



**Aplicación de Algoritmos Cuánticos Variacionales a la Discriminación de  
Estados Cuánticos**

Dayana Andrea Henao Arbeláez

Trabajo de grado para optar al título de Física

Asesor

Boris Anghelo Rodríguez Rey, PhD.

Coasesor

Sebastián Duque Mesa, M.Sc.

Universidad de Antioquia

Instituto de Física

Facultad de Ciencias Exactas y Naturales

Medellín, Colombia

2026

---

|                           |   |
|---------------------------|---|
| <b>Cita</b>               | Henao Arbeláez, 2026 [1]  |
| <b>Referencia</b>         | [1] Henao Arbeláez, D. A., “Aplicación de Algoritmos Cuánticos Variacionales a la Discriminación de Estados Cuánticos”, Trabajo de grado para optar al título de Física, Instituto de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia, Medellín, Colombia, 2026. |
| <b>Estilo IEEE (2020)</b> |   |

---



Grupo de Investigación: Grupo de Física Atómica y Molecular.  
Instituto de Física, Universidad de Antioquia.



**Repositorio Institucional:** <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia – [www.udea.edu.co](http://www.udea.edu.co)

**Rector:** Héctor Iván García García.

**Decana:** Adriana Echavarría Isaza.

**Director de Instituto:** Jorge Mario Osorio Guillen.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

## **Dedicatoria**

A quien ha sido mi hogar y mi refugio durante más de seis años.

## **Agradecimientos**

Quisiera comenzar agradeciéndole al profesor que conocí en primer semestre y que, sin darse cuenta, me acompañó durante toda la carrera de diferentes formas. Me acompañó con frases reflexivas al inicio de las clases; con consejos dirigidos a todo un curso que a mí me ayudaron muchísimo; con recomendaciones de libros; con oportunidades para mejorar y, finalmente, como mi asesor. Muchas gracias, profesor Boris, por su paciencia y por su ayuda durante todo este tiempo.

También quisiera agradecer especialmente al profesor Sebastián Duque por su disposición, su tiempo y la orientación que me brindó a lo largo de este trabajo. Finalmente, agradezco a mi familia y a las personas que me acompañaron y apoyaron durante los momentos más difíciles de este proceso.

## TABLA DE CONTENIDO

|   |    |
|---|----|
| RESUMEN . . . . .   | 9  |
| ABSTRACT . . . . .  | 10 |
| I Introducción . . . . .  | 11 |
| II Fundamentos Teóricos de la Discriminación de Estados . . . . .                 | 16 |
| A. Estados Cuánticos, Operadores y Operadores Densidad . . . . .                  | 16 |
| B. Mediciones Generalizadas (POVMs) . . . . .                                     | 17 |
| C. Discriminación de estados mediante POVMs . . . . .                             | 18 |
| D. Rango y Soporte de Operadores Hermíticos . . . . .                             | 19 |
| E. Fidelidad entre estados cuánticos . . . . .                                    | 20 |
| F. Teorema de Naimark. . . . .  | 21 |
| III Estrategias de Discriminación de Estados . . . . .                            | 23 |
| A. Discriminación de Estados Cuánticos. . . . .                                   | 23 |
| B. Optimización Convexa y Programación Semidefinida . . . . .                     | 25 |
| IV Algoritmos Cuánticos Variacionales y Diseño del Clasificador . . . . .         | 29 |
| A. Algoritmos Cuánticos Variacionales . . . . .                                   | 29 |
| B. Implementación de POVMs mediante VQCs. . . . .                                 | 34 |
| C. Alcances y limitaciones de los VQAs . . . . .                                  | 36 |
| V Definición de las Familias de Estados Cuánticos y Diseño Experimental . . . . . | 37 |
| A. Definición de las Familias de Estados Cuánticos . . . . .                      | 37 |
| B. Generación de Datos. . . . .   | 41 |
| C. Implementación del Clasificador . . . . .                                      | 43 |
| D. Función de Costo . . . . .   | 46 |
| E. Elección de Hiperparámetros . . . . .  | 46 |
| F. Estrategia de Optimización . . . . .   | 47 |
| G. Herramientas Computacionales . . . . .   | 49 |
| VI Resultados y Discusión . . . . .   | 51 |
| A. Clasificador para las Familias de Validación . . . . .                         | 51 |
| B. Clasificador para las Familias de Estados EJM, Popescu y Werner . . . . .      | 72 |
| C. Resultados del Clasificador en el Escenario de Tres Familias . . . . .         | 78 |
| VII Conclusiones y Perspectivas . . . . .   | 82 |
| REFERENCIAS . . . . .   | 85 |

## LISTA DE TABLAS

|          |   |    |
|----------|---|----|
| Tabla 1  | Escenarios de clasificación considerados . . . . .                          | 38 |
| Tabla 2  | Cotas de Helstrom para la discriminación de mínimo error . . . . .          | 61 |
| Tabla 3  | Cotas teóricas para la discriminación inequívoca (USD) . . . . .            | 61 |
| Tabla 4  | Resultados promedio del clasificador (USD) . . . . .                        | 65 |
| Tabla 5  | Resultados promedio del clasificador en el régimen MESD . . . . .           | 68 |
| Tabla 6  | Resultados promedio en MESD con clasificador modificado . . . . .           | 72 |
| Tabla 7  | Probabilidades óptimas por pares de familias . . . . .                      | 73 |
| Tabla 8  | Probabilidades óptimas para discriminación triple . . . . .                 | 73 |
| Tabla 9  | Resultados del clasificador en MESD para discriminación por pares . . . . . | 76 |
| Tabla 10 | Resultados en MESD sin salidas inconclusas . . . . .                        | 78 |
| Tabla 11 | Resultado del clasificador en discriminación triple (MESD) . . . . .        | 79 |

## LISTA DE FIGURAS

|         |  |    |
|---------|--|----|
| Fig. 1  | Representación Geométrica de Estados No Ortogonales . . . . .  | 11 |
| Fig. 2  | Ciclo híbrido CPU–QPU en un VQA . . . . .  | 30 |
| Fig. 3  | Implementación unitaria de una POVM . . . . .  | 35 |
| Fig. 4  | Estructura tetraédrica de la base EJM . . . . .  | 40 |
| Fig. 5  | Arquitectura del clasificador variacional . . . . .  | 43 |
| Fig. 6  | Descomposición explícita de los bloques $U$ y $V_{1,2}$ . . . . .  | 44 |
| Fig. 7  | Distribución de fidelidad — Caso 1, $N=10-100$ . . . . .   | 52 |
| Fig. 8  | Distribución de fidelidad — Caso 1, $N$ 200-1000 . . . . .   | 53 |
| Fig. 9  | Distribución de fidelidad — Caso 2, $N=10-100$ . . . . .   | 54 |
| Fig. 10 | Distribución de fidelidad — Caso 2, $N$ 200-1000 . . . . .   | 55 |
| Fig. 11 | Distribución de fidelidad — Caso 3, $N=10-100$ . . . . .   | 56 |
| Fig. 12 | Distribución de fidelidad — Caso 3, $N$ 200-1000 . . . . .   | 57 |
| Fig. 13 | Distribución de fidelidad — Caso 4, $N=10-100$ . . . . .   | 58 |
| Fig. 14 | Distribución de fidelidad — Caso 4, $N$ 200-1000 . . . . .   | 59 |
| Fig. 15 | Desempeño del clasificador en el régimen USD . . . . .   | 63 |
| Fig. 16 | Matriz de confusión: Caso 1, $N = 1000$ con $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (40, 2)$ . . . . .                   | 64 |
| Fig. 17 | Curvas de entrenamiento y validación: Caso 4, $N=10$ con $(\alpha_{\text{err}}, \alpha_{\text{inc}}) =$<br>$(35, 2)$ . . . . . | 66 |
| Fig. 18 | Matriz de confusión: Caso 1, $N=10$ con $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (5, 2)$ . . . . .                        | 67 |
| Fig. 19 | Matriz de confusión: Caso 4, $N=20$ con $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (5, 20)$ . . . . .                       | 69 |
| Fig. 20 | Matriz de confusión: Caso 4, $N = 20$ en régimen MESD . . . . .  | 71 |
| Fig. 21 | Matriz de confusión: EJM–Werner con salida inconclusa . . . . .  | 75 |
| Fig. 22 | Evolución de la función de costo: EJM–Werner . . . . .   | 77 |
| Fig. 23 | Evolución de tasas: EJM–Werner con $(5, 10)$ . . . . .   | 77 |
| Fig. 24 | Matriz de confusión: EJM–Popescu–Werner con $(5, 20)$ . . . . .  | 80 |

## Siglas, acrónimos y abreviaturas

|              |                                     |
|--------------|-------------------------------------|
| <b>CPU</b>   | Central Processing Unit             |
| <b>EJM</b>   | Elegant Joint Measurements          |
| <b>FRIO</b>  | Fixed Rate of Inconclusive Outcomes |
| <b>MESD</b>  | Minimum Error State Discrimination  |
| <b>NISQ</b>  | Noisy Intermediate-Scale Quantum    |
| <b>PBS</b>   | Polarizing Beam Splitter            |
| <b>POVMs</b> | Positive Operator-Valued Measures   |
| <b>QPU</b>   | Quantum Processing Unit             |
| <b>QSD</b>   | Quantum State Discrimination        |
| <b>SDP</b>   | Semidefinite Programming            |
| <b>USD</b>   | Unambiguous State Discrimination    |
| <b>VQAs</b>  | Variational Quantum Algorithms      |
| <b>VQCs</b>  | Variational Quantum Circuits        |

---

## RESUMEN

En este trabajo se estudia el uso de algoritmos cuánticos variacionales para la discriminación de estados cuánticos, mediante el entrenamiento de un clasificador cuántico simulado capaz de implementar esquemas de medición que aproximan mediciones óptimas. Este enfoque permite estudiar estrategias de medición desde una perspectiva híbrida clásico–cuántica, aprovechando las capacidades de los dispositivos NISQ y evitando la complejidad asociada a implementaciones experimentales directas.

En una primera parte, se valida el desempeño del clasificador mediante la reproducción parcial de un esquema de redes neuronales cuánticas variacionales propuesto previamente en la literatura. Se analizan tanto el régimen de discriminación de mínimo error como el de discriminación inequívoca, comparando los resultados numéricos con las cotas teóricas obtenidas a partir del límite de Helstrom y de técnicas de programación semidefinida. En el régimen de discriminación inequívoca, el clasificador logra aproximarse de manera robusta a las cotas teóricas, incluso para conjuntos de entrenamiento reducidos.

En una segunda parte, el clasificador se generaliza para abordar problemas más complejos que involucran familias de estados asociadas a la Medición Conjunta Elegante (EJM), así como a familias de estados de Popescu y de Werner, tanto en escenarios binarios como en multiclase. Los resultados muestran que, para estas familias, la discriminación inequívoca es imposible y que, en el régimen de mínimo error, el desempeño del clasificador se ve limitado por restricciones geométricas asociadas al rango y al soporte de los operadores de medición implementados.

***Palabras clave*** — Discriminación de estados cuánticos, Algoritmos cuánticos variacionales, Positive Operator-Valued Measures, Teorema de Naimark.

## ABSTRACT

In this work, we study the use of variational quantum algorithms for the discrimination of quantum states by training a simulated quantum classifier capable of implementing measurement schemes that approximate optimal measurements. This approach allows the study of measurement strategies from a hybrid classical–quantum perspective, leveraging the capabilities of NISQ devices while avoiding the complexity associated with direct experimental implementations.

In a first part, the classifier performance is validated by partially replicating a previously proposed variational quantum classification scheme. Both the minimum-error discrimination regime and the unambiguous state discrimination regime are analyzed, and the numerical results are compared with the corresponding theoretical bounds derived from the Helstrom limit and semidefinite programming techniques. In the unambiguous discrimination regime, the classifier robustly approaches the theoretical bounds, even when trained with relatively small datasets.

In the second part, the classifier is generalized to address more complex problems involving families of states associated with the Elegant Joint Measurement (EJM), as well as Popescu and Werner state families, in both binary and multiclass scenarios. The results show that, for these families, unambiguous discrimination is impossible and that, in the minimum-error regime, the classifier performance is limited by geometric constraints related to the rank and support of the implemented measurement operators.

***Keywords*** — Quantum state discrimination, Variational quantum algorithms, Positive Operator-Valued Measures, Naimark’s theorem.

# CAPÍTULO I: INTRODUCCIÓN

La discriminación de estados cuánticos es un problema fundamental en la teoría de la información cuántica que consiste en determinar en cuál de varios posibles estados ha sido preparado un sistema. La mecánica cuántica prohíbe de manera fundamental la discriminación determinista (sin error) de estados cuando los vectores que los representan no son ortogonales entre sí, es decir, cuando su solapamiento es distinto de cero [1, 2], como se ilustra en la Fig. 1. Este problema establece los límites teóricos de la información que puede extraerse de un sistema sin perturbar significativamente su estado original, y la capacidad, o incapacidad, de distinguir perfectamente entre diferentes estados cuánticos es la base de muchas tecnologías cuánticas [3]. Por ejemplo, en áreas como la criptografía cuántica, la imposibilidad de distinguir perfectamente estados no ortogonales garantiza la seguridad de los protocolos de comunicación [4]; en metrología cuántica, la detección de cambios sutiles en un sistema está ligada a la capacidad de diferenciar estados con alta precisión [5]; en la computación cuántica, el éxito de ciertos algoritmos depende de la distinción eficiente entre familias de estados [1]; y en la comunicación cuántica, las señales se codifican frecuentemente en estados no ortogonales [6].

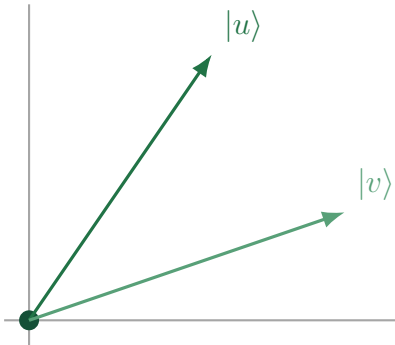


Figura 1: Representación geométrica de dos estados cuánticos puros no ortogonales,  $|u\rangle$  y  $|v\rangle$ , que comparten el mismo origen pero forman un ángulo distinto de  $\pi/2$ . La no ortogonalidad implica un solapamiento interno no nulo  $\langle u|v\rangle \neq 0$ .

Desde un punto de vista formal, dos estados cuánticos  $|u\rangle$  y  $|v\rangle$  pueden discri-

minarse perfectamente mediante una medición proyectiva adecuada siempre que sean ortogonales [3], es decir, cuando  $\langle u|v\rangle = 0$ , ya que en este caso no existe proyección de uno sobre el otro [7]. Por el contrario, cuando los estados no son ortogonales ( $\langle u|v\rangle \neq 0$ ), cualquier procedimiento de medición conlleva inevitablemente una probabilidad no nula de error o la aparición de resultados inconclusos [8]. Esta limitación surge del solapamiento entre los estados, el cual impide identificar con certeza en cuál de ellos fue preparado el sistema, y puede comprenderse de manera intuitiva a través del análisis presentado en [2].

Se considera un fotón preparado con polarización horizontal ( $|H\rangle$ ) o vertical ( $|V\rangle$ ), estados que son ortogonales entre sí. En este caso, es posible distinguir el estado del fotón de forma determinista mediante un divisor de haz polarizante (PBS, del inglés Polarizing Beam Splitter) orientado convenientemente. El fotón es transmitido si se encuentra en el estado vertical y reflejado si se encuentra en el estado horizontal, lo que permite una identificación exacta del estado inicial. Sin embargo, si el fotón ha sido preparado en el estado horizontal ( $|H\rangle$ ) o en el estado de polarización circular izquierda ( $|L\rangle$ ), la situación es completamente diferente. Estos estados no son ortogonales, ya que la polarización circular puede expresarse como una superposición de las componentes horizontal y vertical,

$$|L\rangle = \frac{1}{\sqrt{2}} (|H\rangle + i|V\rangle), \quad (1)$$

lo que implica un solapamiento no nulo entre ambos estados,

$$\langle H|L\rangle = \frac{1}{\sqrt{2}} \neq 0. \quad (2)$$

Si se intenta discriminar estos estados utilizando el mismo montaje experimental basado en el PBS, es decir, realizando la medición en la base  $\{|H\rangle, |V\rangle\}$ , un fotón preparado en el estado  $|L\rangle$  será detectado de manera aleatoria como horizontal o vertical, con probabilidades iguales. Por lo tanto, un resultado correspondiente al estado  $|H\rangle$  no permite determinar de forma inequívoca si el fotón fue preparado originalmente en  $|H\rangle$  o en  $|L\rangle$ , lo que evidencia la imposibilidad de una discriminación perfecta cuando se emplean mediciones proyectivas sobre estados no ortogonales.

Existen principalmente dos estrategias fundamentales para la discriminación de estados cuánticos no ortogonales: la discriminación de mínimo error y la discriminación inequívoca. La *discriminación de mínimo error* (MESD) tiene como objetivo identificar el estado preparado minimizando la probabilidad total de error, aceptando una tasa finita de clasificaciones incorrectas. Este enfoque fue desarrollado en trabajos de Helstrom y Holevo [9, 10], quienes establecieron los límites teóricos fundamentales para la probabilidad máxima de éxito en la discriminación de estados cuánticos.

Por su parte, la *discriminación inequívoca de estados* (USD) permite identificar sin error estados cuánticos no ortogonales, a costa de aceptar una probabilidad finita de resultados inconclusos. Esta estrategia fue introducida en los trabajos de Ivanovic, Dieks y Peres [11, 12, 13], donde se mostró que el uso de mediciones generalizadas permite superar las limitaciones impuestas por las mediciones proyectivas estándar.

La discriminación de estados cuánticos no ortogonales ha sido abordada mediante diversas configuraciones experimentales, particularmente en sistemas ópticos. En el contexto de la discriminación de mínimo error, Barnett y Riis en 1997 [8] presentaron un diseño experimental para la discriminación entre dos estados de polarización no ortogonales, alcanzando el límite teórico de mínimo error mediante una medición óptima. Posteriormente, Clarke et al. en 2001 [14] extendieron estas ideas a conjuntos simétricos de tres y cuatro estados de polarización, implementando redes ópticas interferométricas que permitieron realizar mediciones generalizadas a través de una ampliación efectiva del espacio de estados accesible al sistema.

Más recientemente, se han propuesto estrategias intermedias como la discriminación con tasa fija de resultados inconclusos (FRIO, *Fixed Rate of Inconclusive Outcomes*) [15], que permiten interpolar de manera controlada entre los regímenes de discriminación de mínimo error y de discriminación inequívoca mediante el ajuste de la probabilidad de resultados inconclusos. La realización experimental se basa en una configuración interferométrica que permite llevar a cabo mediciones no proyectivas con tres posibles resultados, mostrando resultados acordes con las predicciones teóricas.

No obstante, la implementación experimental de estos esquemas suele requerir un

---

equipamiento óptico especializado y de alta precisión. Estas exigencias técnicas pueden limitar su escalabilidad y dificultar su integración en arquitecturas más generales de procesamiento de información. Como consecuencia, han surgido enfoques alternativos que abordan el problema de la discriminación de estados cuánticos desde una perspectiva más flexible y escalable, apoyada en herramientas computacionales.

En este contexto, la computación cuántica surge como un nuevo modelo de procesamiento de información que aprovecha propiedades propiamente cuánticas, como la superposición y el entrelazamiento. Originalmente, esta área se desarrolló a partir de la idea de que la simulación de sistemas físicos complejos podría realizarse de manera más eficiente en un computador cuántico [16]. Con el paso del tiempo, esta motivación inicial, fuertemente ligada a la física, se ha ampliado hacia otras áreas, incluyendo la optimización, la criptografía y, más recientemente, la inteligencia artificial. En particular, dentro del modelo de computación basado en circuitos, se han propuesto algoritmos de aprendizaje automático cuántico (QML) que, al menos a nivel teórico, apuntan a posibles ventajas computacionales frente a métodos clásicos en tareas de procesamiento de información [17]

No obstante, la implementación práctica de estos algoritmos enfrenta importantes desafíos tecnológicos. La etapa actual del desarrollo de la computación cuántica se conoce como la era NISQ (*Noisy Intermediate-Scale Quantum*) [18], caracterizada por procesadores cuánticos con un número limitado de cúbits y por tasas de error significativas [19]. Esto se debe a que, la profundidad de los circuitos que pueden ejecutarse está fuertemente restringida por el tiempo de coherencia de los cúbits; si el cálculo se extiende más allá de este límite, la información cuántica se degrada como consecuencia de la decoherencia [20, 21, 22].

En este contexto, los Algoritmos Cuánticos Variacionales (VQAs) surgen como una alternativa para trabajar dentro de las limitaciones de la era NISQ. En lugar de implementar circuitos profundos, estos algoritmos emplean circuitos cuánticos parametrizados de baja profundidad, cuyos parámetros se ajustan de manera iterativa a partir de mediciones realizadas sobre el propio dispositivo. Este ajuste se lleva a cabo me-

---

diante un esquema híbrido clásico–cuántico, en el cual un optimizador clásico utiliza la información obtenida de las mediciones cuánticas para guiar la evolución del circuito [23, 24].

Dado que los estados cuánticos constituyen datos intrínsecamente cuánticos, resulta natural considerar que los procesadores cuánticos puedan ofrecer ventajas frente a métodos clásicos en tareas de discriminación. Estudios recientes han mostrado la posibilidad de discriminar estados no ortogonales mediante clasificadores cuánticos [17], así como el impacto de distintos modelos de ruido en su desempeño [19].

Estos antecedentes motivan el desarrollo del presente trabajo, cuyo objetivo es entrenar y evaluar un clasificador simulado basado en algoritmos cuánticos variacionales (VQAs) para la discriminación de estados cuánticos. Este enfoque permite estudiar la medición óptima desde una perspectiva programable, aprovechando las capacidades de los dispositivos NISQ y evitando la complejidad asociada a montajes experimentales en un laboratorio.

