10^{-8}$
8	$9.9596 \times 10^{-2}$	4.7251	$9.9552 \times 10^{-2}$	4.7251	$4.3874 \times 10^{-5}$	$2.5703 \times 10^{-6}$
9	$9.9596 \times 10^{-2}$	4.7251	$9.9552 \times 10^{-2}$	4.7251	$4.3873 \times 10^{-5}$	$2.5702 \times 10^{-6}$
10	$3.0680 \times 10^{-2}$	8.1223	$3.0823 \times 10^{-2}$	8.1223	$1.4361 \times 10^{-4}$	$2.3049 \times 10^{-5}$
11	$1.9339 \times 10^{-1}$	8.1908	$1.9123 \times 10^{-1}$	8.1907	$2.1634 \times 10^{-3}$	$1.3138 \times 10^{-5}$
12	$2.6034 \times 10^{-2}$	7.9987	$2.6025 \times 10^{-2}$	7.9987	$9.1080 \times 10^{-6}$	$1.8121 \times 10^{-5}$
13	$1.4888 \times 10^{-1}$	5.7051	$1.4861 \times 10^{-1}$	5.7051	$2.6255 \times 10^{-4}$	$5.7260 \times 10^{-6}$
14	$1.4888 \times 10^{-1}$	5.7051	$1.4861 \times 10^{-1}$	5.7051	$2.6255 \times 10^{-4}$	$5.7260 \times 10^{-6}$
15	1.7874	9.2906	1.7864	9.2906	$9.7942 \times 10^{-4}$	$1.5968 \times 10^{-5}$
16	1.8474	9.0071	1.8478	9.0070	$3.7883 \times 10^{-4}$	$8.0451 \times 10^{-5}$
17	1.1907	9.2974	1.1907	9.2973	$4.4400 \times 10^{-5}$	$1.7132 \times 10^{-5}$
18	1.4273	9.3036	1.4312	9.3036	$3.9690 \times 10^{-3}$	$4.0016 \times 10^{-5}$
19	$4.5661 \times 10^{-2}$	6.5170	$4.6136 \times 10^{-2}$	6.5170	$4.7511 \times 10^{-4}$	$1.0480 \times 10^{-5}$
20	$4.5661 \times 10^{-2}$	6.5170	$4.6136 \times 10^{-2}$	6.5170	$4.7511 \times 10^{-4}$	$1.0480 \times 10^{-5}$
21	$7.3154 \times 10^{-2}$	9.3796	$7.3130 \times 10^{-2}$	9.3796	$2.3593 \times 10^{-5}$	$6.1168 \times 10^{-5}$
22	$1.8281 \times 10^{-1}$	9.6668	$1.8374 \times 10^{-1}$	9.6667	$9.3587 \times 10^{-4}$	$9.0419 \times 10^{-5}$
23	$1.5559 \times 10^{-1}$	$1.0060 \times 10^1$	$1.5476 \times 10^{-1}$	$1.0060 \times 10^1$	$8.3027 \times 10^{-4}$	$2.9070 \times 10^{-5}$
24	$2.9611 \times 10^{-1}$	$1.0186 \times 10^1$	$2.9216 \times 10^{-1}$	$1.0186 \times 10^1$	$3.9539 \times 10^{-3}$	$3.4910 \times 10^{-5}$
25	$3.0912 \times 10^{-1}$	$1.0313 \times 10^1$	$3.0943 \times 10^{-1}$	$1.0313 \times 10^1$	$3.1155 \times 10^{-4}$	$4.5900 \times 10^{-5}$
26	$3.5805 \times 10^{-2}$	7.1919	$3.6253 \times 10^{-2}$	7.1919	$4.4757 \times 10^{-4}$	$1.6426 \times 10^{-5}$
27	$3.5805 \times 10^{-2}$	7.1919	$3.6253 \times 10^{-2}$	7.1919	$4.4757 \times 10^{-4}$	$1.6426 \times 10^{-5}$
28	$5.3593 \times 10^{-1}$	$1.0489 \times 10^1$	$5.3680 \times 10^{-1}$	$1.0489 \times 10^1$	$8.7275 \times 10^{-4}$	$1.0000 \times 10^{-4}$
29	$5.3897 \times 10^{-1}$	$1.0459 \times 10^1$	$5.3548 \times 10^{-1}$	$1.0459 \times 10^1$	$3.4919 \times 10^{-3}$	$1.6000 \times 10^{-7}$
30	1.4786	$1.0867 \times 10^1$	1.4707	$1.0867 \times 10^1$	$7.9120 \times 10^{-3}$	$7.5150 \times 10^{-5}$
31	1.6815	$1.0719 \times 10^1$	1.6811	$1.0719 \times 10^1$	$4.3181 \times 10^{-4}$	$4.0080 \times 10^{-5}$
32	$2.9234 \times 10^{-1}$	$1.1154 \times 10^1$	$2.9226 \times 10^{-1}$	$1.1154 \times 10^1$	$8.1016 \times 10^{-5}$	$5.7540 \times 10^{-5}$
33	$2.8677 \times 10^{-1}$	$1.0889 \times 10^1$	$2.8495 \times 10^{-1}$	$1.0889 \times 10^1$	$1.8207 \times 10^{-3}$	$4.1580 \times 10^{-5}$
34	$3.4269 \times 10^{-2}$	7.7459	$3.4306 \times 10^{-2}$	7.7459	$3.7046 \times 10^{-5}$	$2.3606 \times 10^{-5}$
35	$3.4269 \times 10^{-2}$	7.7459	$3.4306 \times 10^{-2}$	7.7459	$3.7041 \times 10^{-5}$	$2.3606 \times 10^{-5}$
36	$2.2239 \times 10^{-1}$	$1.1502 \times 10^1$	$2.2334 \times 10^{-1}$	$1.1502 \times 10^1$	$9.5129 \times 10^{-4}$	$2.1370 \times 10^{-5}$