El trabajo se estructura en dos partes principales. En la primera, se replican parcialmente los resultados presentados en el artículo *Universal Discriminative Quantum Neural Networks* [17], con el objetivo de establecer un punto de referencia que permita validar el desempeño del clasificador implementado. En este esquema, el entrenamiento se realiza mediante el algoritmo de optimización Adam dentro de un marco híbrido clásico–cuántico, ajustando los parámetros del circuito variacional de tal forma que el mismo implemente de manera efectiva una Medida Valuada en Operadores Positivos (POVM) adecuada para la tarea de clasificación. En la segunda parte, el clasificador se generaliza para abordar problemas más complejos que involucran tres familias de estados cuánticos: estados asociados a la Medición Conjunta Elegante (*Elegant Joint Measurement*, EJM), familias de estados de Werner y familias de estados de Popescu, lo que permite analizar la robustez y versatilidad del modelo propuesto más allá del caso binario inicialmente considerado.

# CAPÍTULO II: FUNDAMENTOS DE LA DISCRIMINACIÓN DE ESTADOS

## *A. Estados Cuánticos, Operadores y Operadores Densidad*

El estado de un sistema cuántico contiene toda la información que puede obtenerse mediante mediciones sobre el sistema, y se representa mediante un objeto matemático llamado vector de estado o ket  $|\psi\rangle$ . Los vectores de estado son vectores normalizados que pertenecen a un espacio vectorial conocido como espacio de Hilbert [1, 2].

En mecánica cuántica, los observables físicos se describen mediante operadores lineales que actúan sobre los vectores de estado. En particular, los observables están asociados a operadores hermíticos  $\hat{O}$ , que satisfacen

$$\hat{O} = \hat{O}^\dagger, \quad (3)$$

donde  $\hat{O}^\dagger$  denota el operador adjunto. La hermiticidad garantiza que los posibles resultados de una medición sean valores reales [7].

En este marco se distinguen dos tipos de estados cuánticos. Un *estado puro* corresponde a un conocimiento completo del sistema y puede describirse mediante un vector de estado  $|\psi\rangle$ . En contraste, un *estado mezclado* representa una mezcla estadística de varios estados posibles  $|\psi_i\rangle$ , cada uno asociado a una probabilidad  $p_i$  [4].

Tanto los estados puros como los estados mezclados pueden representarse mediante un operador denominado *operador densidad*  $\rho$ , el cual es un operador positivo semidefinido y de traza unitaria,

$$\rho \geq 0, \quad \text{Tr}(\rho) = 1, \quad (4)$$

donde la condición  $\rho \geq 0$  indica que

$$\langle \phi | \rho | \phi \rangle \geq 0, \quad (5)$$

para todo vector de estado  $|\phi\rangle$ .

Para un estado puro  $|\psi\rangle$ , el operador densidad se define como

$$\rho = |\psi\rangle\langle\psi|, \quad (6)$$

y satisface la condición  $\rho^2 = \rho$ . En el caso de un estado mezclado, la matriz densidad se escribe como

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|, \quad 0 \leq p_i \leq 1, \quad \sum_i p_i = 1. \quad (7)$$

con  $|\psi_i\rangle$  estados puros normalizados.

La formulación en términos de matrices densidad resulta esencial para describir situaciones en las que el sistema interactúa con un entorno o cuando el conocimiento sobre su estado es incompleto, y constituye el marco natural para el estudio de procesos de medición generalizados y problemas de discriminación de estados cuánticos [4].

### B. Mediciones Generalizadas (POVMs)

En mecánica cuántica, el proceso de medición se describe mediante un conjunto de operadores que actúan sobre el espacio de Hilbert del sistema. El caso más habitual corresponde a las mediciones proyectivas (o de von Neumann), en las cuales cada resultado posible se asocia a un proyector ortogonal  $\hat{P}_i$  que satisface las condiciones de ortogonalidad e idempotencia:

$$\hat{P}_i \hat{P}_j = \delta_{ij} \hat{P}_i, \quad \sum_i \hat{P}_i = \mathbb{I}. \quad (8)$$

En este formalismo, los resultados de la medición corresponden a subespacios ortogonales que cubren la totalidad del espacio de Hilbert. Cuando los estados cuánticos a discriminar son ortogonales, este tipo de mediciones permite distinguirlos de manera determinista [4, 2].

Sin embargo, para abordar la discriminación de estados no ortogonales, es necesario recurrir a las mediciones generalizadas, conocidas como **POVM** (*Positive Operator-Valued Measure*). Una POVM se define como un conjunto de operadores hermíticos  $\hat{E}_i = \hat{E}_i^\dagger$ , positivos semidefinidos  $\langle\psi|\hat{E}_i|\psi\rangle \geq 0$ , que satisfacen la condición de comple-

titud:

$$\sum_i \hat{E}_i = \mathbb{I}, \quad (9)$$

La probabilidad de obtener el resultado  $i$  al medir un sistema en el estado  $\rho$  viene dada por la regla de la traza:

$$p(i) = \text{Tr}(\hat{E}_i \rho). \quad (10)$$

Es común describir los elementos de una POVM mediante un conjunto de operadores  $\{\hat{K}_m\}$ , denominados operadores de Kraus, de tal forma que:

$$\hat{E}_m = \hat{K}_m^\dagger \hat{K}_m, \quad (11)$$

cumpléndose la relación  $\sum_m \hat{K}_m^\dagger \hat{K}_m = \mathbb{I}$ .

A diferencia de las mediciones proyectivas, los elementos de una POVM no requieren ser proyectores ( $\hat{E}_i^2 \neq \hat{E}_i$ ), ni ser ortogonales entre sí. Esta flexibilidad permite que el número de resultados posibles de la medición supere la dimensión del espacio de Hilbert del sistema ( $m \geq d$ ), una característica fundamental para implementar estrategias avanzadas de clasificación y discriminación de estados, como la discriminación inequívoca o la minimización del error [2, 4].

### C. Discriminación de estados mediante POVMs

Para ilustrar cómo las mediciones generalizadas permiten abordar tareas de discriminación de estados cuánticos, se describe el ejemplo de discriminación inequívoca entre dos estados no ortogonales propuesto en la Sección 2.2.6 de Nielsen & Chuang (2010) [4].

Supóngase que un sistema cuántico se prepara, con igual probabilidad, en uno de los estados

$$|\psi_1\rangle = |0\rangle, \quad |\psi_2\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}. \quad (12)$$

Dado que estos estados no son ortogonales, no es posible distinguirlos con certeza perfecta mediante mediciones proyectivas. Sin embargo, es posible diseñar una medición que permita identificarlos correctamente en algunos casos, sin cometer errores de identificación, a costa de admitir resultados inconclusos.

Este objetivo puede alcanzarse mediante una medición descrita por una *POVM* con tres elementos  $\{\hat{E}_1, \hat{E}_2, \hat{E}_?\}$ , definidos como

$$\hat{E}_1 = \frac{\sqrt{2}}{1 + \sqrt{2}} |1\rangle \langle 1|, \quad (13)$$

$$\hat{E}_2 = \frac{\sqrt{2}}{1 + \sqrt{2}} \frac{(|0\rangle - |1\rangle)(\langle 0| - \langle 1|)}{2}, \quad (14)$$

$$\hat{E}_? = I - \hat{E}_1 - \hat{E}_2. \quad (15)$$

Puede verificarse que estos operadores son positivos semidefinidos y satisfacen la relación de completitud

$$\sum_m \hat{E}_m = I, \quad (16)$$

por lo que constituyen una *POVM* válida.

El diseño de los operadores  $\hat{E}_1$  y  $\hat{E}_2$  garantiza que nunca se produzcan errores de identificación. En particular, si el sistema se encuentra en el estado  $|\psi_1\rangle = |0\rangle$ , la probabilidad de obtener el resultado asociado a  $\hat{E}_1$  es nula,

$$\langle \psi_1 | \hat{E}_1 | \psi_1 \rangle = 0. \quad (17)$$

Por lo tanto, si el resultado de la medición es  $\hat{E}_1$ , se puede concluir inequívocamente que el estado preparado fue  $|\psi_2\rangle$ . De manera análoga, si el resultado obtenido es  $\hat{E}_2$ , el estado debe haber sido  $|\psi_1\rangle$ .

El operador  $\hat{E}_?$  corresponde a un resultado inconcluso, en el cual no es posible inferir la identidad del estado preparado. No obstante, la característica fundamental de este esquema es que nunca conduce a una identificación errónea, la ausencia de errores se obtiene al precio de aceptar que, en algunos casos, la medición no proporciona información útil (resultado inconcluso).

#### *D. Rango y Soporte de Operadores Hermíticos*

El *rango* de un operador hermítico  $\hat{O}$ , denotado por  $\text{rank}(\hat{O})$ , se define como la dimensión del subespacio generado por sus vectores propios asociados a valores propios no nulos. En el caso de una matriz densidad  $\rho$ , el rango coincide con la dimensión del

soporte del estado, es decir, del subespacio del espacio de Hilbert en el cual el estado tiene soporte no nulo [25].

Un estado puro  $\rho = |\psi\rangle\langle\psi|$  es un operador de rango uno, mientras que un estado mezclado puede presentar un rango mayor, dependiendo del número de estados puros linealmente independientes que contribuyen a la mezcla. Desde un punto de vista geométrico, esto implica que los estados mezclados ocupan regiones de mayor dimensión en el espacio de Hilbert que los estados puros.

En el contexto de una medición generalizada, cada elemento de una POVM  $\{\hat{\Pi}_i\}$  define el efecto asociado a un resultado particular. La probabilidad de identificar un estado  $\rho$  mediante el resultado correspondiente a  $\Pi_i$  está dada por

$$p(i|\rho) = \text{Tr}(\hat{\Pi}_i\rho). \quad (18)$$

Cuando  $\hat{\Pi}_i$  es un proyector de rango  $r$ , esta probabilidad depende únicamente de la componente de  $\rho$  contenida en el soporte de  $\hat{\Pi}_i$ . En particular, si el soporte de  $\rho$  no está contenido en el soporte de  $\hat{\Pi}_i$ , existe necesariamente una fracción del estado ortogonal a dicho subespacio, lo que impide que la probabilidad de identificación alcance la unidad. Por consiguiente, la relación geométrica entre los soportes de  $\rho$  y  $\hat{\Pi}_i$  puede imponer restricciones sobre la probabilidad máxima de identificación alcanzable dentro de una arquitectura de medición dada [26, 27].

### *E. Fidelidad entre estados cuánticos*

La *fidelidad* es una medida del grado de indistinguibilidad entre estados cuánticos. En el contexto de la discriminación de estados cuánticos, la fidelidad proporciona una medida cuantitativa de cuán difícil es distinguir dos estados. Estados con alta fidelidad presentan una mayor superposición efectiva y, en consecuencia, conducen a probabilidades de error más elevadas en esquemas de discriminación [4, 28, 29].

Dadas dos matrices de densidad  $\rho$  y  $\sigma$ , la fidelidad cuántica se define como

$$F(\rho, \sigma) \equiv \text{tr}\left(\sqrt{\rho^{1/2} \sigma \rho^{1/2}}\right). \quad (19)$$

Aunque esta definición no es manifiestamente simétrica en  $\rho$  y  $\sigma$ , puede demostrarse que la fidelidad satisface

$$F(\rho, \sigma) = F(\sigma, \rho), \quad (20)$$

y toma valores en el intervalo  $0 \leq F(\rho, \sigma) \leq 1$ . En particular,  $F(\rho, \sigma) = 1$  si y solo si  $\rho = \sigma$ , mientras que valores cercanos a cero indican estados casi perfectamente distinguibles [4, 28, 29].

Existen situaciones en las que la expresión de la fidelidad se simplifica considerablemente. En el caso en que uno de los estados sea puro,  $\rho = |\psi\rangle\langle\psi|$ , mientras que el otro es un estado arbitrario  $\sigma$ , la fidelidad adopta la forma

$$F(|\psi\rangle, \sigma) = \sqrt{\langle\psi|\sigma|\psi\rangle}. \quad (21)$$

Por lo tanto, la fidelidad corresponde a la raíz cuadrada de la probabilidad de obtener el resultado asociado al proyector  $|\psi\rangle\langle\psi|$  al realizar una medición sobre el estado  $\sigma$ .

### F. Teorema de Naimark

El teorema de Naimark proporciona el vínculo entre las mediciones generalizadas y las implementaciones experimentales. Este teorema establece que cualquier medición POVM puede realizarse físicamente mediante una medición proyectiva actuando sobre un espacio de Hilbert de mayor dimensión [30].

Específicamente, dada una POVM  $\{\hat{E}_m\}_{m=0}^{l-1}$  definida sobre el sistema objetivo, donde  $l$  denota el número de resultados posibles, es posible introducir un sistema auxiliar (o *ancilla*) y un operador unitario  $\hat{U}$  que actúa sobre el sistema conjunto (objetivo + auxiliar), de modo que una medición proyectiva sobre el sistema auxiliar reproduce exactamente las estadísticas de la POVM original [31, 32, 30].

Formalmente, sea  $|\psi\rangle_T$  el estado del sistema objetivo y  $|0\rangle_A$  el estado inicial del sistema auxiliar, el cual debe tener una dimensión de al menos  $l$ . El operador unitario  $\hat{U}$  que acopla ambos sistemas se construye de tal manera que su acción sobre el estado inicial es:

$$\hat{U} |\psi\rangle_T |0\rangle_A = \sum_{m=0}^{l-1} (\hat{K}_m |\psi\rangle_T) \otimes |m\rangle_A, \quad (22)$$

donde  $\{\hat{K}_m\}$  son los operadores de Kraus asociados a la POVM ( $\hat{E}_m = \hat{K}_m^\dagger \hat{K}_m$ ) y  $\{|m\rangle_A\}$  constituye una base ortonormal del espacio de Hilbert del sistema auxiliar [31].

Desde un punto de vista computacional, la implementación del sistema auxiliar implica el uso de qubits adicionales. Dado que un sistema de  $n$  qubits existe en un espacio de dimensión  $2^n$ , el número de qubits auxiliares  $n_A$  necesarios para representar  $l$  resultados posibles de una POVM está dado por:

$$n_A = \lceil \log_2(l) \rceil. \quad (23)$$

Por ejemplo, para un problema de clasificación con 4 clases (que requiere una POVM de 4 elementos), el circuito cuántico deberá incorporar al menos 2 qubits auxiliares [32].

Al realizar una medición proyectiva en la base computacional  $\{|m\rangle_A\}$  de los qubits auxiliares, la probabilidad de obtener el resultado  $m$  es:

$$P(m) = \left( \sum_n {}_T\langle \psi | \hat{K}_n^\dagger \otimes \langle n |_A \right) (I_T \otimes {}_A\langle m | \langle m |_A) \left( \sum_{m'} \hat{K}_{m'} |\psi\rangle_T \otimes |m'\rangle_A \right) \quad (24)$$

$$P(m) = {}_T\langle \psi | \hat{K}_m^\dagger \hat{K}_m |\psi\rangle_T \cdot \underbrace{{}_A\langle m | m \rangle_A}_1 \cdot \underbrace{{}_A\langle m | m \rangle_A}_1 \quad (25)$$

$$p(m) = {}_T\langle \psi | \hat{K}_m^\dagger \hat{K}_m |\psi\rangle_T = {}_T\langle \psi | \hat{E}_m |\psi\rangle_T. \quad (26)$$

Este resultado confirma que medir proyectivamente la ancilla tras la evolución unitaria  $\hat{U}$  es matemáticamente equivalente a medir la POVM  $\{\hat{E}_m\}$  en el sistema original. Esto proporciona la base teórica para diseñar circuitos cuánticos variacionales que aprendan a clasificar estados optimizando los parámetros del operador  $\hat{U}$  [31, 32, 30].

# CAPÍTULO III: ESTRATEGIAS DE DISCRIMINACIÓN DE ESTADOS

## *A. Discriminación de Estados Cuánticos*

La discriminación de estados cuánticos (*Quantum State Discrimination*, QSD) aborda el problema de diseñar una estrategia de medición óptima que permita identificar en qué estado  $\rho_i$  ha sido preparado un sistema, sabiendo que dicho estado pertenece a un conjunto conocido de estados no ortogonales  $\{\rho_i\}$  [1, 2, 33].

Formalmente, se considera una fuente que emite uno de varios estados posibles  $\rho_i$  (donde  $i = 1, \dots, m$ ) con una probabilidad a priori  $q_i$ . De esta forma, la fuente queda caracterizada por el ensamble cuántico:

$$\mathcal{E} = \{q_i, \rho_i\}_{i=1}^m, \quad \text{con} \quad \sum_{i=1}^m q_i = 1. \quad (27)$$

El problema puede interpretarse como un proceso de recuperación de información clásica codificada en soportes cuánticos. La variable aleatoria clásica  $i$  se codifica en el estado  $\rho_i$ . Para decodificar esta información, se realiza una medición POVM  $\{\hat{E}_k\}$  sobre el sistema y, basándose en el resultado  $k$ , se infiere el valor de  $i$ . Debido a que los estados no son ortogonales, no es posible distinguirlos perfectamente mediante una medición única. En consecuencia, el objetivo fundamental es maximizar una figura de mérito, usualmente la probabilidad de acierto o la información mutua [33].

Las dos estrategias principales para abordar este problema son la discriminación de error mínimo (MESD) y la discriminación inequívoca (USD).

1) *Discriminación de Error Mínimo*. La discriminación de error mínimo (*Minimum Error Discrimination*) es una estrategia determinista en la cual el observador siempre debe emitir una estimación sobre el estado del sistema. El objetivo es minimizar la probabilidad de error global o, equivalentemente, maximizar la probabilidad de éxito

$P_{\text{succ}}$ . En este esquema, la POVM óptima  $\{E_i\}$  debe maximizar:

$$P_{\text{succ}} = \sum_i q_i \text{Tr}(\hat{E}_i \rho_i). \quad (28)$$

Para el caso particular de discriminar entre dos estados  $\rho_0$  y  $\rho_1$  con probabilidades a priori  $q_0$  y  $q_1$ , la probabilidad máxima de éxito está acotada superiormente por el **límite de Helstrom**, dado por:

$$P_{\text{Helstrom}} = \frac{1}{2} + \frac{1}{2} \|q_0 \rho_0 - q_1 \rho_1\|_1, \quad (29)$$

donde  $\|A\|_1 = \text{Tr}(\sqrt{\hat{A}^\dagger \hat{A}})$  es la norma de traza, que se define como la suma de los valores absolutos de los autovalores de la matriz resultante. Alcanzar este límite teóricamente implica encontrar la medición óptima, la cual es, en general, difícil de determinar analíticamente para  $m > 2$  estados, justificando así el uso de métodos variacionales aproximados [1, 8, 4, 34].

2) *Discriminación Inequívoca*. La discriminación inequívoca (*Unambiguous Discrimination*) es una estrategia en la cual no se permite cometer errores de identificación. Para garantizar que, cuando se obtiene un resultado, este sea correcto, se introduce la posibilidad de un resultado inconcluso o incierto (?).

La POVM asociada contiene elementos  $\{\hat{E}_1, \dots, \hat{E}_m, \hat{E}_?\}$ , donde  $\hat{E}_i$  (con  $i \neq ?$ ) corresponde a la identificación certera del estado  $\rho_i$ . Esto impone la restricción:

$$\text{Tr}(\hat{E}_i \rho_j) = 0 \quad \text{para todo } i \neq j. \quad (30)$$

Si el sistema colapsa en el operador  $\hat{E}_?$ , el observador sabe que no ha podido identificar el estado, pero no ha cometido un error de clasificación falsa. El objetivo de esta estrategia es maximizar la probabilidad de obtener un resultado concluyente (minimizar la probabilidad de  $\hat{E}_?$ ). La probabilidad máxima de éxito depende intrínsecamente del solapamiento (overlap) entre los estados; si los estados son linealmente dependientes, la discriminación inequívoca perfecta es imposible [1, 8, 4].

### B. Optimización Convexa y Programación Semidefinida

La optimización convexa es una rama de la optimización matemática en la que tanto la función objetivo como el conjunto de restricciones definen un problema convexo [35]. Un problema se denomina convexo si el dominio factible es un conjunto convexo y la función objetivo es convexa sobre dicho dominio.

Un conjunto se dice convexo si, para cualesquiera dos puntos que pertenecen a él, el segmento de recta que los une permanece completamente contenido en el conjunto [36]. De manera análoga, una función es convexa si su valor en cualquier combinación convexa de dos puntos no excede la combinación convexa de los valores de la función en dichos puntos; en otras palabras, el gráfico de una función convexa siempre se encuentra por debajo del segmento que une dos puntos cualesquiera [35].

Estas propiedades estructurales aseguran que cualquier mínimo local de la función objetivo es también un mínimo global. Esta característica elimina la presencia de mínimos locales espurios y permite el uso de algoritmos numéricos capaces de encontrar soluciones globales de manera eficiente, constituyendo una de las principales ventajas de los problemas convexos y fundamentando la robustez de los métodos algorítmicos empleados para su resolución [35].

Dentro de este marco, la *programación cónica* corresponde a una familia general de problemas de optimización convexa en los cuales las variables de optimización están restringidas a pertenecer a un cono convexo. En términos generales, estos problemas consisten en optimizar una función lineal sujeta a restricciones lineales y a la condición de pertenencia a un cono. Esta formulación permite unificar bajo un mismo marco diversas clases de problemas convexos, tales como la programación lineal, la programación cónica de segundo orden y la programación semidefinida [35].

La *programación semidefinida* (SDP) es un caso particular de programación cónica en el cual la variable de optimización es una matriz hermítica positiva semidefinida. En este contexto, la restricción fundamental  $\hat{X} \succeq 0$  indica que la matriz  $X$  pertenece al cono de matrices semidefinidas positivas. Esta estructura resulta especialmente natural en información cuántica, donde tanto los estados cuánticos como los operadores de

medición se describen mediante operadores hermíticos positivos [33].

Un programa semidefinido típico consiste en optimizar una función objetivo lineal real de la forma  $\text{tr}(\hat{A}\hat{X})$ , donde  $A$  es un operador hermítico dado, sujeto a restricciones lineales de igualdad y desigualdad, además de la condición de semidefinitud positiva [33]. La formulación general de un problema de maximización puede escribirse como

$$\text{maximizar } \text{tr}(\hat{A}\hat{X}), \quad (31)$$

$$\text{sujeto a } \Phi_i(\hat{X}) = B_i, \quad i = 1, \dots, m, \quad (32)$$

$$\Gamma_j(\hat{X}) \preceq C_j, \quad j = 1, \dots, n, \quad (33)$$

donde  $\Phi_i$  y  $\Gamma_j$  son mapas lineales que preservan la hermiticidad, y  $B_i, C_j$  son operadores hermíticos.

Una consecuencia clave de la convexidad de los SDPs es que sus soluciones óptimas corresponden a óptimos globales del problema. Por esta razón, los valores obtenidos mediante programación semidefinida pueden interpretarse como cotas teóricas exactas, contra las cuales es posible comparar el desempeño de métodos heurísticos o variacionales, como los clasificadores cuánticos implementados en dispositivos NISQ [33, 28].

Esta herramienta es particularmente idónea para la discriminación de estados cuánticos, ya que las propiedades físicas de los estados ( $\rho \succeq 0$ ,  $\text{tr}(\rho) = 1$ ) y de las mediciones POVM ( $\sum_a \hat{M}_a = \mathbb{I}$ ,  $\hat{M}_a \succeq 0$ ) se traducen de manera natural en restricciones semidefinidas [33].

1) *Formulación para Discriminación de Mínimo Error.* En el problema de discriminación de mínimo error, el objetivo es maximizar la probabilidad promedio de acierto  $P_{\text{guess}}$ . Como se estableció previamente, esto equivale a encontrar la POVM ( $M = \{\hat{M}_a\}$ ) que maximiza la correlación entre el resultado de la medición y el estado preparado [33].

La optimización sobre todas las mediciones posibles se traduce directamente en el siguiente programa semidefinido, donde las variables a optimizar son los operadores

de la POVM  $\{\hat{M}_a\}$ :

$$P_{\text{guess}}^* = \text{maximizar} \quad \sum_i q(i) \text{tr}(\hat{M}_i \rho_i), \quad (34)$$

$$\text{sujeto a} \quad \hat{M}_a \succeq 0 \quad \forall a, \quad (35)$$

$$\sum_a \hat{M}_a = \mathbb{I}. \quad (36)$$

La solución a este problema ( $P_{\text{guess}}^*$ ) proporciona el límite teórico máximo alcanzable (límite de Helstrom generalizado para  $N$  estados), el cual servirá como cota superior para evaluar el desempeño del circuito variacional [33].

2) *Formulación para Discriminación Inequívoca.* Para la discriminación inequívoca, se busca maximizar la probabilidad de obtener un resultado concluyente, sujeto a la restricción estricta de no cometer errores. El conjunto de medición se expande a  $M = \{\hat{M}_1, \dots, \hat{M}_m, \hat{M}_\emptyset\}$ , donde  $\hat{M}_\emptyset$  corresponde al resultado inconcluso [33].

La condición de *no error* implica que la probabilidad de detectar el estado  $i$  dado que se preparó el estado  $j$  (con  $i \neq j$ ) debe ser nula. Matemáticamente, y considerando el peso estadístico de cada estado, esto se impone mediante:

$$q(i) \text{tr}(\hat{M}_a \rho_i) = 0, \quad \forall a \neq i. \quad (37)$$

El uso del peso  $q(i)$  asegura que la restricción sea consistente incluso si la probabilidad a priori de algún estado es nula. Incorporando estas condiciones, el SDP que maximiza la probabilidad de éxito inequívoco es:

$$P_{\text{USD}}^* = \text{maximizar} \quad \sum_{i=1}^m q(i) \text{tr}(\hat{M}_i \rho_i), \quad (38)$$

$$\text{sujeto a} \quad q(i) \text{tr}(\hat{M}_a \rho_i) = 0, \quad i, a = 1, \dots, m; a \neq i, \quad (39)$$

$$\sum_{a=1}^m \hat{M}_a = \mathbb{I}, \quad (40)$$

$$\hat{M}_a \succeq 0. \quad a = 1, \dots, m, \emptyset \quad (41)$$

Es importante notar que la discriminación inequívoca no siempre es posible. En particular, si el soporte de un estado  $\rho_i$  está contenido en el soporte de otro, o si el

conjunto de estados presenta una dependencia estructural fuerte (como ocurre típicamente para estados de rango completo), la solución óptima del problema de programación semidefinida puede ser trivial. En este caso, los operadores asociados a resultados concluyentes se anulan ( $\hat{M}_a^* = 0$  para todo  $a \neq \emptyset$ ) y el operador correspondiente al resultado inconcluso es la identidad ( $\hat{M}_\emptyset^* = \mathbb{I}$ ), lo que implica una probabilidad de éxito nula,  $P_{\text{USD}}^* = 0$  [33].

Desde un punto de vista teórico, la discriminación inequívoca impone restricciones más estrictas que la discriminación de error mínimo. En efecto, la condición de ausencia total de errores exige que

$$\text{tr}(\hat{M}_a \rho_i) = 0 \quad \text{para } a \neq i,$$

lo cual, dado que  $\rho_i \succeq 0$ , implica necesariamente la ortogonalidad operatorial  $\hat{M}_a \rho_i = 0$  [33].

Si un estado  $\rho_i$  es de **rango completo**, su soporte abarca todo el espacio de Hilbert. Esto significa que no existe ningún subespacio en el que el estado sea nulo; en otras palabras, todos los vectores del espacio tienen componente en  $\rho_i$ . En consecuencia, el único operador positivo que satisface la condición necesaria para discriminarlo inequívocamente es el operador nulo. Por ello, si el conjunto a discriminar contiene dos o más estados de rango completo, no es posible construir una POVM que los discrimine inequívocamente [33].

---

# CAPÍTULO IV: ALGORITMOS CUÁNTICOS VARIACIONALES Y DISEÑO DEL CLASIFICADOR

## *A. Algoritmos Cuánticos Variacionales*

Los *Algoritmos Cuánticos Variacionales* (VQAs) constituyen una familia de algoritmos diseñados para aprovechar de manera óptima los dispositivos NISQ actuales mediante una estrategia de separación funcional entre recursos cuánticos y clásicos. En este esquema híbrido, el procesador cuántico (QPU, del inglés Quantum Processing Unit) se emplea exclusivamente para la preparación de estados y la estimación de valores esperados, tareas intrínsecamente costosas para un ordenador clásico. Por su parte, un procesador clásico (CPU, del inglés Central Processing Unit) se encarga de optimizar los parámetros del circuito. Como resultado, la evolución cuántica queda completamente determinada por un conjunto de parámetros ajustables, los cuales son actualizados iterativamente a partir de una función de costo definida a nivel clásico [23, 37, 38].

Este modelo funciona mediante un ciclo de retroalimentación que itera constantemente, tal como se ilustra en la Fig. 2. En cada iteración, la QPU ejecuta el circuito y realiza las mediciones necesarias para estimar los observables de interés. Los resultados estadísticos se transfieren a la CPU, que evalúa la función de costo y, mediante un algoritmo de optimización, determina una nueva configuración de parámetros. Estos parámetros actualizados se envían nuevamente a la QPU, repitiéndose el ciclo hasta alcanzar un criterio de convergencia previamente establecido.

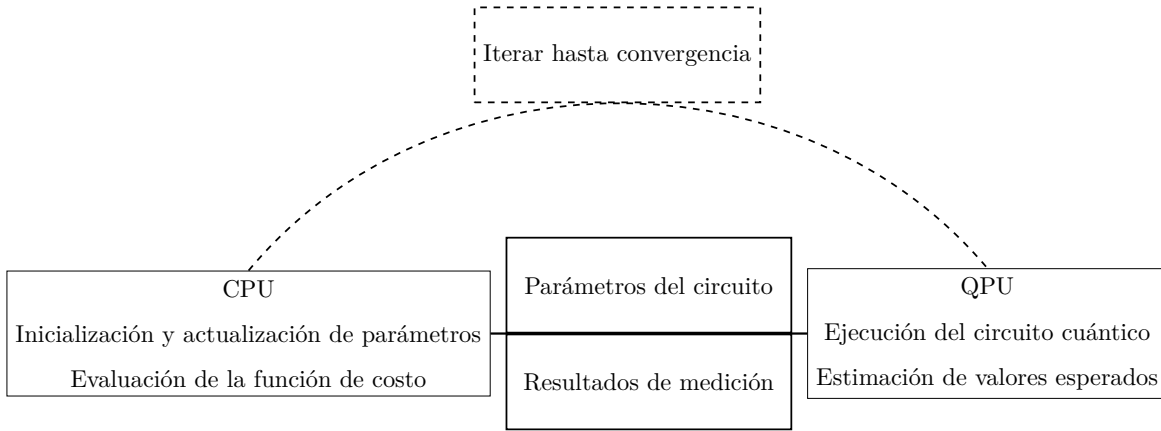


Figura 2: Esquema del ciclo de retroalimentación clásico-cuántico en un Algoritmo Cuántico Variacional (VQA). La CPU inicializa y actualiza los parámetros del circuito mediante la evaluación de una función de costo, mientras que la QPU ejecuta el circuito parametrizado y estima los valores esperados asociados a las mediciones. Los resultados obtenidos se devuelven a la CPU para la optimización iterativa, proceso que se repite hasta alcanzar la convergencia.

Los algoritmos cuánticos variacionales se componen principalmente de tres elementos: un *ansatz* parametrizado, que define la estructura del circuito y contiene los parámetros entrenables; una función de costo, que evalúa el desempeño del circuito; y un bucle de optimización clásico, que actualiza los parámetros según los resultados obtenidos [37, 38]. Cada uno de estos elementos será detallado a continuación.

1) *Preparación de Estados y Ansatz Variacional.* El componente central del algoritmo es el *Circuito Cuántico Variacional* (VQC), también denominado *ansatz*. Este consiste en un circuito cuántico parametrizado por un conjunto de variables reales  $\theta$ , las cuales definen rotaciones unitarias entrenables. La variación de estos parámetros permite explorar una subregión del espacio de Hilbert con el objetivo de optimizar una función de costo asociada al problema de interés [37, 38].

El proceso físico comienza con la preparación del estado de entrada. El registro cuántico se inicializa en el estado de referencia  $|0\rangle$ , el cual es transformado mediante una operación de codificación  $U_R$  que implementa la carga del dato cuántico. En general, esta codificación prepara un estado  $\rho$  que representa el estado a discriminar. En el caso

de estados puros, esta transformación puede escribirse como

$$|0\rangle \xrightarrow{\hat{U}_R} |\rho\rangle, \quad (42)$$

mientras que, de forma más general, a nivel de operadores densidad se tiene

$$\rho = \hat{U}_R |0\rangle\langle 0| \hat{U}_R^\dagger. \quad (43)$$

Posteriormente, el estado codificado es procesado por el circuito variacional parametrizado  $\hat{U}_V(\boldsymbol{\theta})$ , dando lugar al estado de salida

$$\rho(\boldsymbol{\theta}) = \hat{U}_V(\boldsymbol{\theta}) \rho \hat{U}_V^\dagger(\boldsymbol{\theta}), \quad (44)$$

o, en el caso puro,

$$|\psi(\boldsymbol{\theta})\rangle = \hat{U}_V(\boldsymbol{\theta}) |\rho\rangle. \quad (45)$$

El operador global del circuito puede definirse como

$$\hat{U}_A(\boldsymbol{\theta}) := \hat{U}_V(\boldsymbol{\theta}) \hat{U}_R, \quad (46)$$

el cual genera el estado candidato cuya medición proporciona la información necesaria para evaluar la función de costo [23].

Desde un punto de vista estructural, el ansatz variacional  $U_V(\boldsymbol{\theta})$  se construye como una secuencia de capas unitarias:

$$\hat{U}_V(\boldsymbol{\theta}) = \hat{U}_L(\boldsymbol{\theta}_L) \cdots \hat{U}_2(\boldsymbol{\theta}_2) \hat{U}_1(\boldsymbol{\theta}_1), \quad (47)$$

donde cada bloque  $U_\ell(\boldsymbol{\theta}_\ell)$  depende de un subconjunto de parámetros entrenables. Típicamente, estas capas están formadas por rotaciones unitarias generadas por operadores hermíticos  $\hat{H}_m$  (por ejemplo, matrices de Pauli), intercaladas con compuertas fijas de entrelazamiento:

$$\hat{U}_\ell(\boldsymbol{\theta}_\ell) = \prod_m e^{-i\theta_{\ell m} \hat{H}_m} W_m, \quad (48)$$

donde  $W_m$  representa compuertas no parametrizadas responsables de generar correlaciones cuánticas entre los qubits [37].

En este trabajo, el ansatz implementado es lo suficientemente flexible como para aproximar mediciones óptimas de tipo POVM asociadas a estrategias de discriminación de estados cuánticos no ortogonales, manteniendo al mismo tiempo una profundidad compatible con las limitaciones de coherencia y ruido propias del régimen NISQ.

Dentro de este ciclo híbrido, el circuito cuántico parametrizado puede interpretarse de manera análoga a una *red neuronal cuántica*. La evolución unitaria de los estados cuánticos desempeña un rol similar al de capas de transformación lineal con parámetros entrenables, mientras que el proceso de medición y el post-procesamiento clásico de los resultados introducen una no linealidad efectiva. Esta no linealidad es indispensable para tareas de clasificación o discriminación de estados cuánticos. De este modo, los circuitos pueden entrenarse mediante algoritmos híbridos clásico-cuánticos, optimizando sus parámetros con respecto a una función de costo, y evaluarse posteriormente en configuraciones de parámetros no utilizadas durante el entrenamiento, permitiendo cuantificar su desempeño generalizado [24, 39].

2) *Función de costo en VAQs*. En los algoritmos cuánticos variacionales, la función de costo  $J(\boldsymbol{\theta})$  es una cantidad real asociada al estado de salida del circuito cuántico parametrizado, cuya optimización determina la solución del problema físico considerado. En la práctica, dicha función se define a partir de valores esperados de observables medidos sobre el estado preparado por el ansatz variacional [23, 38].

Sea  $\{\rho_k\}$  un conjunto de estados de entrada y  $U(\boldsymbol{\theta})$  el operador unitario parametrizado que define el circuito variacional. Para cada estado  $\rho_k$ , el estado de salida viene dado por

$$\rho_k(\boldsymbol{\theta}) = \hat{U}(\boldsymbol{\theta}) \rho_k \hat{U}^\dagger(\boldsymbol{\theta}), \quad (49)$$

y los valores esperados de observables hermíticos  $\{\hat{O}_k\}$  se obtienen como

$$\langle \hat{O}_k \rangle_{\boldsymbol{\theta}} = \text{Tr}[\hat{O}_k \rho_k(\boldsymbol{\theta})]. \quad (50)$$

La función de costo se construye entonces como una combinación de estos valores

esperados,

$$J(\boldsymbol{\theta}) = \sum_k f_k(\langle \hat{O}_k \rangle_{\boldsymbol{\theta}}), \quad (51)$$

donde las funciones  $f_k(\cdot)$  dependen del problema específico que se desea resolver.

En problemas de discriminación de estados cuánticos, la función de costo se elige de modo que cuantifique el desempeño del esquema de medición implementado por el circuito. En particular, el circuito variacional seguido de una medición proyectiva final induce una medida POVM efectiva  $\{E_m(\boldsymbol{\theta})\}$ , definida por

$$E_m(\boldsymbol{\theta}) = \hat{U}^\dagger(\boldsymbol{\theta}) \hat{\Pi}_m \hat{U}(\boldsymbol{\theta}), \quad (52)$$

donde  $\{\Pi_m\}$  son los proyectores asociados a la medición final [37].

Cabe destacar que, en los problemas de discriminación cuántica, la función de costo no se define a partir de etiquetas clásicas conocidas, sino directamente en términos de probabilidades de medición y cantidades físicas bien definidas [38].

3) *Bucles de Optimización y Entrenamiento de Parámetros.* Una vez definidos el *ansatz* variacional y la función de costo  $J(\boldsymbol{\theta})$ , el siguiente paso consiste en la optimización de los parámetros entrenables. En este trabajo se emplean optimizadores clásicos locales basados en gradientes, los cuales ajustan iterativamente los parámetros a partir de la información proporcionada por la función de costo [23, 38].

En este enfoque, la actualización de los parámetros se realiza siguiendo la dirección de descenso del gradiente,

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \eta \nabla J(\boldsymbol{\theta}_n), \quad (53)$$

donde  $\eta$  denota la tasa de aprendizaje. En particular, se utiliza el optimizador **Adam**, el cual combina estimaciones adaptativas del primer y segundo momento del gradiente, proporcionando una mayor robustez frente al ruido estadístico asociado a la estimación de valores esperados mediante un número finito de mediciones [24, 40].

Un desafío fundamental en el entrenamiento de algoritmos cuánticos variacionales es la aparición de **mesetas áridas** (*barren plateaus*), es decir, regiones del paisaje de

optimización en las que la norma del gradiente decrece exponencialmente con el número de qubits, dificultando la convergencia del algoritmo [23].

### B. Implementación de POVMs mediante VQCs

Como se mencionó en el Capítulo II, el teorema de Naimark garantiza que cualquier medida generalizada (POVM) puede implementarse como una medición proyectiva en un espacio de Hilbert extendido, introduciendo grados de libertad auxiliares y una evolución unitaria apropiada [30, 31]. En el contexto de los algoritmos cuánticos variacionales, esta idea se materializa mediante la construcción de un circuito unitario  $\hat{U}_A(\boldsymbol{\theta})$  que actúa conjuntamente sobre el sistema objetivo y un registro ancilla, seguido de una medición proyectiva sobre los qubits auxiliares.

Cada posible resultado de la medición del ancilla induce un operador de Kraus sobre el sistema, lo que permite implementar una POVM arbitraria. En particular, si el sistema objetivo se encuentra inicialmente en el estado  $\rho$  y el registro auxiliar en el estado  $|0\rangle_A$ , la acción del operador unitario global seguida de la medición proyectiva del ancilla en la base computacional produce un conjunto de operadores de Kraus  $\{\hat{K}_m\}$  tales que

$$\hat{K}_m = {}_A\langle m | \hat{U}_A | 0 \rangle_A, \quad (54)$$

donde  $|m\rangle_A$  denota un estado base del registro auxiliar.

En la práctica, la medición se realiza sobre  $n_A$  qubits auxiliares en la base computacional, produciendo una cadena binaria

$$\mathbf{z} = z_1 z_2 \cdots z_{n_A} \in \{0, 1\}^{n_A}. \quad (55)$$

Cada cadena binaria  $\mathbf{z}$  se asocia de manera unívoca con un índice decimal

$$m \in \{0, \dots, 2^{n_A} - 1\}, \quad (56)$$

el cual etiqueta el resultado de la POVM. De este modo, la probabilidad de obtener el resultado  $m$  al medir el estado  $\rho$  viene dada por

$$p(m|\rho) = \text{Tr}[\hat{K}_m \rho \hat{K}_m^\dagger] = \text{Tr}[\hat{E}_m \rho], \quad (57)$$

La implementación explícita de estas medidas requiere descomponer el operador unitario global en compuertas elementales. Existen distintos esquemas de descomposición para circuitos multiqubit, en particular, la CSD permite expresar el operador unitario como una secuencia de compuertas controladas que actúan de forma jerárquica sobre el sistema y los qubits auxiliares. Desde una perspectiva operacional, estas compuertas controladas generan factores diagonales que, condicionados al estado del ancilla, se traducen en operadores de Kraus de rango bajo [30, 31, 41].

Esta arquitectura en cascada, donde cada etapa entrelaza el sistema objetivo con un nuevo qubit auxiliar, se ilustra esquemáticamente en la Fig. 3.

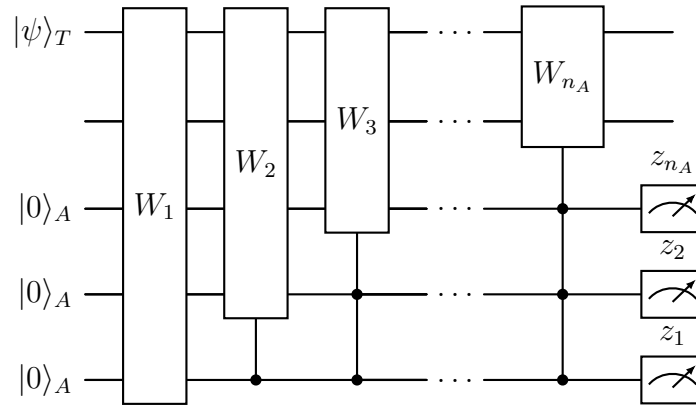


Figura 3: Esquema de la implementación unitaria de una medida generalizada (POVM) mediante dilatación en un espacio extendido. El estado objetivo  $|\psi\rangle_T$  se acopla secuencialmente a un conjunto de qubits auxiliares inicializados en  $|0\rangle_A$  a través de una arquitectura jerárquica de operadores unitarios  $W_i$ . Esta estructura en cascada genera, tras la medición proyectiva de los ancillas, una descomposición efectiva en operadores de Kraus asociados a los distintos resultados de medición.

En este esquema, cada bloque unitario controlado contribuye a la implementación efectiva de los operadores de Kraus, los cuales emergen tras la medición de los qubits auxiliares. Las POVMs más generales pueden obtenerse incrementando el número de ancillas y concatenando capas controladas adicionales [31]. No obstante, este procedimiento implica un aumento significativo en la complejidad del circuito, particularmente en el número de compuertas de dos qubits requeridas, que en el caso más general crece exponencialmente con el número de qubits del sistema [41].

Para mitigar este problema en dispositivos NISQ, resulta conveniente restringir la

---

estructura del operador unitario global mediante un *ansatz* variacional eficiente desde el punto de vista del hardware [38].

Para más detalles sobre la descomposición explícita del operador unitario global en compuertas elementales, incluyendo la estructura jerárquica asociada a la descomposición coseno-seno, se remite al lector a Refs. [30, 31, 41].