**Table 13** Statistical summary of X and Y aberration components in the base sample.

Coefficient	$M_{\text{base},X}(\lambda)$	$\sigma_{\text{base},X}(\lambda)$	$M_{\text{base},Y}(\lambda)$	$\sigma_{\text{base},Y}(\lambda)$
1	0.0000	$8.3552 \times 10^{-12}$	2.0000	$9.2704 \times 10^{-12}$
2	2.0000	$9.9600 \times 10^{-12}$	0.0000	$8.4682 \times 10^{-12}$
3	$-9.5059 \times 10^{-2}$	3.5122	$1.6580 \times 10^{-2}$	3.5461
4	$1.1724 \times 10^{-2}$	2.5074	$-6.7217 \times 10^{-2}$	2.4835
5	$6.7217 \times 10^{-2}$	2.4835	$1.1724 \times 10^{-2}$	2.5074
6	$2.0154 \times 10^0$	4.7464	$8.6427 \times 10^{-2}$	3.3167
7	$8.6427 \times 10^{-2}$	3.3167	$1.9491 \times 10^0$	4.8684
8	$-4.9495 \times 10^{-2}$	3.3655	$8.6427 \times 10^{-2}$	3.3167
9	$-8.6427 \times 10^{-2}$	5.6675	$-4.9495 \times 10^{-2}$	3.3655
10	$2.9352 \times 10^{-2}$	5.6805	$-8.9299 \times 10^{-3}$	5.8182
11	$1.3606 \times 10^{-1}$	5.6266	$-1.3743 \times 10^{-1}$	5.9009
12	$-9.7521 \times 10^{-4}$	4.0291	$-2.6015 \times 10^{-2}$	5.6851
13	$1.4887 \times 10^{-1}$	4.0391	$8.5169 \times 10^{-4}$	4.0391
14	$8.5169 \times 10^{-4}$	4.6087	$-1.4887 \times 10^{-1}$	4.0291
15	$1.8338 \times 10^{-2}$	6.4110	$-1.7873 \times 10^0$	8.0670
16	$-1.8473 \times 10^0$	6.5613	$1.8338 \times 10^{-2}$	4.6087
17	$-2.8345 \times 10^{-2}$	4.5895	$-1.1904 \times 10^0$	6.7335
18	$1.4272 \times 10^0$	4.6268	$-1.5751 \times 10^{-2}$	6.5960
19	$3.5556 \times 10^{-2}$	6.5772	$-2.8648 \times 10^{-2}$	4.6268
20	$-2.8648 \times 10^{-2}$	6.7687	$-3.5556 \times 10^{-2}$	4.5895
21	$4.8305 \times 10^{-2}$	7.0124	$-5.4937 \times 10^{-2}$	6.6871
22	$-4.2951 \times 10^{-2}$	7.0695	$1.7769 \times 10^{-1}$	6.9015
23	$1.5372 \times 10^{-1}$	7.2236	$2.4037 \times 10^{-2}$	7.2127
24	$1.1842 \times 10^{-2}$	5.0513	$2.9587 \times 10^{-1}$	7.3336
25	$-3.0491 \times 10^{-1}$	5.1194	$5.0850 \times 10^{-2}$	7.3603
26	$-1.4733 \times 10^{-3}$	9.1247	$-3.5775 \times 10^{-2}$	5.1194
27	$3.5775 \times 10^{-2}$	5.1734	$-1.4733 \times 10^{-3}$	5.0513
28	$-5.3410 \times 10^{-1}$	7.7177	$4.4254 \times 10^{-2}$	5.1734
29	$4.4254 \times 10^{-2}$	7.6937	$-5.3715 \times 10^{-1}$	9.0904
30	$-1.4782 \times 10^0$	7.8611	$3.6075 \times 10^{-2}$	7.6499
31	$-3.9753 \times 10^{-2}$	7.5189	$1.6810 \times 10^0$	7.6937
32	$-2.9166 \times 10^{-1}$	5.5475	$1.9900 \times 10^{-2}$	7.9128
33	$3.6213 \times 10^{-2}$	5.4059	$-2.8447 \times 10^{-1}$	7.8765
34	$3.3067 \times 10^{-2}$	8.1908	$8.9973 \times 10^{-3}$	5.4059
35	$-8.9973 \times 10^{-3}$	8.0747	$3.3067 \times 10^{-2}$	5.5475
36	$-2.2229 \times 10^{-1}$	7.8398	$-6.6376 \times 10^{-3}$	8.0747