### *C. Alcances y limitaciones de los VQAs*

Los algoritmos cuánticos variacionales presentan varias ventajas que los convierten en candidatos naturales para su implementación en dispositivos cuánticos ruidosos de escala intermedia (NISQ). Entre sus principales fortalezas se encuentran su versatilidad, al permitir adaptar tanto la estructura del *ansatz* como la función de costo al problema de interés; su resiliencia parcial al ruido, debido a la posibilidad de absorber efectos imperfecciones mediante el ajuste variacional de parámetros; y su eficiencia en el uso de recursos, al requerir profundidades de circuito compatibles con el hardware cuántico actual [18, 38].

No obstante, estas ventajas vienen acompañadas de limitaciones importantes. En particular, la optimización de los parámetros puede verse severamente afectada por la aparición de mesetas áridas en el paisaje de la función de costo, así como por la dificultad de analizar rigurosamente la complejidad y escalabilidad de estos algoritmos. En consecuencia, la obtención de una ventaja cuántica demostrable mediante VQAs permanece como una cuestión abierta [18, 38].

# CAPÍTULO V: DEFINICIÓN DE LAS FAMILIAS DE ESTADOS CUÁNTICOS Y DISEÑO EXPERIMENTAL

## A. Definición de las Familias de Estados Cuánticos

En este capítulo se describen los conjuntos de estados cuánticos utilizados para evaluar el desempeño del clasificador variacional. Para ello, se consideran dos tipos de familias de estados con roles metodológicos diferentes. En primer lugar, se introducen *familias de validación*, tomadas del trabajo de Chen *et al.* [17], cuyo propósito es validar el correcto funcionamiento y la estabilidad del clasificador bajo condiciones controladas. En segundo lugar, se analizan *familias de generalización*, que presentan distintas estructuras de correlación y niveles de entrelazamiento, y que permiten evaluar la capacidad del modelo para generalizar más allá de las distribuciones empleadas durante el entrenamiento.

1) *Familias de Validación.* Para evaluar el desempeño del clasificador, se definen dos familias de estados cuánticos en un sistema de dos qubits (dimensión  $d = 4$ ). Estos estados se generan a partir de parámetros escalares  $a$  y  $b$ , los cuales siguen distribuciones de probabilidad específicas. Las familias se definen de la siguiente manera:

- **Familia 1** ( $\rho_1(a)$ ): Corresponde a una familia de **estados puros**, parametrizados por  $a \in [0, 1]$ . El vector de estado se define como:

$$|\psi_1(a)\rangle = \left(\sqrt{1-a^2}, 0, a, 0\right)^T, \quad (58)$$

cuyo operador densidad asociado es:

$$\rho_1(a) = |\psi_1(a)\rangle \langle \psi_1(a)|. \quad (59)$$

- **Familia 2** ( $\rho_2(b)$ ): Corresponde a una familia de **estados mixtos**, construida como una mezcla estadística equiprobable de dos vectores de estado ortogonales (para  $b \neq$

$0, 1)$ :

$$|\psi_2(b)\rangle = (0, \sqrt{1-b^2}, b, 0)^T, \quad (60)$$

$$|\psi_3(b)\rangle = (0, -\sqrt{1-b^2}, b, 0)^T. \quad (61)$$

El estado mixto resultante viene dado por:

$$\rho_2(b) = \frac{1}{2} |\psi_2(b)\rangle \langle \psi_2(b)| + \frac{1}{2} |\psi_3(b)\rangle \langle \psi_3(b)|. \quad (62)$$

El objetivo es clasificar un estado cuántico desconocido  $\rho_x$  perteneciente a una de estas familias. Para ello, se generaron conjuntos de datos variando los parámetros  $a$  y  $b$  según dos tipos de distribuciones de probabilidad:

1. **Distribución Uniforme ( $\mathcal{U}$ ):** El parámetro se elige uniformemente en el intervalo  $[0, 1]$ .
2. **Distribución Normal ( $\mathcal{N}$ ):** El parámetro se elige siguiendo una distribución normal con desviación estándar  $\sigma = 0.05$  y media  $\mu$ , truncada al intervalo  $[0, 1]$ .

En la Tabla 1 se resumen los cuatro escenarios de clasificación analizados en la primera etapa de este trabajo.

Cuadro 1: Resumen de los cuatro escenarios de clasificación analizados. En cada caso se indica la distribución de probabilidad utilizada para generar los parámetros  $a$  y  $b$  que definen las familias de estados  $\rho_1(a)$  y  $\rho_2(b)$ . Las distribuciones normales  $\mathcal{N}$  tienen desviación estándar  $\sigma = 0.05$ , media  $\mu$  indicada en la tabla y están truncadas al intervalo  $[0, 1]$ .

|               | $\rho_1(a)$               | $\rho_2(b)$                     |
|---------------|---------------------------|---------------------------------|
| <b>Caso 1</b> | $\mathcal{U}[0, 1]$       | $\mathcal{U}[0, 1]$             |
| <b>Caso 2</b> | $\mathcal{N}(\mu = 0.25)$ | $\mathcal{N}(\mu = 1/\sqrt{2})$ |
| <b>Caso 3</b> | $\mathcal{N}(\mu = 0.25)$ | $\mathcal{U}[0, 1]$             |
| <b>Caso 4</b> | $\mathcal{U}[0, 1]$       | $\mathcal{N}(\mu = 1/\sqrt{2})$ |

2) *Familias de Generalización.* Para evaluar la capacidad de generalización del clasificador variacional, se consideraron tres familias de estados cuánticos definidos en un sistema bipartito de dos qubits ( $d = 4$ ). Estas familias presentan estructuras de correlación y niveles de entrelazamiento distintos, lo que permite analizar el desempeño del clasificador frente a distribuciones de estados con características físicas diversas.

- **Familia EJM (Elegant Joint Measurements):**

La familia de mediciones conjuntas elegantes (EJM) se construye a partir de una base de estados puros en el espacio de Hilbert de dos qubits, cuya estructura geométrica está asociada a un tetraedro regular en la esfera de Bloch.

Se introducen cuatro vectores unitarios  $\{\vec{\mathbf{m}}_b\}_{b=1}^4$  que apuntan hacia los vértices de dicho tetraedro, como se muestra en la Fig. 4, definidos como

$$\vec{\mathbf{m}}_1 = (+1, +1, +1), \quad (63)$$

$$\vec{\mathbf{m}}_2 = (+1, -1, -1), \quad (64)$$

$$\vec{\mathbf{m}}_3 = (-1, +1, -1), \quad (65)$$

$$\vec{\mathbf{m}}_4 = (-1, -1, +1), \quad (66)$$

normalizados de modo que  $|\vec{\mathbf{m}}_b| = 1$ . A cada vector se asocia su dirección antipodal  $-\vec{\mathbf{m}}_b$ .

Cada vector  $\vec{\mathbf{m}}_b$  puede parametrizarse en coordenadas cilíndricas de la esfera de Bloch como

$$\vec{\mathbf{m}}_b = \sqrt{3} \left( \sqrt{1 - n_b^2} \cos \varphi_b, \sqrt{1 - n_b^2} \sin \varphi_b, n_b \right), \quad (67)$$

donde  $n_b$  y  $\varphi_b$  corresponden a la componente longitudinal y al ángulo azimutal, respectivamente.

A partir de estos vectores se definen los estados puros de un qubit

$$|\pm \vec{\mathbf{m}}_b\rangle = \sqrt{\frac{1 \pm n_b}{2}} e^{-i\varphi_b/2} |0\rangle + \sqrt{\frac{1 \mp n_b}{2}} e^{i\varphi_b/2} |1\rangle, \quad (68)$$

que corresponden a puntos diametralmente opuestos en la esfera de Bloch.

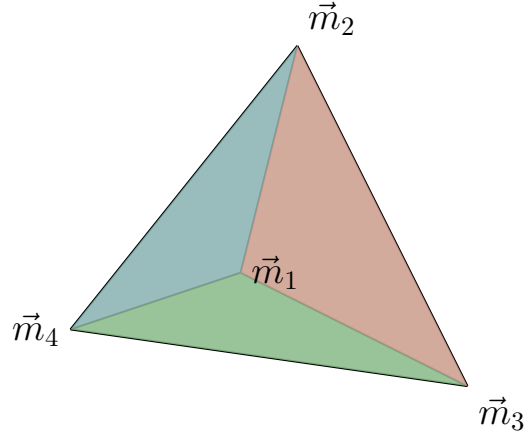


Figura 4: Representación geométrica en la esfera de Bloch de los cuatro vectores unitarios  $\vec{\mathbf{m}}_b = \frac{1}{\sqrt{3}}(\pm 1, \pm 1, \pm 1)$  asociados a la familia Elegant Joint Measurements (EJM). Estos vectores apuntan hacia los vértices de un tetraedro regular y determinan las direcciones locales de los estados reducidos obtenidos al trazar uno de los qubits en la base bipartita  $\{|\Phi_b^\theta\rangle\}$ . La simetría tetraédrica garantiza que los cuatro estados presentan el mismo grado de entrelazamiento y una distribución uniforme en el espacio de Bloch.

La familia de estados EJM se obtiene como una interpolación continua entre distintas bases conjuntas entrelazadas, parametrizada por un ángulo  $\theta \in [0, \pi/2]$ . Los estados bipartitos se definen como

$$|\Phi_b^\theta\rangle = \frac{\sqrt{3} + e^{i\theta}}{2\sqrt{2}} |\vec{\mathbf{m}}_b, -\vec{\mathbf{m}}_b\rangle + \frac{\sqrt{3} - e^{i\theta}}{2\sqrt{2}} |-\vec{\mathbf{m}}_b, \vec{\mathbf{m}}_b\rangle, \quad (69)$$

para  $b = 1, \dots, 4$ .

Todos los estados  $|\Phi_b^\theta\rangle$  presentan el mismo grado de entrelazamiento y, al trazar uno de los subsistemas, los estados locales resultantes forman un tetraedro regular dentro de la esfera de Bloch. En los límites  $\theta = 0$  y  $\theta = \pi/2$ , se recuperan respectivamente la medición EJM original y una base equivalente a la medición de Bell [42, 43, 44, 45].

- **Familia de Estados de Popescu:**

Introducida por Massar y Popescu, esta familia se construye utilizando los mismos vectores  $\vec{\mathbf{m}}_b$  definidos en la Ec. 68, combinados con el estado singlete. A diferencia de las EJM, donde los estados bipartitos se obtienen como superposiciones simétricas de pares localmente opuestos, en este caso los estados surgen de la combinación entre estados alineados y una componente entrelazada.

$$|\psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle). \quad (70)$$

Los estados se definen como

$$|\Psi_1\rangle = \frac{\sqrt{3}}{2} |\vec{\mathbf{m}}_1, \vec{\mathbf{m}}_1\rangle + \frac{1}{2} |\psi^-\rangle, \quad (71)$$

$$|\Psi_2\rangle = \frac{\sqrt{3}}{2} |\vec{\mathbf{m}}_2, \vec{\mathbf{m}}_2\rangle - \frac{1}{2} |\psi^-\rangle, \quad (72)$$

$$|\Psi_3\rangle = \frac{\sqrt{3}}{2} |\vec{\mathbf{m}}_3, \vec{\mathbf{m}}_3\rangle - \frac{1}{2} |\psi^-\rangle, \quad (73)$$

$$|\Psi_4\rangle = \frac{\sqrt{3}}{2} |\vec{\mathbf{m}}_4, \vec{\mathbf{m}}_4\rangle + \frac{1}{2} |\psi^-\rangle, \quad (74)$$

donde los vectores  $\vec{\mathbf{m}}_b$  definen direcciones tetraédricas en la esfera de Bloch. Esta familia posee una alta simetría y combina de manera no trivial correlaciones clásicas y cuánticas [46].

- **Familia de Estados de Werner:**

Los estados de Werner corresponden a una familia de estados mixtos bipartitos definidos como una combinación convexa entre el estado singlete puro y el estado completamente mezclado,

$$\rho_i = V_i |\psi^-\rangle \langle \psi^-| + \frac{1 - V_i}{4} \mathbb{I}, \quad (75)$$

donde  $V_i \in [0, 1]$  es el parámetro de visibilidad que cuantifica la fracción del estado singlete presente en la preparación experimental. Físicamente,  $V_i$  mide el grado de pureza y entrelazamiento del estado, mientras que el término proporcional a la identidad modela ruido blanco y efectos de decoherencia [43, 47].

## B. Generación de Datos

1) *Familias de Validación.* Los datos de entrenamiento y testeo se generaron de manera aleatoria a partir de dos parámetros reales  $a$  y  $b$ , asociados respectivamente a  $\rho_1$  y a  $\rho_2$ . Para cada conjunto de datos de tamaño total  $N$ , los parámetros se muestrearon de forma independiente a partir de las cuatro distribuciones consideradas, las cuales se detallan en la Tabla 1. En todos los casos se mantuvo una proporción fija entre ambas familias:  $\frac{1}{3}N$  estados corresponden a  $\rho_1$  y los  $\frac{2}{3}N$  restantes a  $\rho_2$ . A partir de estos parámetros se construyeron los estados cuánticos junto con sus etiquetas correspondientes.

---

El clasificador se entrenó utilizando conjuntos de datos de tamaño  $N = 10, 20, 50, 100, 200, 400, 600$  y  $1000$ . En cada caso, el conjunto total se dividió en subconjuntos de entrenamiento y validación, preservando la proporción original entre ambas familias. Para evaluar el desempeño del modelo, se utilizó un conjunto de prueba fijo compuesto por  $10\,000$  estados, generados con la misma proporción  $\frac{1}{3}$  a  $\frac{2}{3}$ , lo que permite una comparación consistente de los resultados obtenidos para distintos tamaños del conjunto de entrenamiento.

Con el fin de garantizar la reproducibilidad de los experimentos, se fijaron explícitamente semillas para los generadores de números aleatorios empleados en la generación de los datos. Este procedimiento se aplicó de manera análoga para cada una de las distribuciones consideradas en este trabajo.

2) *Familias de Generalización.* Para evaluar la capacidad de generalización del clasificador variacional, se realizaron distintos experimentos empleando las familias de estados EJM, Popescu y Werner. En una primera etapa, se estudió el comportamiento del clasificador al considerar pares de familias de estados. En particular, se analizaron los siguientes escenarios de clasificación: EJM–Popescu, EJM–Werner y Popescu–Werner. En cada caso, se generaron conjuntos de prueba compuestos por  $10\,000$  estados cuánticos, distribuidos equitativamente entre ambas familias ( $5\,000$  estados por familia). La generación de los datos se realizó de forma aleatoria a partir de los parámetros característicos de cada familia.

Para el entrenamiento del clasificador, se utilizaron conjuntos de datos de tamaño total  $N = 10, 20, 100, 600$  y  $1000$ . En todos los casos, los datos se distribuyeron de manera equitativa entre las dos familias consideradas, de modo que a cada una le correspondió la mitad de los estados. Cada conjunto se dividió a su vez en subconjuntos de entrenamiento y validación, utilizando una proporción del  $80\%$  para entrenamiento y  $20\%$  para validación.

En una segunda etapa, se consideró el escenario en el que el clasificador debe discriminar simultáneamente entre las tres familias de estados. En este caso, el conjunto

de prueba estuvo compuesto por 10 000 estados, distribuidos uniformemente entre las familias EJM, Popescu y Werner, de modo que a cada una le correspondió un tercio del total. De manera análoga, los conjuntos de entrenamiento y validación se generaron de forma aleatoria, manteniendo la misma proporción entre las tres familias, con una división del 80 % y 20 %, respectivamente.

### C. Implementación del Clasificador

La arquitectura del *ansatz* implementada en este trabajo es la misma utilizada en [17] y [19], y se muestra en las Fig. 5 y 6. Consta de un sistema de 4 qubits, dos qubits auxiliares (ancillas), inicializados en  $|0\rangle$ , y dos qubits que portan el estado de entrada  $|\psi_{in}\rangle$ .

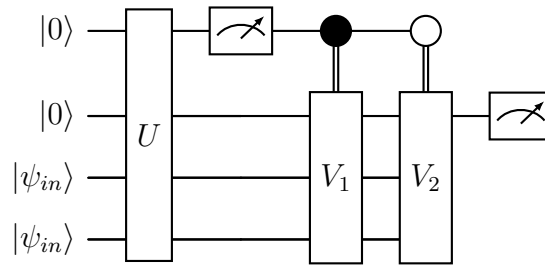


Figura 5: Arquitectura del clasificador cuántico variacional utilizada en este trabajo. El estado de entrada  $|\psi_{in}\rangle$  se codifica en los dos qubits inferiores, mientras que dos qubits auxiliares inicializados en  $|0\rangle$  permiten extender el espacio de Hilbert. El bloque unitario global  $U$  actúa sobre los cuatro qubits, seguido de la medición del primer qubit, que introduce una capa no lineal efectiva (tipo *dropout*) mediante control clásico condicional. Dependiendo del resultado de esta medición, se aplican los subcircuitos variacionales  $V_1$  o  $V_2$ , cuyos resultados determinan la salida del clasificador. El punto negro (blanco) indica activación condicionada al resultado  $|1\rangle$  ( $|0\rangle$ ).

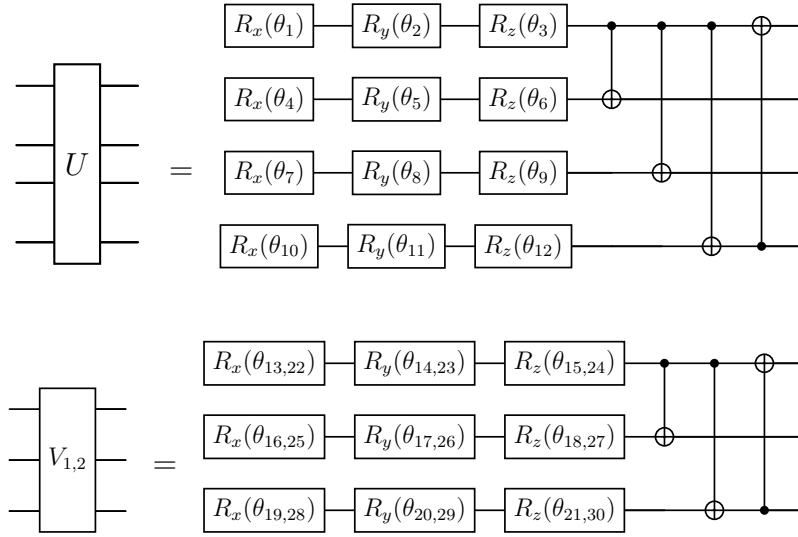


Figura 6: Descomposición explícita del ansatz variacional. Arriba: el bloque unitario global  $U$  de cuatro qubits se implementa mediante rotaciones parametrizadas  $R_x$ ,  $R_y$ ,  $R_z$  en cada qubit, seguidas de una capa de compuertas controladas que introducen entrelazamiento. Abajo: los bloques variacionales  $V_1$  y  $V_2$  de tres qubits poseen una estructura análoga, con parámetros independientes que se optimizan durante el entrenamiento. Los ángulos  $\theta_i$  constituyen los parámetros libres del modelo y determinan la clase de mediciones realizables por el clasificador.

El circuito implementa una dinámica controlada por mediciones intermedias. Inicialmente, los qubits son procesados mediante una compuerta unitaria  $U$ , seguida de compuertas condicionales  $V_1$  y  $V_2$ , cuya aplicación depende de los resultados de mediciones clásicas realizadas sobre qubits de control. En el diagrama del circuito, los puntos negros indican controles activados cuando el qubit se encuentra en el estado  $|1\rangle$ , mientras que los puntos blancos corresponden a controles activados en el estado  $|0\rangle$ . Los resultados de las mediciones se transmiten a través de canales clásicos, lo que permite la retroalimentación clásica necesaria para ejecutar las operaciones siguientes.

Cada bloque unitario de un solo qubit se parametriza mediante una descomposición de Euler de la forma

$$U = R_z(\theta_1) R_y(\theta_2) R_z(\theta_3), \quad (76)$$

donde los ángulos  $\theta_i$  constituyen los parámetros variacionales del circuito. La arquitectura alterna capas de rotaciones de un solo qubit con capas de entrelazamiento implementadas mediante compuertas CNOT.

Tras aplicar el bloque  $U$ , se mide el qubit 1; dependiendo del resultado de la medición, se aplica condicionalmente el bloque  $V_1$  si se obtiene el estado  $|1\rangle$  o el bloque  $V_2$  si se obtiene el estado  $|0\rangle$ . Finalmente, se realiza una medición sobre el qubit 2.

La salida del clasificador es un vector de probabilidad sobre la base computacional de los qubits de medida,  $\vec{P} = [P_{00}, P_{01}, P_{10}, P_{11}]$ . La regla de decisión para la clasificación es la siguiente:

- **Familia 1:** Se asigna si el resultado más probable es  $|00\rangle$  o  $|10\rangle$ .
- **Familia 2:** Se asigna si el resultado más probable es  $|01\rangle$ .
- **Inconcluso:** Se asigna si el resultado más probable es  $|11\rangle$ .

La asignación de resultados descrita anteriormente fue diseñada para implementar un esquema de discriminación inequívoca. En este régimen, la presencia de un resultado explícitamente inconcluso (asociado al estado  $|11\rangle$ ) permite al clasificador descartar aquellas realizaciones del estado de entrada cuya identificación no puede realizarse sin ambigüedad.

No obstante, al considerar el régimen de discriminación de mínimo error (MESD), la inclusión de un resultado inconcluso restringe el espacio efectivo de medición, ya que parte de la probabilidad total se asigna a salidas que no conducen a una decisión binaria. Por este motivo, se consideró un esquema alternativo de asignación de resultados, redistribuyendo las salidas del circuito entre las familias consideradas.

En particular, se analizó la siguiente regla de decisión sin resultado inconcluso:

- **Familia 1:** Se asigna si el resultado de medición con mayor probabilidad es  $|00\rangle$  o  $|10\rangle$ .
- **Familia 2:** Se asigna si el resultado de medición con mayor probabilidad es  $|01\rangle$  o  $|11\rangle$ .

En esta configuración se elimina explícitamente la posibilidad de un resultado inconcluso, de modo que cada medición conduce necesariamente a una decisión binaria. Esta modificación amplía el conjunto efectivo de operadores POVM inducidos por el circuito y resulta adecuada para el análisis del régimen de discriminación de mínimo error.

#### D. Función de Costo

La función de costo  $J(\boldsymbol{\theta})$  codifica el objetivo del problema de discriminación. En particular, se busca maximizar la probabilidad de éxito  $P_{\text{suc}}$ , penalizando simultáneamente la probabilidad de error  $P_{\text{err}}$  y, dependiendo de la estrategia considerada (MED o UD), controlando la tasa de resultados inconclusos  $P_{\text{inc}}$ .

En este contexto, la función de costo se define como

$$J(\boldsymbol{\theta}) = \sum_i \frac{1}{|S_i|} \sum_{\rho_i \in S_i} \left( |P_{\text{suc}}(\rho_i) - 1| + \alpha_{\text{err}} |P_{\text{err}}(\rho_i)| + \alpha_{\text{inc}} |P_{\text{inc}}(\rho_i)| \right), \quad (77)$$

donde  $S_i$  es el conjunto de entrenamiento correspondiente a la familia  $i$  y  $|S_i|$  su cardinalidad. Los hiperparámetros  $\alpha_{\text{err}}$  y  $\alpha_{\text{inc}}$  controlan la penalización asociada a cometer un error de clasificación y a obtener un resultado inconcluso, respectivamente. Minimizar el término  $|P_{\text{suc}} - 1|$  es equivalente a maximizar la probabilidad de éxito.

Cabe destacar que en este trabajo se emplea la misma función de costo para ambos regímenes de discriminación considerados. La diferencia entre la discriminación de mínimo error (MESD) y la discriminación inequívoca (USD) no se introduce modificando la forma funcional de  $J(\boldsymbol{\theta})$ , sino a través de la elección de los hiperparámetros.

De acuerdo con Chen *et al.* [17], el régimen de discriminación inequívoca se obtiene cuando  $\alpha_{\text{err}} > \alpha_{\text{inc}}$ , es decir, cuando los errores de clasificación son penalizados más severamente que los resultados inconclusos. En este caso, el clasificador tiende a preferir respuestas inconclusas antes que cometer un error. Por el contrario, cuando  $\alpha_{\text{err}} < \alpha_{\text{inc}}$ , el costo asociado a los resultados inconclusos domina y el clasificador se comporta como un discriminador de mínimo error (MESD).

#### E. Elección de Hiperparámetros

En el presente trabajo, la estrategia de selección de hiperparámetros se simplificó con respecto a la empleada por Chen *et al.* [17], con el objetivo de reducir el costo computacional y adaptar el entrenamiento a las características de los conjuntos de datos considerados. En dicho trabajo, los autores exploran de manera sistemática el espacio de hiperparámetros variando el parámetro asociado al error en el intervalo

$\alpha_{\text{err}} \in [5, 40]$ , en pasos de 5, y el parámetro asociado a los resultados inconclusos en el intervalo  $\alpha_{\text{inc}} \in [2, 20]$ , en pasos de 2.

En este trabajo, se optó por fijar inicialmente el valor del parámetro asociado a los resultados inconclusos,  $\alpha_{\text{inc}}$ , y explorar distintos valores de  $\alpha_{\text{err}}$ , de modo de controlar explícitamente el compromiso entre errores de clasificación e inconclusiones, manteniendo al mismo tiempo un costo computacional manejable.

Como configuración inicial, se fijó  $\alpha_{\text{inc}} = 2$  y se varió  $\alpha_{\text{err}}$  en el intervalo  $\alpha_{\text{err}} \in [5, 40]$ , en pasos de 5. Esta elección permite recorrer de forma progresiva el espacio de parámetros relevante para los regímenes de discriminación inequívoca y de mínimo error, manteniendo una interpretación clara del rol de cada término en la función de costo.

La extensión del espacio de hiperparámetros explorado se ajustó en función del tamaño del conjunto de entrenamiento. Para conjuntos de datos grandes ( $N = 600$  y  $N = 1000$ ), el clasificador mostró un comportamiento estable y una convergencia consistente, por lo que no fue necesario ampliar significativamente el rango de valores considerados. En contraste, para conjuntos de entrenamiento pequeños (por ejemplo,  $N = 10$ ), el entrenamiento resultó más sensible a la elección de los hiperparámetros, requiriendo una exploración más amplia del espacio de parámetros. En estos casos, se consideraron valores mayores de  $\alpha_{\text{inc}}$ , incluyendo  $\alpha_{\text{inc}} = 4, 8$  y superiores, manteniendo la variación de  $\alpha_{\text{err}}$  en el mismo rango.

Esta estrategia permitió adaptar la función de costo a distintos regímenes de datos sin modificar su forma funcional, y facilitó el análisis del desempeño del clasificador en escenarios con diferente cantidad de información disponible. Los efectos de la elección de estos hiperparámetros sobre el desempeño del clasificador se analizan en detalle en la sección de resultados.

### *F. Estrategia de Optimización*

El entrenamiento del circuito cuántico variacional se formuló como un problema de minimización de la función de costo  $J(\boldsymbol{\theta})$  sobre el espacio de parámetros del *ansatz*. Para

---

ello, se empleó un enfoque híbrido cuántico-clásico utilizando la librería *PennyLane*, en el cual la evaluación del costo se realiza mediante simulación cuántica, mientras que la actualización de parámetros se lleva a cabo clásicamente mediante el optimizador **Adam**.

Los parámetros  $\theta$  se inicializaron aleatoriamente a partir de una distribución uniforme en el intervalo  $[0, 2\pi)$ , y se utilizó una tasa de aprendizaje inicial  $\eta = 0.01$ . El proceso de optimización se ejecutó por un máximo de 5000 iteraciones, aunque este número pudo reducirse dinámicamente mediante criterios de convergencia temprana.

Con el fin de acelerar el cálculo del gradiente y reducir el costo computacional, se implementó un esquema de *mini-batching*. En cada iteración, el optimizador se actualizó utilizando un subconjunto aleatorio correspondiente al 20% del conjunto de entrenamiento, generado de manera reproducible mediante un generador de números aleatorios con semilla fija. La función de pérdida se evaluó adicionalmente sobre un conjunto de validación independiente en cada iteración, permitiendo monitorear la generalización del modelo.

Para evitar el sobreajuste y mejorar la estabilidad del entrenamiento, se incorporó una estrategia de *early stopping*. En particular, si la pérdida de validación no presentaba una mejora significativa mayor que  $\delta = 10^{-4}$  durante 500 iteraciones consecutivas (*patience*), el entrenamiento se detenía y se restauraban los parámetros correspondientes al mejor valor de pérdida de validación observado.

Adicionalmente, se implementó un planificador adaptativo de tasa de aprendizaje del tipo *ReduceLROnPlateau*. Cuando la pérdida de validación permanecía estancada durante 150 iteraciones, la tasa de aprendizaje se reducía multiplicándola por un factor constante, hasta alcanzar un valor mínimo  $\eta_{\min} = 10^{-5}$ . Este mecanismo permitió un ajuste más fino de los parámetros en las etapas finales del entrenamiento y facilitó la convergencia hacia mínimos locales más estables.

Adicionalmente, con el objetivo de mitigar la presencia de mínimos locales y regiones planas del paisaje de optimización, se empleó una estrategia de reinicios múltiples. En particular, el proceso de entrenamiento se repitió varias veces utilizando los mis-

mos hiperparámetros, pero con diferentes inicializaciones aleatorias de los parámetros variacionales  $\theta$ . Cada intento de optimización se ejecutó de manera independiente, y al finalizar se seleccionó el conjunto de parámetros que produjo el menor valor de la pérdida de validación.

Este procedimiento permite aumentar la probabilidad de converger hacia soluciones más favorables del paisaje de optimización y constituye una práctica habitual en el entrenamiento de algoritmos cuánticos variacionales, donde la no convexidad de la función de costo puede dificultar la localización del mínimo global.

Durante el proceso de optimización se registraron, tanto en el conjunto de entrenamiento como en el de validación, las métricas de probabilidad de éxito ( $P_{\text{suc}}$ ), error ( $P_{\text{err}}$ ) e inconclusión ( $P_{\text{inc}}$ ), lo que permitió analizar de manera detallada la evolución del desempeño del discriminador. Finalmente, los parámetros óptimos  $\theta^*$  obtenidos se evaluaron sobre un conjunto de prueba independiente para estimar la capacidad de generalización del modelo.

### *G. Herramientas Computacionales*

1) *Implementación del Clasificador Variacional mediante PennyLane.* La implementación y el entrenamiento del clasificador cuántico variacional se realizaron utilizando la librería `PennyLane`, un marco de trabajo en Python diseñado para el desarrollo y la optimización de algoritmos cuánticos variacionales en esquemas híbridos cuántico-clásicos [48].

En este trabajo, `PennyLane` se emplea para definir el circuito cuántico parametrizado que implementa el clasificador, así como para evaluar la función de costo asociada a cada conjunto de parámetros. El circuito se implementa como un `QNode`, lo que permite encapsular la evolución cuántica y la medición en una función diferenciable con respecto a los parámetros variacionales. Esta característica resulta fundamental para la optimización basada en gradientes utilizada durante el entrenamiento.

Las simulaciones cuánticas se llevaron a cabo mediante el dispositivo `default.mixed`, el cual permite trabajar directamente con matrices de densidad y, por tanto, modelar

de manera explícita estados mixtos, mediciones generalizadas y efectos de ruido [48].

A partir de la salida probabilística del circuito, PennyLane permite construir de forma directa las cantidades relevantes para el entrenamiento y la evaluación del clasificador, tales como las probabilidades de éxito, error e inconclusión. Estas magnitudes se combinan posteriormente en la función de costo, la cual se evalúa de manera repetida durante el proceso de optimización.

Finalmente, la integración de PennyLane con bibliotecas clásicas de optimización permite implementar de forma eficiente el bucle híbrido de entrenamiento, incluyendo estrategias como *mini-batching*, *early stopping*, ajuste adaptativo de la tasa de aprendizaje y reinicios múltiples. De este modo, PennyLane actúa como el núcleo cuántico del clasificador variacional, mientras que la lógica de optimización y análisis se gestiona de manera clásica.

2) *Implementación Numérica mediante la Librería PICOS.* Los programas semidefinidos formulados en el Capítulo II pueden resolverse de manera exacta mediante técnicas de optimización convexa. Con el fin de obtener las soluciones óptimas asociadas a estos problemas y evaluar las cotas teóricas correspondientes, se utilizó la librería PICOS, la cual proporciona una interfaz de modelado en Python para problemas de optimización cónica convexa [49].

Desde el punto de vista computacional, PICOS permite modelar de forma directa problemas de programación semidefinida, gestionando internamente las condiciones de semidefinitud ( $X \succeq 0$ ) y las restricciones lineales sobre operadores hermíticos, y traduciéndolas automáticamente a la forma estándar requerida por el solucionador numérico subyacente [49].

En consecuencia, los valores óptimos  $P_{\text{guess}}^*$  y  $P_{\text{USD}}^*$  obtenidos a partir de estos programas semidefinidos deben interpretarse como cotas teóricas exactas, derivadas de optimización convexa, y se utilizan como referencia para el análisis posterior del desempeño de los clasificadores cuánticos variacionales.

---

# CAPÍTULO VI: RESULTADOS Y DISCUSIÓN

## *A. Clasificador para las Familias de Validación*

En esta sección se presentan los resultados obtenidos para la clasificación entre las dos familias de estados cuánticos definidas en [17] ( $\rho_1(a)$ : Ec. (59) y  $\rho_2(b)$ : Ec. (62)). Para evaluar la robustez del modelo, se consideraron cuatro escenarios distintos basados en las distribuciones de probabilidad de los parámetros  $a$  y  $b$ , los cuales se resumen en la Tabla 1. En adelante, estos escenarios se denominarán simplemente Casos 1 al 4.

1) *Fidelidad de las familias de estados.* Se calculó la distribución de la fidelidad  $F$  para cada caso variando el tamaño del conjunto de entrenamiento  $N$ . Las figuras a continuación muestran, en el panel izquierdo, los histogramas de fidelidad junto con su valor medio para distintos  $N$ . En el panel derecho se muestra el mapa de calor de la fidelidad  $F(a, b)$ , sobre el cual se superpone un subconjunto representativo de las muestras de entrenamiento para ilustrar cómo se distribuyen en el espacio de parámetros.

Caso 1

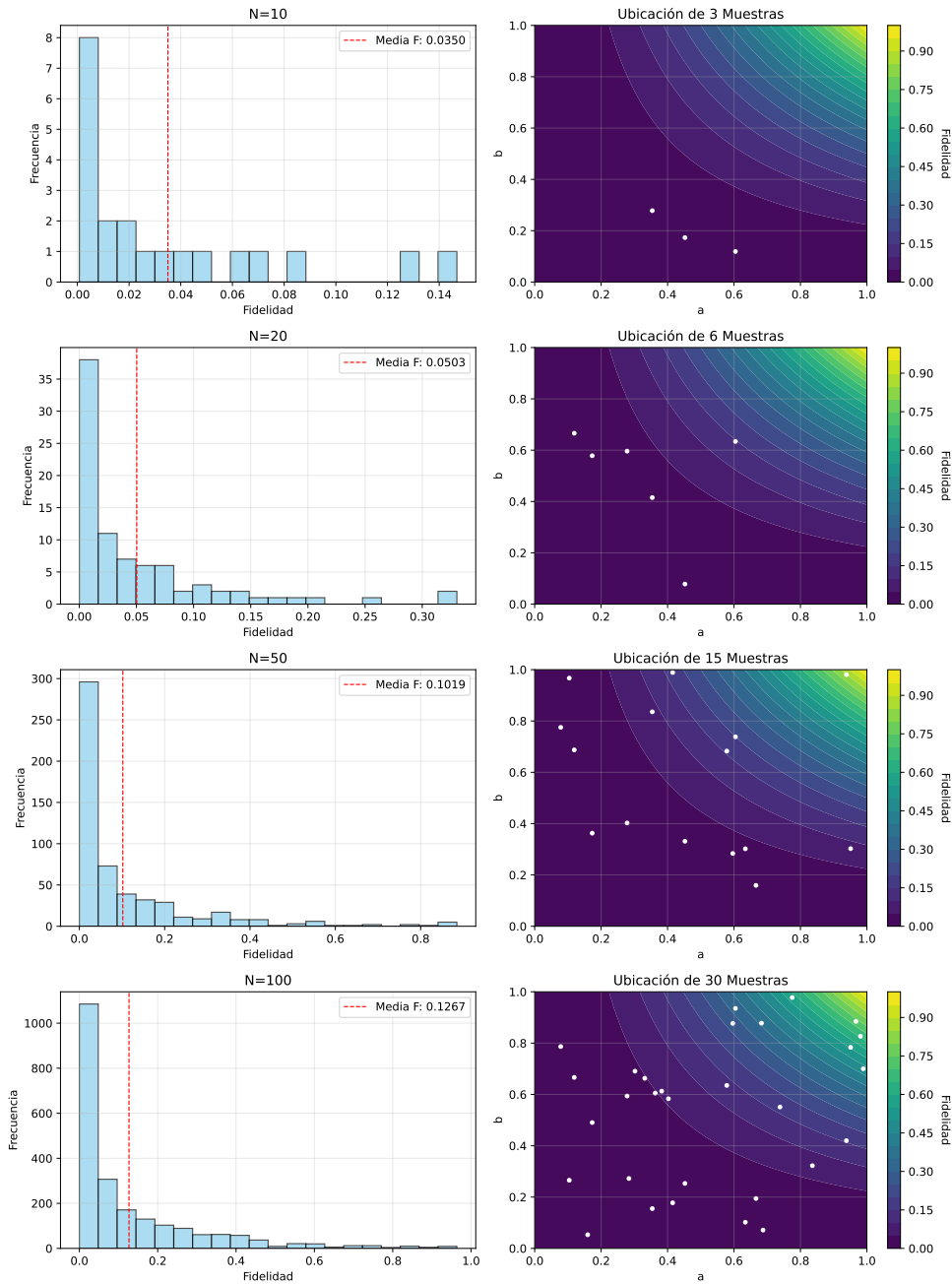


Figura 7: Distribución de la fidelidad  $F$  entre las familias  $\rho_1(a)$  y  $\rho_2(b)$  para el Caso 1 y tamaños de entrenamiento  $N \in \{10, 20, 50, 100\}$ . Panel izquierdo: histogramas de  $F$  junto con su valor medio (indicado por la línea roja punteada). Panel derecho: mapa de calor de  $F(a, b)$  en el espacio de parámetros, donde los puntos blancos representan un subconjunto de las muestras de entrenamiento.

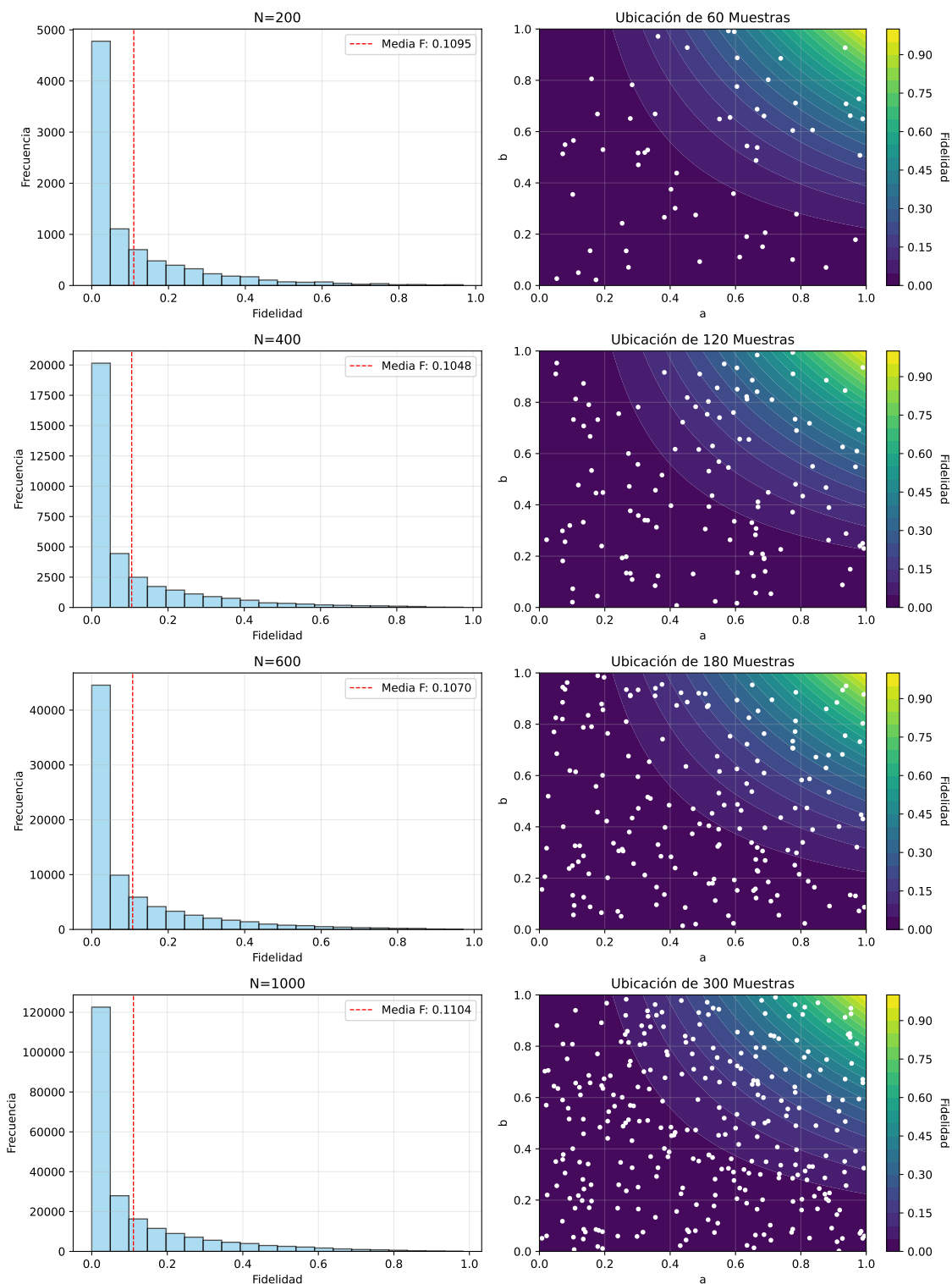


Figura 8: Distribución de la fidelidad  $F$  entre las familias  $\rho_1(a)$  y  $\rho_2(b)$  para el Caso 1 y tamaños de entrenamiento  $N \in \{200, 400, 600, 1000\}$ . Se muestran los histogramas de  $F$  junto con su valor medio (indicado por la línea roja punteada).