**Table 14** Statistical summary of  $X$  and  $Y$  aberration components in the synthetic samples.

Coefficient	$M_{\text{syn},X}(\lambda)$	$\sigma_{\text{syn},X}(\lambda)$	$M_{\text{syn},Y}(\lambda)$	$\sigma_{\text{syn},Y}(\lambda)$
1	0.0000	$8.3554 \times 10^{-12}$	2.0000	$9.2711 \times 10^{-12}$
2	2.0000	$9.9610 \times 10^{-12}$	0.0000	$8.4679 \times 10^{-12}$
3	$-9.5421 \times 10^{-2}$	3.5122	$1.6593 \times 10^{-2}$	3.5461
4	$1.1733 \times 10^{-2}$	2.5074	$-6.7473 \times 10^{-2}$	2.4835
5	$6.7473 \times 10^{-2}$	2.4835	$1.1733 \times 10^{-2}$	2.5074
6	$2.0153 \times 10^0$	4.7463	$8.6388 \times 10^{-2}$	3.3167
7	$8.6388 \times 10^{-2}$	3.3167	$1.9496 \times 10^0$	4.8684
8	$-4.9476 \times 10^{-2}$	3.3655	$8.6388 \times 10^{-2}$	3.3167
9	$-8.6388 \times 10^{-2}$	3.3167	$-4.9476 \times 10^{-2}$	3.3655
10	$2.9485 \times 10^{-2}$	5.6675	$-8.9845 \times 10^{-3}$	5.8182
11	$1.3421 \times 10^{-1}$	5.6804	$-1.3622 \times 10^{-1}$	5.9009
12	$2.9200 \times 10^{-4}$	5.6266	$-2.6023 \times 10^{-2}$	5.6851
13	$1.4861 \times 10^{-1}$	4.0291	$-9.4900 \times 10^{-4}$	4.0391
14	$-9.4900 \times 10^{-4}$	4.0391	$-1.4861 \times 10^{-1}$	4.0291
15	$1.8847 \times 10^{-2}$	4.6087	$-1.7863 \times 10^0$	8.0670
16	$-1.8477 \times 10^0$	7.7386	$-1.1904 \times 10^0$	6.7335
17	$-2.4613 \times 10^{-2}$	6.4110	$-1.7115 \times 10^{-2}$	6.5959
18	$1.4311 \times 10^0$	6.5612	$-2.8449 \times 10^{-2}$	4.6268
19	$3.6321 \times 10^{-2}$	4.5895	$-3.6321 \times 10^{-2}$	4.5895
20	$-2.8449 \times 10^{-2}$	4.6268	$-5.4517 \times 10^{-2}$	6.6871
21	$4.8744 \times 10^{-2}$	6.5771	$1.7867 \times 10^{-1}$	6.9015
22	$-4.2854 \times 10^{-2}$	6.7686	$1.7921 \times 10^{-2}$	7.2128
23	$1.5372 \times 10^{-1}$	7.0123	$2.9191 \times 10^{-1}$	7.3336
24	$1.1899 \times 10^{-2}$	7.0694	$5.1405 \times 10^{-2}$	7.3603
25	$-3.0513 \times 10^{-1}$	7.2236	$-3.6247 \times 10^{-2}$	5.1194
26	$-6.4450 \times 10^{-4}$	5.0513	$-6.4450 \times 10^{-4}$	5.0513
27	$3.6247 \times 10^{-2}$	5.1194	$4.5198 \times 10^{-2}$	5.1733
28	$-5.3489 \times 10^{-1}$	9.1246	$-5.3357 \times 10^{-1}$	9.0904
29	$4.5198 \times 10^{-2}$	5.1733	$4.3189 \times 10^{-2}$	7.6499
30	$-1.4701 \times 10^0$	7.7176	$1.6806 \times 10^0$	7.6937
31	$-4.0455 \times 10^{-2}$	7.4635	$1.8881 \times 10^{-2}$	7.9128
32	$-2.9165 \times 10^{-1}$	7.5189	$-2.8198 \times 10^{-1}$	7.8765
33	$4.1009 \times 10^{-2}$	5.5475	$9.2385 \times 10^{-3}$	5.4059
34	$3.3039 \times 10^{-2}$	5.4059	$3.3039 \times 10^{-2}$	5.5475
35	$-9.2385 \times 10^{-3}$	8.1908	$-5.8090 \times 10^{-3}$	8.0747
36	$-2.2327 \times 10^{-1}$	8.1908	$-5.8090 \times 10^{-3}$	8.0747

## Disclosures

The authors declare no conflicts of interest.