Caso 2

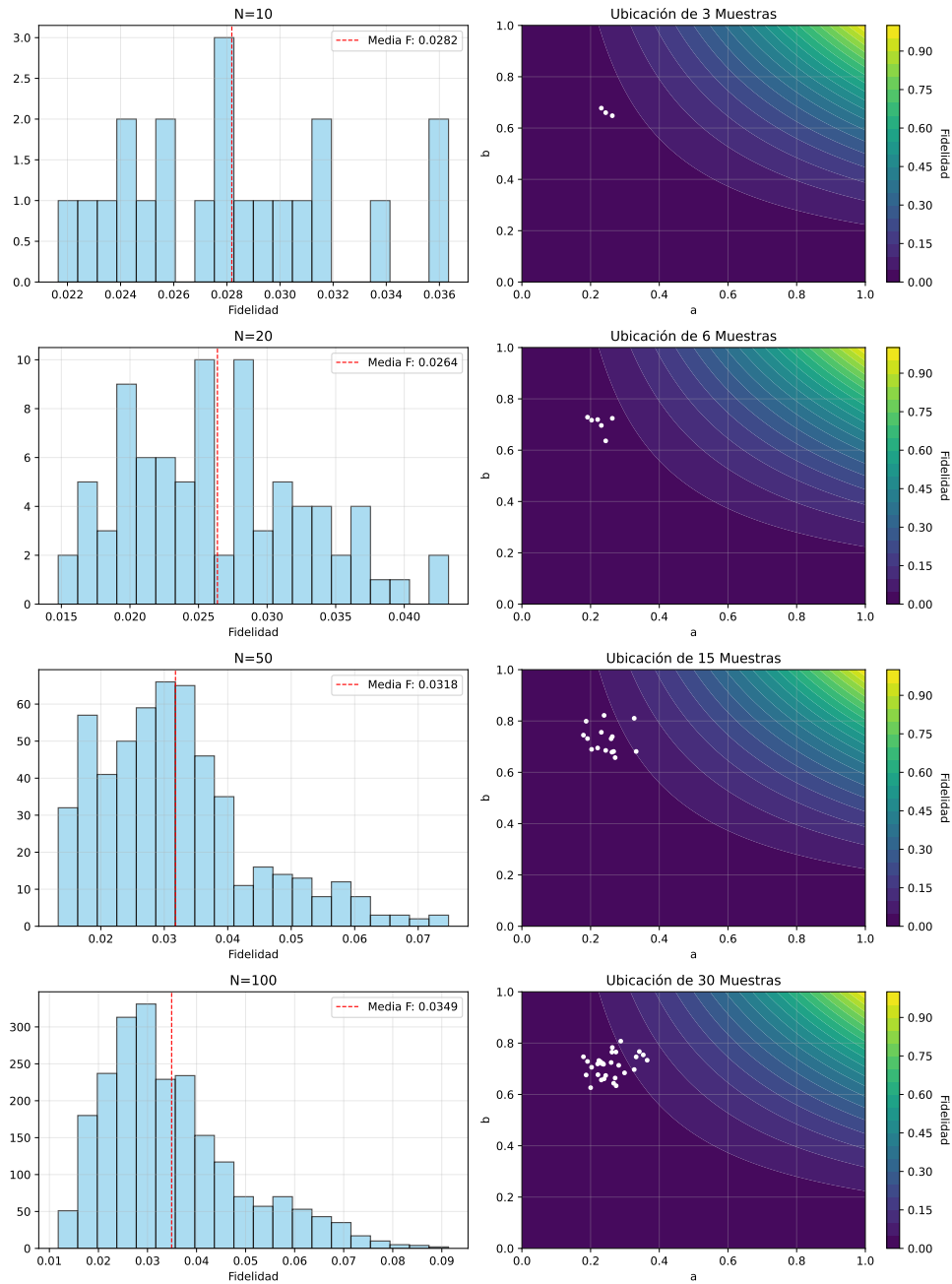


Figura 9: Distribución de la fidelidad  $F$  entre las familias  $\rho_1(a)$  y  $\rho_2(b)$  para el Caso 2 y tamaños de entrenamiento  $N \in \{10, 20, 50, 100\}$ . Panel izquierdo: histogramas de  $F$  junto con su valor medio (indicado por la línea roja punteada). Panel derecho: mapa de calor de  $F(a, b)$  en el espacio de parámetros, donde los puntos blancos representan un subconjunto de las muestras de entrenamiento.

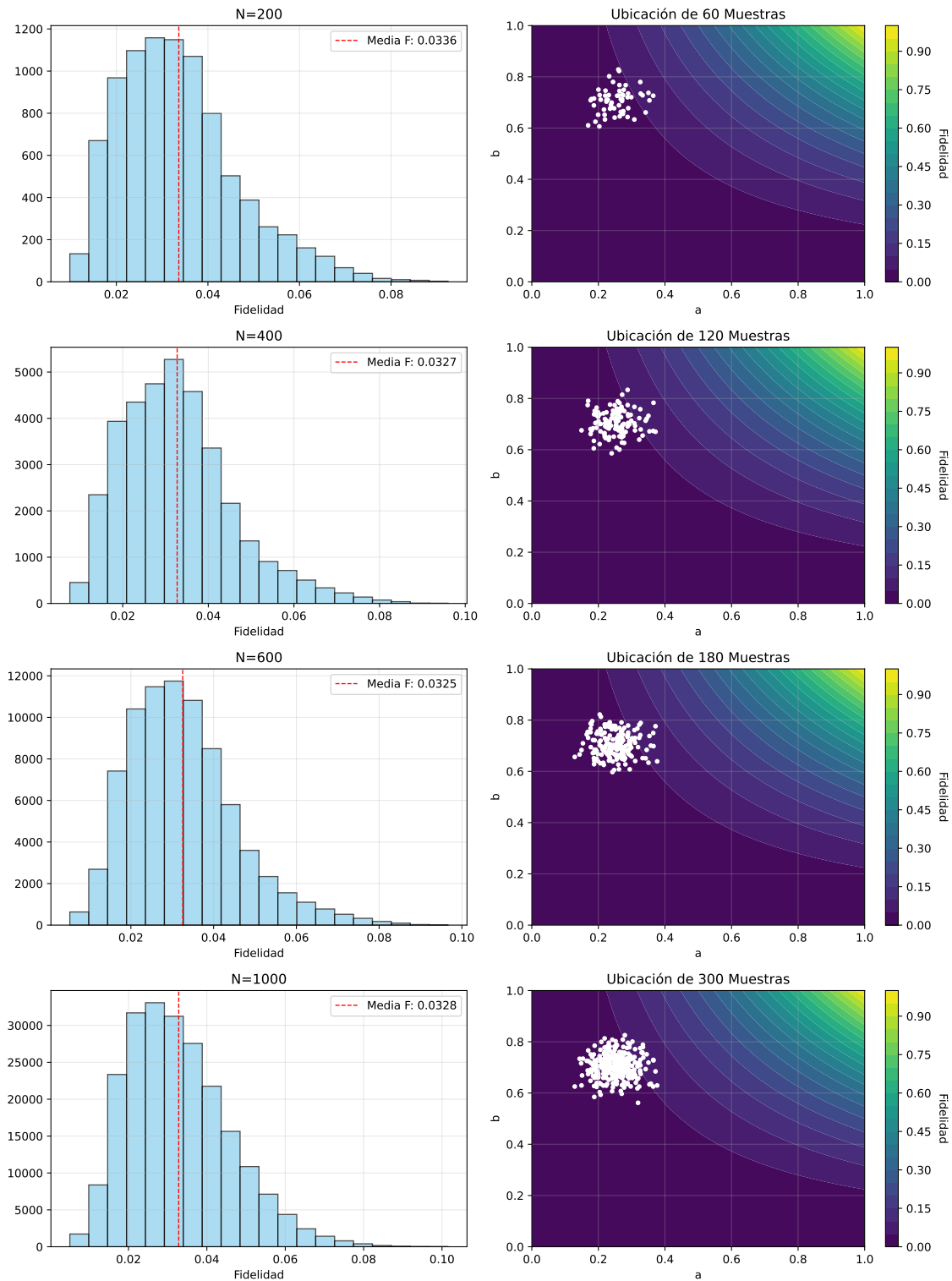


Figura 10: Distribución de la fidelidad  $F$  entre las familias  $\rho_1(a)$  y  $\rho_2(b)$  para el Caso 2 y tamaños de entrenamiento  $N \in \{200, 400, 600, 1000\}$ . Se muestran los histogramas de  $F$  junto con su valor medio (indicado por la línea roja punteada).

Caso 3

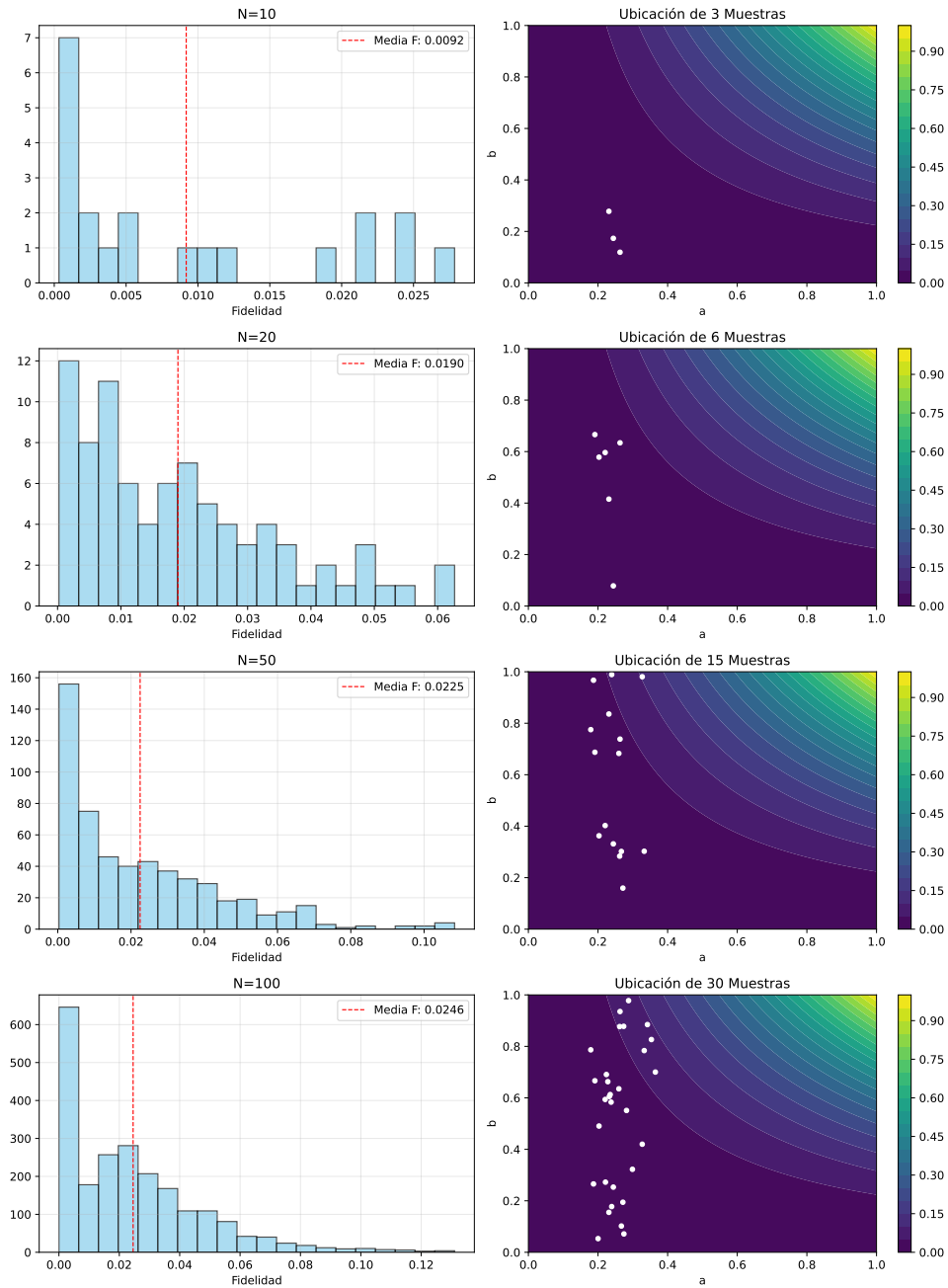


Figura 11: Distribución de la fidelidad  $F$  entre las familias  $\rho_1(a)$  y  $\rho_2(b)$  para el Caso 3 y tamaños de entrenamiento  $N \in \{10, 20, 50, 100\}$ . Panel izquierdo: histogramas de  $F$  junto con su valor medio (indicado por la línea roja punteada). Panel derecho: mapa de calor de  $F(a, b)$  en el espacio de parámetros, donde los puntos blancos representan un subconjunto de las muestras de entrenamiento.

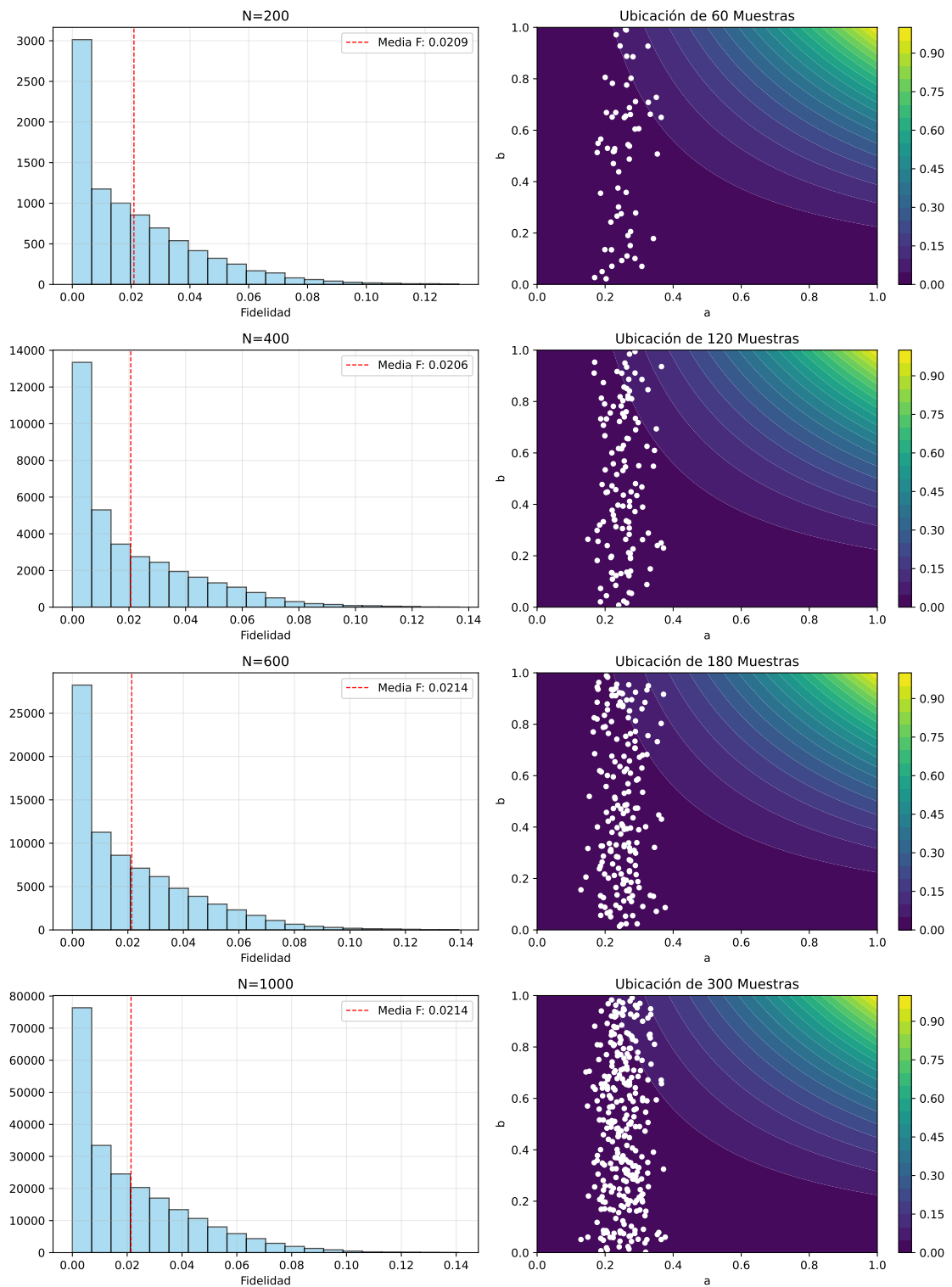


Figura 12: Distribución de la fidelidad  $F$  entre las familias  $\rho_1(a)$  y  $\rho_2(b)$  para el Caso 3 y tamaños de entrenamiento  $N \in \{200, 400, 600, 1000\}$ . Se muestran los histogramas de  $F$  junto con su valor medio (indicado por la línea roja punteada).

Caso 4

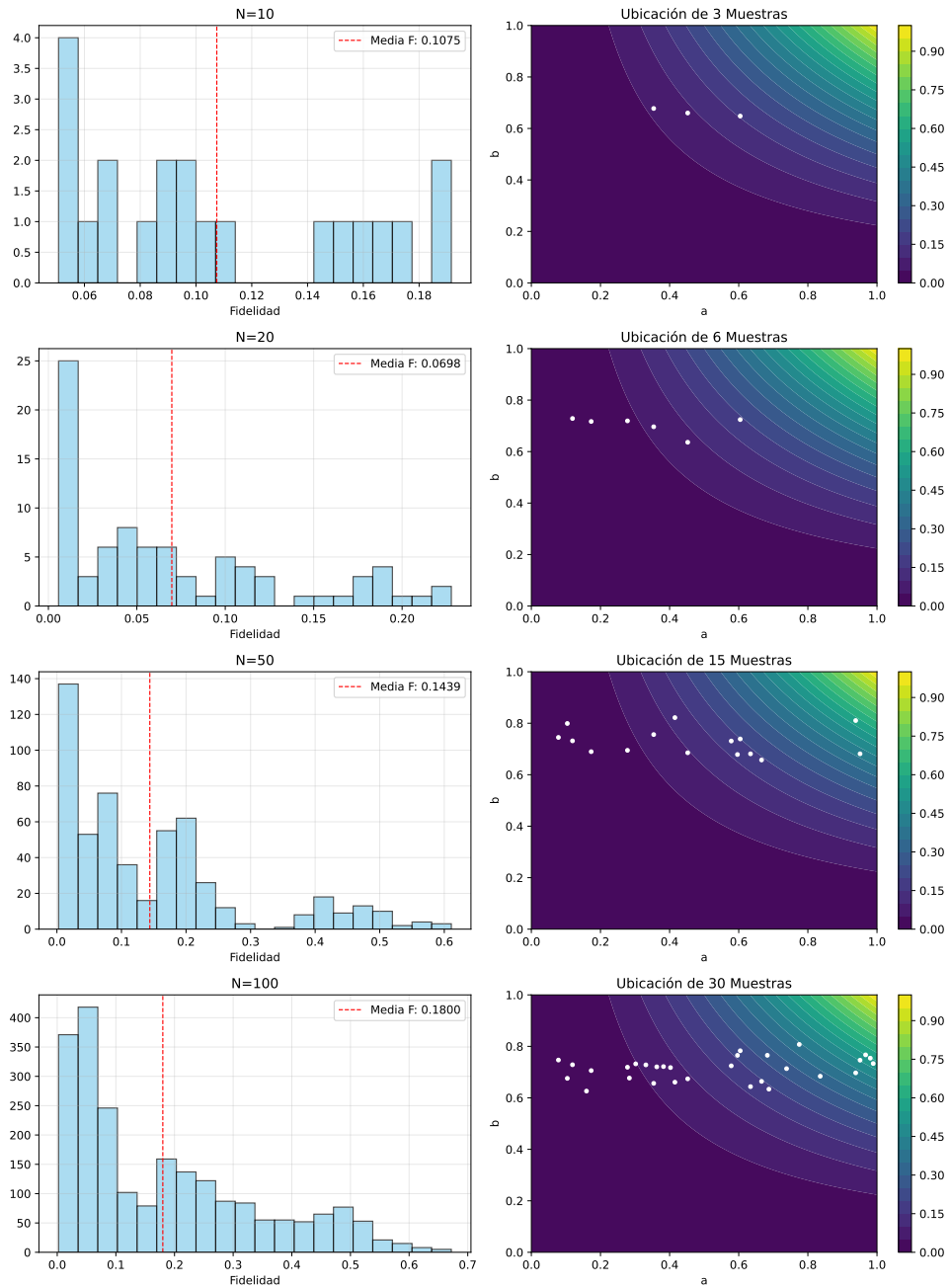


Figura 13: Distribución de la fidelidad  $F$  entre las familias  $\rho_1(a)$  y  $\rho_2(b)$  para el Caso 4 y tamaños de entrenamiento  $N \in \{10, 20, 50, 100\}$ . Panel izquierdo: histogramas de  $F$  junto con su valor medio (indicado por la línea roja punteada). Panel derecho: mapa de calor de  $F(a, b)$  en el espacio de parámetros, donde los puntos blancos representan un subconjunto de las muestras de entrenamiento.

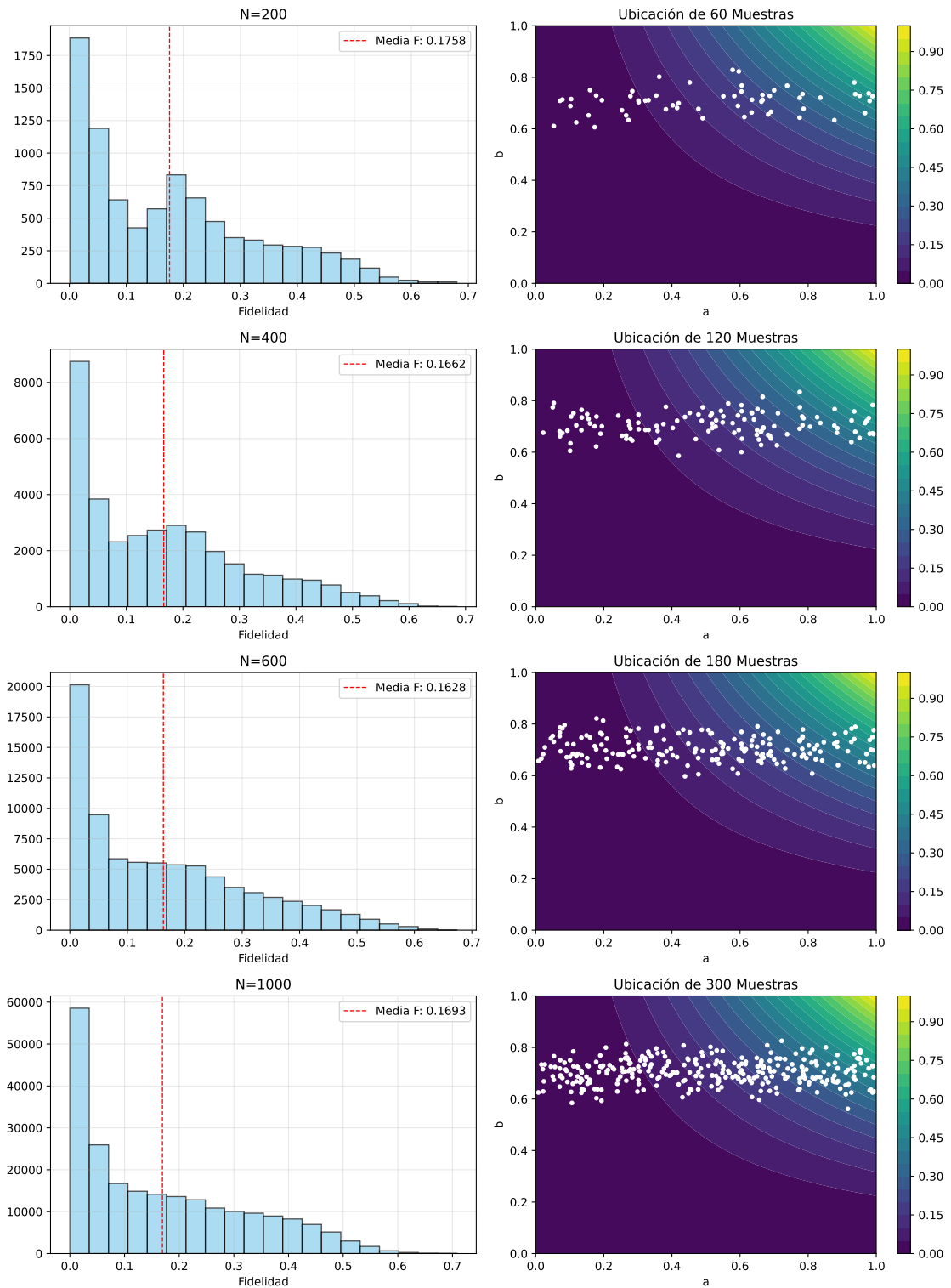


Figura 14: Distribución de la fidelidad  $F$  entre las familias  $\rho_1(a)$  y  $\rho_2(b)$  para el Caso 4 y tamaños de entrenamiento  $N \in \{200, 400, 600, 1000\}$ . Se muestran los histogramas de  $F$  junto con su valor medio (indicado por la línea roja punteada).

El análisis de las Figs. 7–14 muestra que las distribuciones de fidelidad presentan promedios pequeños en todos los escenarios, oscilando entre un mínimo de 0.0092 (Caso 3,  $N = 10$ ) y un máximo de 0.1800 (Caso 4,  $N = 100$ ). La concentración de muestras en la región de baja fidelidad ( $F < 0.2$ ) es consistente con los mapas de calor, donde los puntos se agrupan en las zonas frías del espacio de parámetros.

Esta baja fidelidad sugiere, desde una perspectiva geométrica, una separación significativa entre los estados de ambas familias, lo que implica un solapamiento mínimo. Teóricamente, esta alta distinguibilidad intrínseca debería favorecer el desempeño del discriminador variacional. Asimismo, se observa que el comportamiento estadístico de los histogramas tiende a estabilizarse conforme aumenta el tamaño del conjunto de entrenamiento  $N$ .

2) *Tasas de Éxito Teóricas.* Con el fin de establecer una comparación rigurosa, se calcularon las cotas máximas teóricas de la probabilidad de éxito para la clasificación de estas familias de estados. En particular, se consideró el límite de Helstrom para el caso de mínimo error (Ec. 29), mientras que la discriminación inequívoca se evaluó mediante programación semidefinida. Estos cálculos se realizaron asumiendo las mismas probabilidades a priori empleadas durante el entrenamiento:  $\eta_a = 1/3$  y  $\eta_b = 2/3$ .

*Discriminación de Mínimo Error (MESD).* En el esquema de discriminación de mínimo error, el objetivo es maximizar la probabilidad total de acierto permitiendo errores de clasificación. Las cotas máximas teóricas de la probabilidad de éxito se presentan en la Tabla 2.

|                  | $b \approx \frac{1}{\sqrt{2}} \quad b \in [0, 1]$ |      |
|------------------|---|------|
| $a \approx 0.25$ | 0.99  | 0.99 |
| $a \in [0, 1]$   | 0.91  | 0.92 |

Cuadro 2: Cota máxima teórica de la probabilidad de éxito  $P_{\text{suc}}^{\text{máx}}$  para la discriminación de mínimo error (MESD), obtenida mediante el límite de Helstrom (Ec. 29). Las filas indican el rango del parámetro  $a$  y las columnas el rango del parámetro  $b$ . Los valores se calcularon asumiendo probabilidades a priori  $\eta_a = 1/3$  y  $\eta_b = 2/3$ .

*Discriminación Inequívoca (USD)*. En contraste, la discriminación inequívoca impone la condición estricta de no permitir errores de clasificación, admitiendo únicamente resultados correctos o inconclusos. Los resultados se muestran en la Tabla 3 y son consistentes con los reportados en la literatura [17, 50].

|                  | $b \approx \frac{1}{\sqrt{2}} \quad b \in [0, 1]$ |      |
|------------------|---|------|
| $a \approx 0.25$ | 0.64  | 0.76 |
| $a \in [0, 1]$   | 0.55  | 0.67 |

Cuadro 3: Cota máxima teórica de la probabilidad de éxito  $P_{\text{suc}}^{\text{máx}}$  en el esquema de discriminación inequívoca (USD), obtenida mediante programación semidefinida (SDP). Las filas indican el rango del parámetro  $a$  y las columnas el rango del parámetro  $b$ . Los valores se calcularon asumiendo probabilidades a priori  $\eta_a = 1/3$  y  $\eta_b = 2/3$ .

La diferencia entre los resultados obtenidos para la discriminación de mínimo error y la discriminación inequívoca ayuda a entender mejor las diferencias entre ambos esquemas. En el caso de MESD, las probabilidades de éxito alcanzan valores cercanos al 100 %, lo que concuerda con el análisis de fidelidad presentado previamente. La baja fidelidad entre estados de distintas familias indica que el solapamiento entre ellos es reducido, lo que sugiere una alta distinguibilidad. En este contexto, la discriminación de mínimo error se ve favorecida, ya que puede aprovechar estas diferencias para minimizar

la tasa de error.

En cambio, las probabilidades de éxito asociadas a la discriminación inequívoca son considerablemente menores. Esto no contradice la baja fidelidad observada, sino que refleja las condiciones más estrictas que impone la USD. Al exigir que no se permitan errores de clasificación, se restringe de forma importante el conjunto de mediciones que pueden implementarse, especialmente cuando los estados son mixtos o presentan solapamientos en sus soportes. Como resultado, una fracción significativa de los eventos debe asignarse al resultado inconcluso.

En conjunto, estos resultados muestran que una buena separación entre los estados es suficiente para obtener un excelente desempeño en MESD, pero no garantiza necesariamente una USD eficiente.

*3) Resultados Numéricos del Clasificador en el Régimen USD.* El desempeño del clasificador variacional implementado en el régimen de discriminación inequívoca se resume en la Fig. 15, donde se compara la probabilidad de éxito alcanzada por el clasificador con la cota teórica en función de  $N$ .

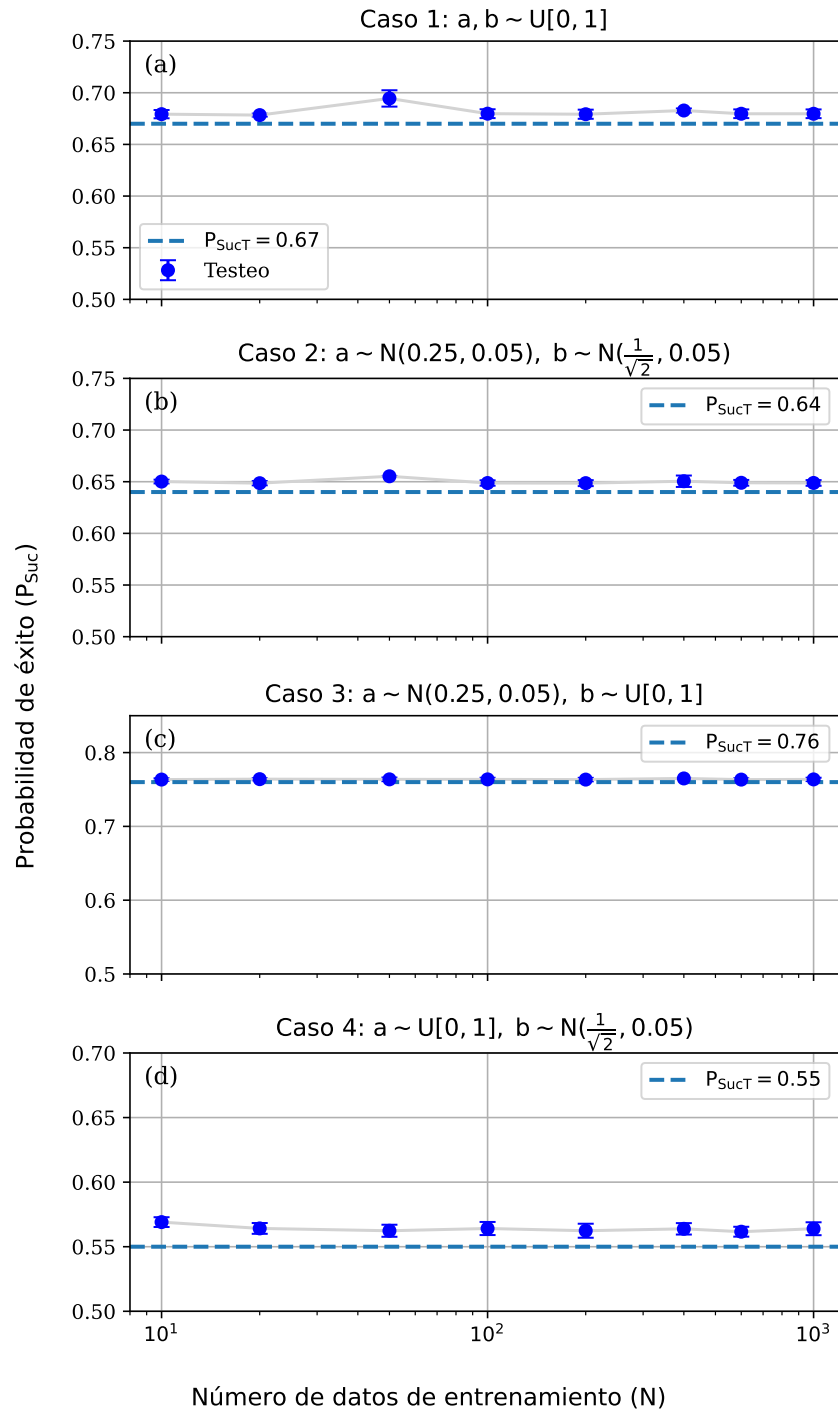


Figura 15: Probabilidad de éxito  $P_{\text{Suc}}$  del clasificador variacional en el régimen de discriminación inequívoca (USD) en función del número de datos de entrenamiento  $N$ . Los paneles (a)–(d) corresponden a los Casos 1–4, respectivamente. Los puntos indican el desempeño promedio obtenido en la fase de testeo (mostrando barras de error), mientras que la línea punteada representa la cota teórica máxima calculada mediante programación semidefinida (SDP) para cada escenario.

En términos generales, el clasificador muestra un desempeño robusto y se aproxima a la cota teórica incluso cuando se entrena con conjuntos de datos relativamente pequeños. En algunos escenarios, la probabilidad de éxito obtenida supera ligeramente la cota calculada mediante programación semidefinida (SDP). Este comportamiento no implica una violación de los límites teóricos. En particular, la cota reportada en la Tabla 3 corresponde a un esquema de discriminación inequívoca estricta, en el cual se impone la condición  $P_{\text{err}} = 0$ . Por el contrario, el clasificador variacional optimiza una función de costo que penaliza los errores a través del parámetro  $\alpha_{\text{err}}$ , sin excluirlos de forma definitiva. Como resultado, el circuito converge hacia una estrategia híbrida, en la que se tolera una tasa de error pequeña a cambio de mejorar la probabilidad de éxito.

Este comportamiento se aprecia en el Caso 1 con  $N = 1000$ , mostrado en la Fig. 16. En este escenario, el clasificador se equivoca una vez al clasificar un estado de  $\rho_2$  como perteneciente a  $\rho_1$ , y alcanza una tasa de éxito  $P_{\text{Suc}} = 0.68$ , ligeramente superior a la cota teórica  $P_{\text{suc}}^{\text{teo}} = 0.67$ .

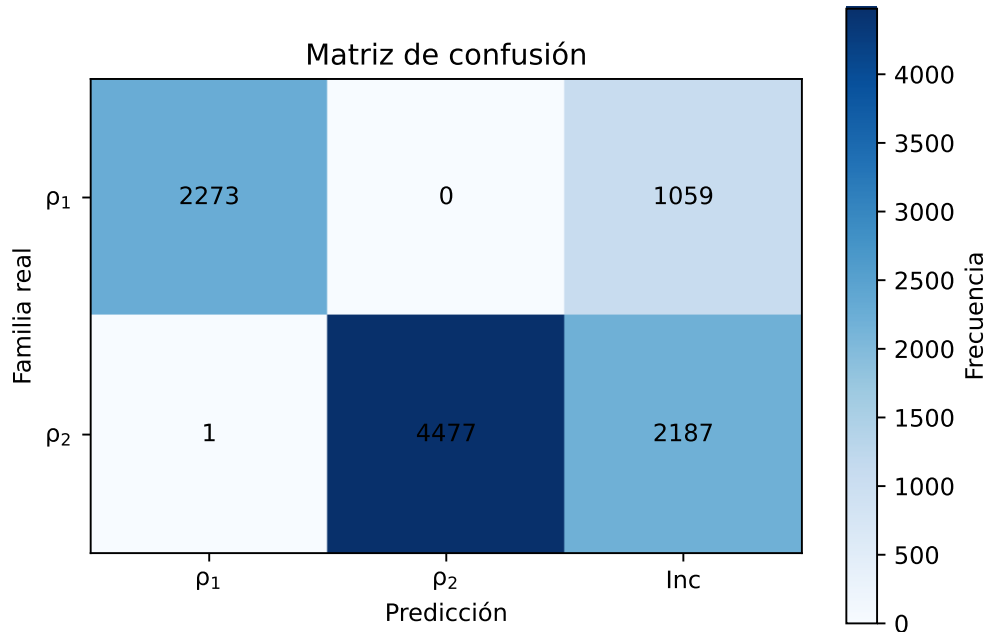


Figura 16: Matriz de confusión correspondiente al Caso 1 con  $N = 1000$ , utilizando los parámetros de penalización  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (40, 2)$ . Se observa un único error de clasificación, en el cual un estado de  $\rho_2$  es identificado como  $\rho_1$ .

Sin embargo, es importante mencionar que, en todos los casos, la tasa de éxito promedio del clasificador sólo supera la cota teórica en aproximadamente 0.01, lo que corresponde a una mejora del orden del 1%. Incluso, para el Caso 3, el valor promedio de la tasa de éxito coincide con la cota teórica. Estos resultados se muestran en 4.

|                  | $b \approx \frac{1}{\sqrt{2}} \quad b \in [0, 1]$ |      |
|------------------|---|------|
| $a \approx 0.25$ | 0.65  | 0.76 |
| $a \in [0, 1]$   | 0.56  | 0.68 |

Cuadro 4: Valores promedio de la probabilidad de éxito  $P_{\text{suc}}$  alcanzados por el clasificador variacional en el régimen de discriminación inequívoca (USD) para los cuatro escenarios considerados. Los resultados muestran una concordancia cercana con las cotas teóricas obtenidas mediante programación semidefinida (SDP), con diferencias del orden del 1%.

En comparación con los resultados reportados en [17], el clasificador presenta un comportamiento similar en los Casos 1 y 4. No obstante, para los Casos 2 y 3, en [17] se reportan cotas de éxito superiores a las teóricas, alcanzando mejoras de hasta un 20% en el Caso 2 y un 11% en el Caso 4. En el marco de este trabajo, no se alcanzaron tasas de éxito tan elevadas para la discriminación inequívoca.

*Sensibilidad a los hiperparámetros en el régimen de pocos datos.* En los cuatro casos considerados, se observó que, para conjuntos de entrenamiento con pocos datos, en particular para  $N = 10$ , la diferencia relativa entre  $\alpha_{\text{err}}$  y  $\alpha_{\text{inc}}$  desempeña un papel importante en la estrategia de convergencia del modelo. Para configuraciones con una penalización al error considerablemente mayor que la asociada a la incertidumbre, como  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (25, 2)$  o  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (35, 2)$ , la tasa de éxito disminuye de forma notable. En los Casos 1 y 3 se obtienen valores  $P_{\text{suc}} \approx 0.44$ , mientras que en los Casos 2 y 4 la tasa de éxito cae hasta  $P_{\text{suc}} \approx 0.33$ , resultados muy por debajo de las predicciones teóricas. Este comportamiento sugiere que una penalización excesiva al error puede congelar el proceso de aprendizaje del clasificador.

Las curvas de entrenamiento y validación del Caso 4 con  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (35, 2)$ , Fig. 17, muestran que, bajo esta elección de hiperparámetros, el clasificador reduce rápidamente la tasa de error hasta valores cercanos a cero. Sin embargo, esta reducción no implica un aumento significativo de la tasa de éxito, la cual permanece prácticamente constante a lo largo del entrenamiento. En contraste, la probabilidad de resultados inconclusos aumenta progresivamente y pasa a dominar el comportamiento del clasificador.

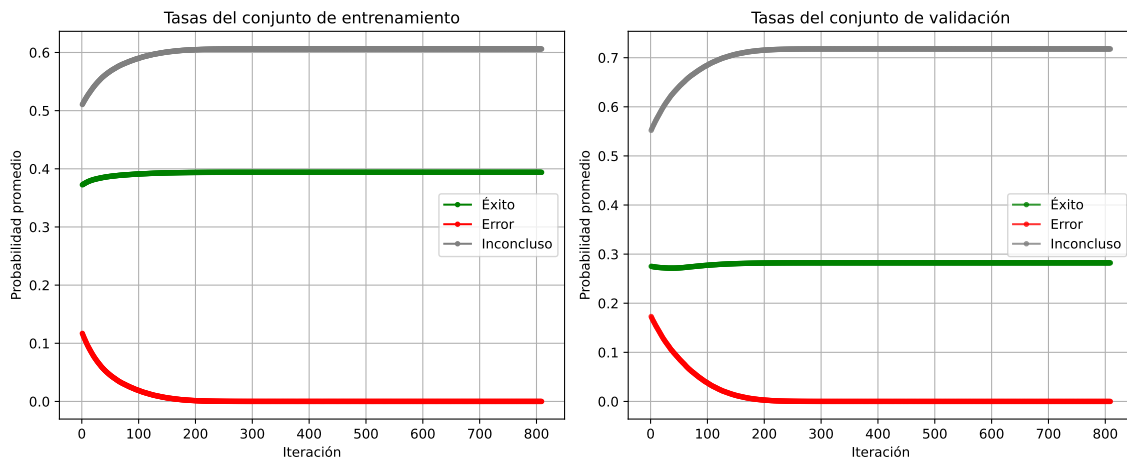


Figura 17: Evolución de las tasas de éxito, error e inconclusión en los conjuntos de entrenamiento (izquierda) y validación (derecha) para el Caso 4 con  $N = 10$  y  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (35, 2)$ . Se observa una rápida reducción de la tasa de error hacia valores cercanos a cero, mientras que la probabilidad de éxito permanece aproximadamente constante y la tasa de resultados inconclusos aumenta, pasando a dominar el comportamiento del clasificador.

Cabe resaltar que, al incrementar el tamaño del conjunto de entrenamiento, la sensibilidad del clasificador a la diferencia relativa entre los hiperparámetros disminuye, y el modelo converge hacia tasas de éxito acordes con las predicciones teóricas.

*Convergencia hacia soluciones MESD.* En el trabajo de Chen *et al.* [17] se establece que la condición  $\alpha_{\text{err}} > \alpha_{\text{inc}}$  es necesaria para que el clasificador pueda converger hacia soluciones de discriminación inequívoca (USD). Sin embargo, los resultados obtenidos en este trabajo indican que dicha condición no es suficiente para garantizar, en la práctica, la convergencia al régimen de USD. En particular, se observa que para ciertos valores

de los hiperparámetros, como  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (5, 2)$ , el clasificador puede converger hacia soluciones características del régimen de discriminación de mínimo error (MESD), aun cuando se satisface la desigualdad  $\alpha_{\text{err}} > \alpha_{\text{inc}}$ .

Este comportamiento se observa, por ejemplo, en el Caso 1 con  $N = 10$  y  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (5, 2)$ , donde el clasificador alcanza una tasa de éxito de  $P_{\text{suc}} = 0.78$ . La matriz de confusión correspondiente, mostrada en la Fig. 18, indica que la asignación de resultados se concentra en decisiones binarias, sin la aparición de una región dominante de resultados inconclusos, lo cual es característico del régimen de discriminación de mínimo error.