## Code and Data Availability

The data and code presented in this paper are available to download in this link provided by Ref. 24.

## Acknowledgments

We thank the anonymous reviewers for their valuable feedback, which allowed us to improve the clarity and accuracy of this paper. Juan J. Tapia acknowledges support from project SIP20241558.

## References

1. J. Rayces, "Exact relation between wave aberration and ray aberration," *Opt. Acta* **11**(2), 85–88 (1964).
2. N. Goel and D. Ganotra, "Sensorless wavefront sensing: techniques, applications, and challenges," *Atmos. Ocean. Opt.* **37**(1), 103–117 (2024).
3. R. V. Shack, "Production and use of a lenticular Hartmann screen," in *Spring Meeting of Optical Society of America*, p. 656 (1971).
4. D. Malacara, *Optical Shop Testing*, 3rd ed., John Wiley & Sons, Hoboken, New Jersey (2007).
5. Z. A. Noel et al., "Robust, regularized wavefront reconstruction of Shack-Hartmann data in discontinuous flow," *Proc. SPIE* **12693**, 1269318 (2023).
6. S. Pérez et al., "Enhancing open-loop wavefront prediction in adaptive optics through 2d-lstm neural network implementation," *Photonics* **11**(3), 240 (2024).
7. Y. Guo et al., "Adaptive optics based on machine learning: a review," *Opto-Electron. Adv.* **5**(7), 200082 (2022).
8. J. Herrmann, "Least-squares wave front errors of minimum norm," *J. Opt. Soc. Am.* **70**, 28–35 (1980).
9. J. Fowler and R. Landman, "Tempestas ex machina: a review of machine learning methods for wavefront control," arXiv:2309.00730 (2023).
10. X. Liu et al., "Wavefront prediction using artificial neural networks for open-loop adaptive optics," *Mon. Not. R. Astron. Soc.* **496**, 456–464 (2020).
11. G. Wetzstein et al., "Deep optics: learning cameras and optical computing systems," in *54th Asilomar Conf. Signals, Syst. Comput.*, IEEE, pp. 1313–1315 (2020).
12. X. Lin et al., "All-optical machine learning using diffractive deep neural networks," *Science* **361**(6406), 1004–1008 (2018).
13. V. Hutterer, R. Ramlau, and I. Shatokhina, "Real-time adaptive optics with pyramid wavefront sensors: part ii. Accurate wavefront reconstruction using iterative methods," *Inverse Probl.* **35**, 045008 (2019).
14. J. Duvall et al., "Design-variable hypernetworks for flowfield emulation and shape optimization of compressor airfoils," *AIAA J.* **62**, 741–757 (2024).
15. Z. Chen et al., "Analysis of telescope wavefront aberration and optical path stability in space gravitational wave detection," *Appl. Sci.* **12**(24), 12697 (2022).
16. R. K. Tyson, *Principles of Adaptive Optics*, CRC Press (2015).
17. R. J. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.* **66**, 207–211 (1976).
18. X. Chai et al., "Zernike polynomials fitting of arbitrary shape wavefront," *Proc. SPIE* **13069**, 1306912 (2024).
19. R. Barakat and G. Newsam, "Numerically stable iterative method for the inversion of wave-front aberrations from measured point-spread-function data," *J. Opt. Soc. Am.* **70**, 1255 (1980).
20. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," Adam Optimization Method, CoRR abs/1412.6980 (2014).
21. V. K. Chauhan et al., "A brief review of hypernetworks in deep learning," *Artif. Intell. Rev.* **57**(9), 250 (2024).
22. Y. Xu et al., "Constraint based Bayesian optimization of bioink precursor: a machine learning framework," *Biofabrication* **16**, 045031 (2024).
23. L. Tani and C. Veelken, "Comparison of Bayesian and particle swarm algorithms for hyperparameter optimisation in machine learning applications in high energy physics," *Comput. Phys. Commun.* **294**, 108955 (2024).
24. M. A. Jara, "Codebase for wavefront aberration calibration using supervised and deep learning (shwfs)," (2024), <https://doi.org/10.5281/zenodo.14964320>

25. Y. Endo, Y. Amako, and T. Fukuchi, "Improving techniques for Shack-Hartmann wavefront sensing," in Adv. Maui Opt. Space Surveill. Technol. Conf., <https://amotech.com/TechnicalPapers/2018/Poster/Endo.pdf> (2018).

**Mario Alberto Jara Reyna** holds a bachelor's degree in actuarial science from the Universidad Autónoma de Nuevo León and a master's degree in artificial intelligence from Universidad Tecmilenio. He is currently pursuing a master's degree in Astrophysics at the Universidad Autónoma de Nuevo León and is a consultant specializing in the implementation of artificial intelligence solutions. His research interests are centered on the application of AI in the field of optics.

**Carlos Alberto Guerrero Peña** is a full-time researcher and professor at the Faculty of Physical and Mathematical Sciences, Universidad Autónoma de Nuevo León (UANL). His expertise encompasses stellar and planetary astrophysics, high-resolution spatial techniques, astronomical instrumentation, and the application of machine learning and big data to astrophysical challenges.

**Joel Herrera Vázquez** obtained a degree in physics from the Faculty of Physics and Artificial Intelligence of the University of Veracruz and also obtained his master's and PhD degrees in optical science at the National Institute of Astrophysics, Optics and Electronics. In 2008, he joined the Astronomy Institute of UNAM and the National Astronomical Observatory in B.C. Mexico. His main line of research is focused on optical systems design and astronomical instrumentation.

**Juan José Tapia Armenta** is a full-time researcher and professor at the Digital Technology Research and Development Center, Instituto Politécnico Nacional. His research interests include high-performance computing, data science analytics, machine learning, and multi-GPU computing. In addition, he has experience in mathematical modeling and evolutionary computing algorithms.

**Eduardo Gerardo Pérez Tijerina** is a leading Mexican researcher in astronomy and nanotechnology. A member of SNI Level III and the Mexican Academy of Sciences, he founded Nuevo León's first Nanoscience and Nanotechnology Laboratory and leads the UANL Astronomical Observatory, collaborating with international space programs. He received the Nuevo León Civic Merit Medal in 2019 and advised the state government on science and technology since 2023.

## A.2. Estadísticas Detalladas de las Aberraciones

Este apéndice presenta tablas estadísticas detalladas de las magnitudes de las aberraciones y sus componentes  $x$  y  $y$  tanto para el conjunto de datos base como para el conjunto de datos sintéticos, proporcionando una visión general numérica completa para la evaluación de la reconstrucción del frente de onda. Las Tablas [A.1](#), [A.2](#) y [A.3](#) enumeran los polinomios de Zernike desde el coeficiente 1 hasta el 36, facilitando la referencia cruzada y garantizando la consistencia en el análisis.

**Tabla A.1:** Resumen estadístico de los coeficientes de aberración en la muestra base y sintéticas.

Coeficiente	$M_{\text{base}}(\lambda)$	$\sigma_{\text{base}}(\lambda)$	$M_{\text{sint}}(\lambda)$	$\sigma_{\text{sint}}(\lambda)$	$M_{\text{dif}}(\lambda)$	$\sigma_{\text{dif}}(\lambda)$
1	2.0000	$1.2480 \times 10^{-11}$	2.0000	$1.2580 \times 10^{-11}$	0.0000	$1.0000 \times 10^{-13}$
2	2.0000	$1.3100 \times 10^{-11}$	2.0000	$1.3200 \times 10^{-11}$	0.0000	$1.0000 \times 10^{-13}$
3	$9.6494 \times 10^{-2}$	4.9910	$9.6853 \times 10^{-2}$	4.9910	$3.5929 \times 10^{-4}$	$1.1530 \times 10^{-6}$
4	$6.8231 \times 10^{-2}$	3.5292	$6.8485 \times 10^{-2}$	3.5292	$2.5419 \times 10^{-4}$	$8.1500 \times 10^{-7}$
5	$6.8231 \times 10^{-2}$	3.5292	$6.8485 \times 10^{-2}$	3.5292	$2.5419 \times 10^{-4}$	$8.1500 \times 10^{-7}$
6	2.0173	5.7903	2.0171	5.7903	$1.3668 \times 10^{-4}$	$8.8240 \times 10^{-6}$
7	1.9510	5.8908	1.9515	5.8908	$4.5151 \times 10^{-4}$	$8.5000 \times 10^{-8}$
8	$9.9596 \times 10^{-2}$	4.7251	$9.9552 \times 10^{-2}$	4.7251	$4.3874 \times 10^{-5}$	$2.5703 \times 10^{-6}$
9	$9.9596 \times 10^{-2}$	4.7251	$9.9552 \times 10^{-2}$	4.7251	$4.3873 \times 10^{-5}$	$2.5702 \times 10^{-6}$
10	$3.0680 \times 10^{-2}$	8.1223	$3.0823 \times 10^{-2}$	8.1223	$1.4361 \times 10^{-4}$	$2.3049 \times 10^{-5}$
11	$1.9339 \times 10^{-1}$	8.1908	$1.9123 \times 10^{-1}$	8.1907	$2.1634 \times 10^{-3}$	$1.3138 \times 10^{-5}$
12	$2.6034 \times 10^{-2}$	7.9987	$2.6025 \times 10^{-2}$	7.9987	$9.1080 \times 10^{-6}$	$1.8121 \times 10^{-5}$
13	$1.4888 \times 10^{-1}$	5.7051	$1.4861 \times 10^{-1}$	5.7051	$2.6255 \times 10^{-4}$	$5.7260 \times 10^{-6}$
14	$1.4888 \times 10^{-1}$	5.7051	$1.4861 \times 10^{-1}$	5.7051	$2.6255 \times 10^{-4}$	$5.7260 \times 10^{-6}$
15	1.7874	9.2906	1.7864	9.2906	$9.7942 \times 10^{-4}$	$1.5968 \times 10^{-5}$
16	1.8474	9.0071	1.8478	9.0070	$3.7883 \times 10^{-4}$	$8.0451 \times 10^{-5}$
17	1.1907	9.2974	1.1907	9.2973	$4.4400 \times 10^{-5}$	$1.7132 \times 10^{-5}$
18	1.4273	9.3036	1.4312	9.3036	$3.9690 \times 10^{-3}$	$4.0016 \times 10^{-5}$
19	$4.5661 \times 10^{-2}$	6.5170	$4.6136 \times 10^{-2}$	6.5170	$4.7511 \times 10^{-4}$	$1.0480 \times 10^{-5}$
20	$4.5661 \times 10^{-2}$	6.5170	$4.6136 \times 10^{-2}$	6.5170	$4.7511 \times 10^{-4}$	$1.0480 \times 10^{-5}$
21	$7.3154 \times 10^{-2}$	9.3796	$7.3130 \times 10^{-2}$	9.3796	$2.3593 \times 10^{-5}$	$6.1168 \times 10^{-5}$
22	$1.8281 \times 10^{-1}$	9.6668	$1.8374 \times 10^{-1}$	9.6667	$9.3587 \times 10^{-4}$	$9.0419 \times 10^{-5}$
23	$1.5559 \times 10^{-1}$	$1.0060 \times 10^1$	$1.5476 \times 10^{-1}$	$1.0060 \times 10^1$	$8.3027 \times 10^{-4}$	$2.9070 \times 10^{-5}$
24	$2.9611 \times 10^{-1}$	$1.0186 \times 10^1$	$2.9216 \times 10^{-1}$	$1.0186 \times 10^1$	$3.9539 \times 10^{-3}$	$3.4910 \times 10^{-5}$
25	$3.0912 \times 10^{-1}$	$1.0313 \times 10^1$	$3.0943 \times 10^{-1}$	$1.0313 \times 10^1$	$3.1155 \times 10^{-4}$	$4.5900 \times 10^{-5}$
26	$3.5805 \times 10^{-2}$	7.1919	$3.6253 \times 10^{-2}$	7.1919	$4.4757 \times 10^{-4}$	$1.6426 \times 10^{-5}$
27	$3.5805 \times 10^{-2}$	7.1919	$3.6253 \times 10^{-2}$	7.1919	$4.4757 \times 10^{-4}$	$1.6426 \times 10^{-5}$
28	$5.3593 \times 10^{-1}$	$1.0489 \times 10^1$	$5.3680 \times 10^{-1}$	$1.0489 \times 10^1$	$8.7275 \times 10^{-4}$	$1.0000 \times 10^{-4}$
29	$5.3897 \times 10^{-1}$	$1.0459 \times 10^1$	$5.3548 \times 10^{-1}$	$1.0459 \times 10^1$	$3.4919 \times 10^{-3}$	$1.6000 \times 10^{-7}$
30	1.4786	$1.0867 \times 10^1$	1.4707	$1.0867 \times 10^1$	$7.9120 \times 10^{-3}$	$7.5150 \times 10^{-5}$
31	1.6815	$1.0719 \times 10^1$	1.6811	$1.0719 \times 10^1$	$4.3181 \times 10^{-4}$	$4.0080 \times 10^{-5}$
32	$2.9234 \times 10^{-1}$	$1.1154 \times 10^1$	$2.9226 \times 10^{-1}$	$1.1154 \times 10^1$	$8.1016 \times 10^{-5}$	$5.7540 \times 10^{-5}$
33	$2.8677 \times 10^{-1}$	$1.0889 \times 10^1$	$2.8495 \times 10^{-1}$	$1.0889 \times 10^1$	$1.8207 \times 10^{-3}$	$4.1580 \times 10^{-5}$
34	$3.4269 \times 10^{-2}$	7.7459	$3.4306 \times 10^{-2}$	7.7459	$3.7046 \times 10^{-5}$	$2.3606 \times 10^{-5}$
35	$3.4269 \times 10^{-2}$	7.7459	$3.4306 \times 10^{-2}$	7.7459	$3.7041 \times 10^{-5}$	$2.3606 \times 10^{-5}$
36	$2.2239 \times 10^{-1}$	$1.1502 \times 10^1$	$2.2334 \times 10^{-1}$	$1.1502 \times 10^1$	$9.5129 \times 10^{-4}$	$2.1370 \times 10^{-5}$

**Tabla A.2:** Resumen estadístico de los componentes de aberración X e Y en la muestra base.

Coeficiente	$M_{\text{base},X}(\lambda)$	$\sigma_{\text{base},X}(\lambda)$	$M_{\text{base},Y}(\lambda)$	$\sigma_{\text{base},Y}(\lambda)$
1	0	$8.3552 \times 10^{-12}$	2.0000	$9.2704 \times 10^{-12}$
2	2.0000	$9.9600 \times 10^{-12}$	0	$8.4682 \times 10^{-12}$
3	$-9.5059 \times 10^{-2}$	3.5122	$1.6580 \times 10^{-2}$	3.5461
4	$1.1724 \times 10^{-2}$	2.5074	$-6.7217 \times 10^{-2}$	2.4835
5	$6.7217 \times 10^{-2}$	2.4835	$1.1724 \times 10^{-2}$	2.5074
6	$2.0154 \times 10^0$	4.7464	$8.6427 \times 10^{-2}$	3.3167
7	$8.6427 \times 10^{-2}$	3.3167	$1.9491 \times 10^0$	4.8684
8	$-4.9495 \times 10^{-2}$	3.3655	$8.6427 \times 10^{-2}$	3.3167
9	$-8.6427 \times 10^{-2}$	5.6675	$-4.9495 \times 10^{-2}$	3.3655
10	$2.9352 \times 10^{-2}$	5.6805	$-8.9299 \times 10^{-3}$	5.8182
11	$1.3606 \times 10^{-1}$	5.6266	$-1.3743 \times 10^{-1}$	5.9009
12	$-9.7521 \times 10^{-4}$	4.0291	$-2.6015 \times 10^{-2}$	5.6851
13	$1.4887 \times 10^{-1}$	4.0391	$8.5169 \times 10^{-4}$	4.0391
14	$8.5169 \times 10^{-4}$	4.6087	$-1.4887 \times 10^{-1}$	4.0291
15	$1.8338 \times 10^{-2}$	6.4110	$-1.7873 \times 10^0$	8.0670
16	$-1.8473 \times 10^0$	6.5613	$1.8338 \times 10^{-2}$	4.6087
17	$-2.8345 \times 10^{-2}$	4.5895	$-1.1904 \times 10^0$	6.7335
18	$1.4272 \times 10^0$	4.6268	$-1.5751 \times 10^{-2}$	6.5960
19	$3.5556 \times 10^{-2}$	6.5772	$-2.8648 \times 10^{-2}$	4.6268
20	$-2.8648 \times 10^{-2}$	6.7687	$-3.5556 \times 10^{-2}$	4.5895
21	$4.8305 \times 10^{-2}$	7.0124	$-5.4937 \times 10^{-2}$	6.6871
22	$-4.2951 \times 10^{-2}$	7.0695	$1.7769 \times 10^{-1}$	6.9015
23	$1.5372 \times 10^{-1}$	7.2236	$2.4037 \times 10^{-2}$	7.2127
24	$1.1842 \times 10^{-2}$	5.0513	$2.9587 \times 10^{-1}$	7.3336
25	$-3.0491 \times 10^{-1}$	5.1194	$5.0850 \times 10^{-2}$	7.3603
26	$-1.4733 \times 10^{-3}$	9.1247	$-3.5775 \times 10^{-2}$	5.1194
27	$3.5775 \times 10^{-2}$	5.1734	$-1.4733 \times 10^{-3}$	5.0513
28	$-5.3410 \times 10^{-1}$	7.7177	$4.4254 \times 10^{-2}$	5.1734
29	$4.4254 \times 10^{-2}$	7.6937	$-5.3715 \times 10^{-1}$	9.0904
30	$-1.4782 \times 10^0$	7.8611	$3.6075 \times 10^{-2}$	7.6499
31	$-3.9753 \times 10^{-2}$	7.5189	$1.6810 \times 10^0$	7.6937
32	$-2.9166 \times 10^{-1}$	5.5475	$1.9900 \times 10^{-2}$	7.9128
33	$3.6213 \times 10^{-2}$	5.4059	$-2.8447 \times 10^{-1}$	7.8765
34	$3.3067 \times 10^{-2}$	8.1908	$8.9973 \times 10^{-3}$	5.4059
35	$-8.9973 \times 10^{-3}$	8.0747	$3.3067 \times 10^{-2}$	5.5475
36	$-2.2229 \times 10^{-1}$	7.8398	$-6.6376 \times 10^{-3}$	8.0747

**Tabla A.3:** Resumen estadístico de los componentes de aberración X e Y en las muestras sintéticas.

Coeficiente	$M_{\text{sint},X}(\lambda)$	$\sigma_{\text{sint},X}(\lambda)$	$M_{\text{sint},Y}(\lambda)$	$\sigma_{\text{sint},Y}(\lambda)$
1	0	$8.3554 \times 10^{-12}$	2.0000	$9.2711 \times 10^{-12}$
2	2.0000	$9.9610 \times 10^{-12}$	0	$8.4679 \times 10^{-12}$
3	$-9.5421 \times 10^{-2}$	3.5122	$1.6593 \times 10^{-2}$	3.5461
4	$1.1733 \times 10^{-2}$	2.5074	$-6.7473 \times 10^{-2}$	2.4835
5	$6.7473 \times 10^{-2}$	2.4835	$1.1733 \times 10^{-2}$	2.5074
6	$2.0153 \times 10^0$	4.7463	$8.6388 \times 10^{-2}$	3.3167
7	$8.6388 \times 10^{-2}$	3.3167	$1.9496 \times 10^0$	4.8684
8	$-4.9476 \times 10^{-2}$	3.3655	$8.6388 \times 10^{-2}$	3.3167
9	$-8.6388 \times 10^{-2}$	3.3167	$-4.9476 \times 10^{-2}$	3.3655
10	$2.9485 \times 10^{-2}$	5.6675	$-8.9845 \times 10^{-3}$	5.8182
11	$1.3421 \times 10^{-1}$	5.6804	$-1.3622 \times 10^{-1}$	5.9009
12	$2.9200 \times 10^{-4}$	5.6266	$-2.6023 \times 10^{-2}$	5.6851
13	$1.4861 \times 10^{-1}$	4.0291	$-9.4900 \times 10^{-4}$	4.0391
14	$-9.4900 \times 10^{-4}$	4.0391	$-1.4861 \times 10^{-1}$	4.0291
15	$1.8847 \times 10^{-2}$	4.6087	$-1.7863 \times 10^0$	8.0670
16	$-1.8477 \times 10^0$	7.7386	$-1.1904 \times 10^0$	6.7335
17	$-2.4613 \times 10^{-2}$	6.4110	$-1.7115 \times 10^{-2}$	6.5959
18	$1.4311 \times 10^0$	6.5612	$-2.8449 \times 10^{-2}$	4.6268
19	$3.6321 \times 10^{-2}$	4.5895	$-3.6321 \times 10^{-2}$	4.5895
20	$-2.8449 \times 10^{-2}$	4.6268	$-5.4517 \times 10^{-2}$	6.6871
21	$4.8744 \times 10^{-2}$	6.5771	$1.7867 \times 10^{-1}$	6.9015
22	$-4.2854 \times 10^{-2}$	6.7686	$1.7921 \times 10^{-2}$	7.2128
23	$1.5372 \times 10^{-1}$	7.0123	$2.9191 \times 10^{-1}$	7.3336
24	$1.1899 \times 10^{-2}$	7.0694	$5.1405 \times 10^{-2}$	7.3603
25	$-3.0513 \times 10^{-1}$	7.2236	$-3.6247 \times 10^{-2}$	5.1194
26	$-6.4450 \times 10^{-4}$	5.0513	$-6.4450 \times 10^{-4}$	5.0513
27	$3.6247 \times 10^{-2}$	5.1194	$4.5198 \times 10^{-2}$	5.1733
28	$-5.3489 \times 10^{-1}$	9.1246	$-5.3357 \times 10^{-1}$	9.0904
29	$4.5198 \times 10^{-2}$	5.1733	$4.3189 \times 10^{-2}$	7.6499
30	$-1.4701 \times 10^0$	7.7176	$1.6806 \times 10^0$	7.6937
31	$-4.0455 \times 10^{-2}$	7.4635	$1.8881 \times 10^{-2}$	7.9128
32	$-2.9165 \times 10^{-1}$	7.5189	$-2.8198 \times 10^{-1}$	7.8765
33	$4.1009 \times 10^{-2}$	5.5475	$9.2385 \times 10^{-3}$	5.4059
34	$3.3039 \times 10^{-2}$	5.4059	$3.3039 \times 10^{-2}$	5.5475
35	$-9.2385 \times 10^{-3}$	8.1908	$-5.8090 \times 10^{-3}$	8.0747
36	$-2.2327 \times 10^{-1}$	8.1908	$-5.8090 \times 10^{-3}$	8.0747