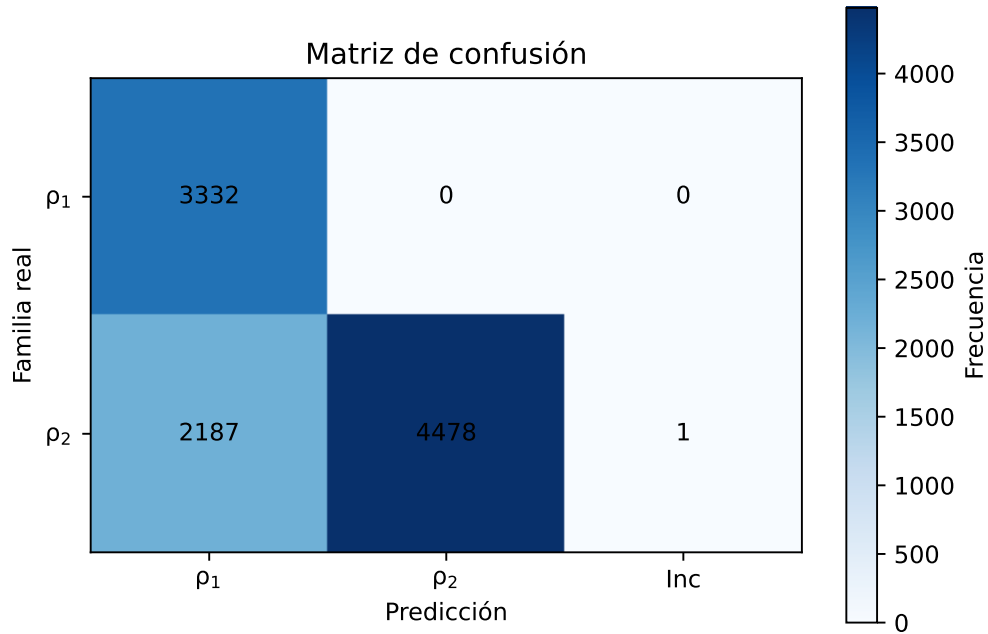


Figura 18: Matriz de Confusión correspondiente al Caso 1 con  $N = 10$  y  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (5, 2)$ . Se observa un único evento de la familia  $\rho_2$  clasificado como inconcluso, mientras que las decisiones se concentran casi exclusivamente en asignaciones binarias. Este comportamiento es característico del régimen de discriminación de mínimo error (MESD).

Por otro lado, en los experimentos realizados no se observaron casos en los que el clasificador convergiera a soluciones dentro del régimen de discriminación inequívoca cuando se cumple  $\alpha_{\text{err}} < \alpha_{\text{inc}}$ . En todos estos escenarios, las soluciones obtenidas corresponden al régimen de discriminación de mínimo error (MESD), en concordancia con lo reportado en [17]. De manera similar, para los casos en los que  $\alpha_{\text{err}} = \alpha_{\text{inc}}$ , el

clasificador converge a soluciones de tipo MESD.

4) *Resultados numéricos del clasificador en el régimen MESD.* Los valores promedio de la probabilidad de éxito del clasificador en el régimen de discriminación de mínimo error (MESD) se presentan en la Tabla 5.

|                  | $b \approx \frac{1}{\sqrt{2}} \quad b \in [0, 1]$ |      |
|------------------|---|------|
| $a \approx 0.25$ | 0.66  | 0.78 |
| $a \in [0, 1]$   | 0.65  | 0.78 |

Cuadro 5: Promedio de la probabilidad de éxito  $P_{\text{succ}}$  alcanzada por el clasificador en el régimen de discriminación de mínimo error (MESD), para las distintas configuraciones de los parámetros  $a$  y  $b$ .

El hecho de que el clasificador opere en el régimen de mínimo error se evidencia en el comportamiento de las matrices de confusión. En particular, en el Caso 4 con  $N = 20$  y  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (5, 20)$ , la matriz de confusión mostrada en la Fig. 19 presenta una ausencia total de resultados inconclusos, lo que indica que el modelo adopta una estrategia de decisión determinista característica del régimen de discriminación de mínimo error (MESD).

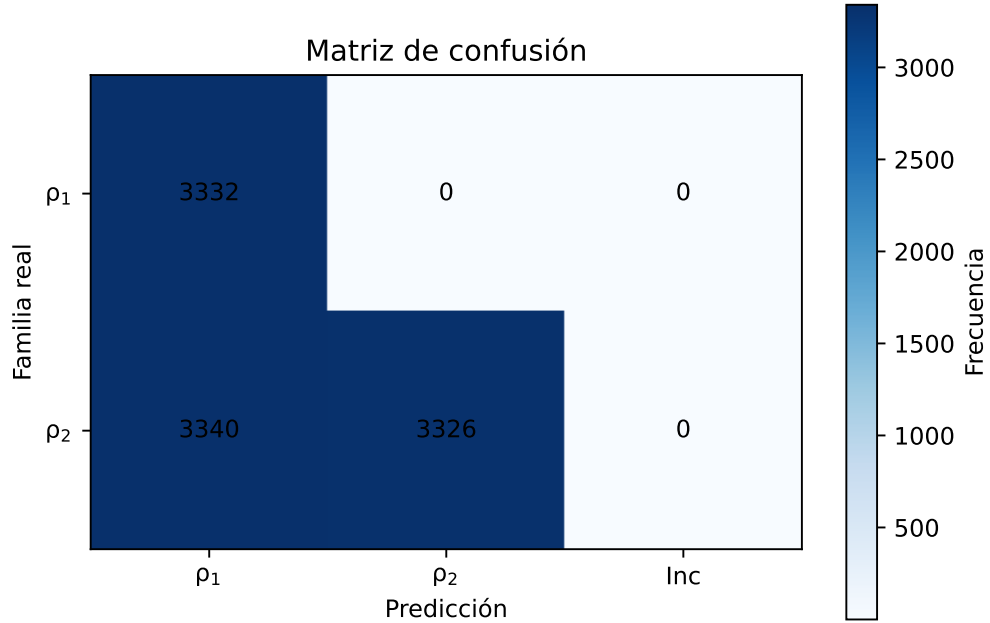


Figura 19: Matriz de confusión correspondiente al Caso 4 con  $N = 20$  y  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (5, 20)$ . Se observa que las decisiones son asignaciones exclusivamente binarias. Este comportamiento es característico del régimen de discriminación de mínimo error (MESD).

No obstante, las tasas de éxito obtenidas no alcanzan las cotas teóricas máximas esperadas para la discriminación de mínimo error. En todos los casos se observa una discrepancia superior al 10% entre el desempeño del clasificador y las cotas teóricas, alcanzando incluso una diferencia cercana al 33% en el Caso 2. Para el caso mostrado en la Fig. 19, la probabilidad de éxito del clasificador es  $P_{\text{suc}} = 0.66$ , mientras que la cota teórica correspondiente al Caso 4 es  $P_{\text{suc}}^{\text{teo}} = 0.91$ .

Esta diferencia puede atribuirse a la relación entre la estructura geométrica de las familias de estados y la forma en que el clasificador asigna sus salidas. Mientras que  $\rho_1$  está compuesta por estados puros de rango uno,  $\rho_2$  está formada por estados mixtos de rango dos, cuyo soporte ocupa un subespacio de mayor dimensión en el espacio de Hilbert. Sin embargo, en la implementación considerada se asigna un mayor número de resultados de medición a  $\rho_1$  que a  $\rho_2$ , introduciendo una asimetría en la representación de ambas familias. Esta elección restringe la capacidad del modelo para capturar completamente el soporte de los estados mixtos, lo que limita la tasa de éxito

alcanzable incluso en el régimen de discriminación de mínimo error.

Esta misma asimetría permite comprender por qué el clasificador sí logra aproximarse a las cotas teóricas en el régimen de discriminación inequívoca. En este régimen, la presencia de resultados inconclusos funciona como un mecanismo que absorbe la ambigüedad asociada a los estados mixtos, permitiendo que las componentes del estado que no pueden ser identificadas sin ambigüedad se asignen al resultado inconcluso. En contraste, en el régimen de discriminación de mínimo error (MESD) esta flexibilidad desaparece. Al no disponerse de una salida inconclusa que actúe como canal de compensación, la arquitectura del modelo impone una restricción sobre la tasa de éxito alcanzable, incluso cuando el proceso de entrenamiento converge de forma estable.

Desde un punto de vista geométrico, en la configuración original del clasificador el operador de medición asociado a la familia  $\rho_2$ , denotado por  $\Pi_2$ , es un proyector de rango uno, mientras que los estados de dicha familia son estados mixtos de rango dos. La probabilidad de identificación correcta está dada por  $\text{Tr}[\Pi_2\rho_2]$ , lo que implica que únicamente la componente de  $\rho_2$  contenida en el soporte de  $\Pi_2$  contribuye directamente al resultado de clasificación. En consecuencia, si el soporte de  $\rho_2$  no está contenido en el soporte de  $\Pi_2$ , la medición impone una restricción estructural sobre el subespacio accesible para su identificación. Desde esta perspectiva, la discrepancia observada respecto a la cota teórica puede interpretarse como resultado de una limitación en la clase de efectos de medición (POVMs) realizables por la arquitectura considerada.

Para ilustrar el análisis anterior, se estudia nuevamente el caso presentado en la Fig. 19, ahora considerando una configuración del clasificador en la que se asignó una de las salidas originalmente asociada a la discriminación inequívoca como una segunda salida para  $\rho_2$ . En este escenario se considera el caso  $N = 20$  con  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (5, 0)$ , lo que da lugar a una salida puramente binaria y permite estudiar el régimen de discriminación de mínimo error, como se muestra en la Fig. 20. Con esta nueva asignación, la tasa de éxito del clasificador alcanza un valor de  $P_{\text{suc}} = 0.90$ , mucho más cercano a la cota teórica máxima.

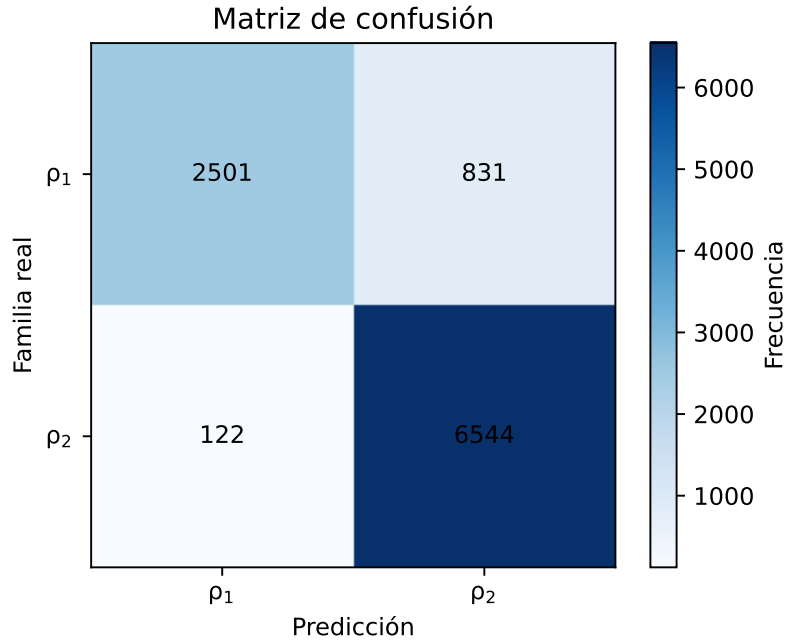


Figura 20: Matriz de confusión correspondiente al Caso 4 con  $N = 20$  y  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (5, 0)$ , lo que da lugar a una salida puramente binaria y permite analizar el régimen de discriminación de mínimo error (MESD). En esta configuración, una de las salidas originalmente asociadas a la discriminación inequívoca fue reasignada como segunda salida para la familia  $\rho_2$ , ampliando el subespacio accesible para su identificación.

Este comportamiento sugiere que la configuración original imponía una restricción en el conjunto de mediciones accesibles por la arquitectura. Al aumentar el rango efectivo de  $\Pi_2$ , se amplía el subespacio proyectivo disponible. En consecuencia, la mejora observada no debe interpretarse como una simple igualdad entre los rangos de  $\Pi_2$  y  $\rho_2$ , sino como una expansión efectiva del conjunto de POVMs realizables. Esto sugiere que la limitación original no estaba vinculada a la capacidad de aprendizaje del *ansatz*, sino a una restricción en la clase de mediciones que la arquitectura podía implementar.

Por lo anterior, se optó por mantener la modificación en la asignación de las salidas del clasificador. Los resultados obtenidos con esta modificación se muestran en la Tabla 6.

|                  | $b \approx \frac{1}{\sqrt{2}} \quad b \in [0, 1]$ |      |
|------------------|---|------|
| $a \approx 0.25$ | 0.98  | 0.99 |
| $a \in [0, 1]$   | 0.90  | 0.90 |

Cuadro 6: Promedio de la probabilidad de éxito  $P_{\text{suc}}$  en el régimen de discriminación de mínimo error (MESD) para el clasificador con la reasignación de una de las salidas originalmente asociadas a la discriminación inequívoca como segunda salida para la familia  $\rho_2$ .

*B. Clasificador para las Familias de Estados EJM, Popescu y Werner*

Con el fin de evaluar la capacidad de generalización del clasificador, antes de abordar simultáneamente las tres familias de estados consideradas, se analizó su desempeño en escenarios más controlados, correspondientes a la discriminación por pares de familias. En particular, se estudió el comportamiento del clasificador para los pares EJM-Popescu, EJM-Werner y Popescu-Werner.

Para cada uno de estos casos, se realizó el mismo análisis descrito en la sección anterior, calculando las tasas óptimas de discriminación tanto en el régimen de error mínimo como en el de discriminación inequívoca. Cabe destacar que, para los tres pares de familias considerados, la fidelidad promedio entre los estados resultó ser la misma, con un valor  $F = 0.25$ , lo que indica un solapamiento comparable entre los distintos escenarios de discriminación.

1) *Tasas de éxito teóricas.* A continuación, se resumen las probabilidades óptimas de éxito obtenidas a partir de los programas semidefinidos para cada par de familias considerado.

Los resultados obtenidos para la discriminación por pares de familias se resumen en la Tabla 7. En todos los casos considerados, la discriminación inequívoca resulta imposible, mientras que en el régimen de error mínimo se observan diferencias significativas en la probabilidad óptima de éxito según el par de familias analizado.

Cuadro 7: Probabilidades óptimas de éxito para la discriminación por pares de familias de estados (EJM–Popescu, EJM–Werner y Popescu–Werner), obtenidas mediante programación semidefinida. Las columnas corresponden a la probabilidad máxima de éxito en el régimen de discriminación inequívoca ( $P_{\text{USD}}^*$ ) y en el régimen de mínimo error ( $P_{\text{MESD}}^*$ ), respectivamente. En todos los pares considerados se obtiene  $P_{\text{USD}}^* = 0$ , lo que indica que no es posible realizar discriminación inequívoca entre estas familias, mientras que los valores de  $P_{\text{MESD}}^*$  reflejan la distinguibilidad alcanzable en el régimen de error mínimo.

| Par de familias  | $P_{\text{USD}}^*$ | $P_{\text{MESD}}^*$ |
|------------------|--------------------|---------------------|
| EJM - Popescu    | 0                  | 0.50                |
| EJM - Werner     | 0                  | 0.69                |
| Popescu - Werner | 0                  | 0.69                |

Además del análisis por pares de familias, se estudió el caso en el que el discriminador debe identificar simultáneamente estados pertenecientes a las tres familias EJM, Popescu y Werner. Los resultados correspondientes se muestran en la Tabla 11. Este escenario representa una tarea de clasificación más exigente, al involucrar un mayor número de hipótesis y una mayor superposición entre los conjuntos de estados considerados.

En este caso, la formulación semidefinida indica que la discriminación inequívoca es nuevamente imposible. Por su parte, en el régimen de discriminación de mínimo error se observa una disminución significativa en la probabilidad óptima de éxito, reflejando la dificultad intrínseca del problema al intentar discriminar simultáneamente entre las tres familias.

Cuadro 8: Probabilidades óptimas de éxito para la discriminación simultánea entre las tres familias de estados EJM, Popescu y Werner, obtenidas mediante programación semidefinida. Las columnas corresponden a la probabilidad máxima en el régimen de discriminación inequívoca ( $P_{\text{USD}}$ ) y en el régimen de discriminación de mínimo error ( $P_{\text{MESD}}$ ), respectivamente. El resultado  $P_{\text{USD}} = 0$  indica que no es posible realizar discriminación inequívoca entre las tres familias, mientras que  $P_{\text{MESD}} = 0.47$  establece el límite óptimo alcanzable cuando se permite una tasa de error mínima.

| Escenario              | $P_{\text{USD}}$ | $P_{\text{MESD}}$ |
|------------------------|------------------|-------------------|
| EJM - Popescu - Werner | 0                | 0.47              |

---

Estos resultados establecen las cotas teóricas de referencia para el análisis posterior del comportamiento del clasificador variacional. Asimismo, ponen de manifiesto cómo el aumento en el número de familias a discriminar impacta negativamente en la probabilidad máxima de éxito alcanzable en el régimen de mínimo error.

2) *Clasificación por pares de familias.* El comportamiento del clasificador para los distintos pares de familias (EJM-Popescu, EJM-Werner y Popescu-Werner) no es muy diferente. En todos los escenarios considerados, la discriminación inequívoca resulta imposible y se identifican dos comportamientos característicos del clasificador.

En primer lugar, para ciertos valores de los hiperparámetros, el proceso de entrenamiento converge hacia soluciones en las que el clasificador asigna sistemáticamente la mayoría de los estados a resultados inconclusos. Si bien este comportamiento permite minimizar la función de costo, no corresponde a un esquema de clasificación informativo, ya que no se produce una discriminación efectiva entre las familias consideradas. Un ejemplo representativo de este fenómeno se muestra en la Fig. 17, donde se observa una maximización de la tasa de inconclusión sin evidencia de aprendizaje significativo.

En segundo lugar, se identifican escenarios en los que el clasificador no converge claramente hacia un régimen de discriminación de mínimo error ni hacia un régimen de discriminación inequívoca. En estos casos, el modelo presenta una mezcla considerable de resultados erróneos e inconclusos, sin una preferencia definida por un régimen de operación específico. Este comportamiento se ilustra en la matriz de confusión de la Fig. 21, correspondiente al caso EJM-Werner con  $N = 1000$  y  $(\alpha_{\text{err}}, \alpha_{\text{inc}}) = (10, 4)$ .

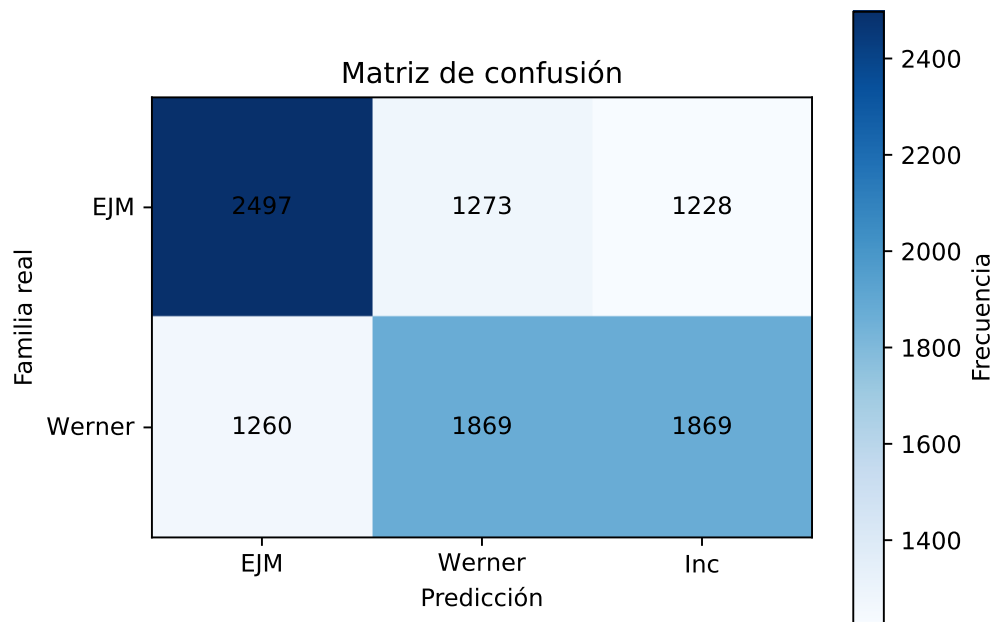


Figura 21: Matriz de confusión correspondiente al escenario EJM–Werner ( $N = 1000$ ) con  $(\alpha_{err}, \alpha_{inc}) = (10, 4)$ . Se observa una distribución significativa de resultados inconclusos en ambas familias, junto con una fracción no despreciable de errores de clasificación cruzada. Este comportamiento indica que, bajo esta elección de hiperparámetros, el clasificador no converge a una discriminación inequívoca estricta, sino que adopta una estrategia intermedia en la que coexisten decisiones binarias e inconclusas.

Para aquellas configuraciones del clasificador que lograron converger hacia soluciones de discriminación de mínimo error (MESD), se analizaron las tasas de éxito alcanzadas en la discriminación por pares de familias. En la Tabla 9 se resumen los valores promedio obtenidos para cada uno de los pares considerados.

Sin embargo, de manera análoga a lo observado para las familias de validación, al comparar estos resultados con las cotas teóricas presentadas en la Tabla 7 se evidencia una discrepancia en las tasas de éxito alcanzadas por el clasificador. En particular, únicamente para el par EJM-Popescu se obtiene un valor consistente con la cota teórica, mientras que para los pares EJM-Werner y Popescu-Werner se observa una diferencia del orden del 10% en la probabilidad de éxito.

Cabe destacar que las tasas de éxito obtenidas se mantienen cercanas al valor esperado para una clasificación aleatoria binaria, lo que indica que el clasificador no logra realizar una discriminación efectiva entre las familias consideradas.

Cuadro 9: Tasas de éxito  $P_{\text{suc}}$  obtenidas por el clasificador variacional en el régimen de discriminación de mínimo error (MESD) para los distintos pares de familias de estados. Los valores pueden compararse con las probabilidades óptimas  $P_{\text{MESD}}^*$  obtenidas mediante programación semidefinida (véase Tabla 7). Se observa que el clasificador alcanza el valor óptimo en el caso EJM–Popescu y presenta una desviación significativa respecto al límite teórico en los pares EJM–Werner y Popescu–Werner, lo que indica una mayor dificultad de aprendizaje en estos escenarios.

| <b>Par de familias</b> | <b><math>P_{\text{suc}}</math></b> |
|------------------------|------------------------------------|
| EJM - Popescu          | 0.50                               |
| EJM - Werner           | 0.50                               |
| Popescu - Werner       | 0.51                               |

No obstante, este comportamiento no puede atribuirse a una falla del proceso de entrenamiento ni a la convergencia hacia una meseta del paisaje de optimización. El clasificador fue entrenado en múltiples ocasiones utilizando los mismos hiperparámetros, pero con distintas inicializaciones aleatorias de los parámetros variacionales, observándose en todos los casos una convergencia consistente hacia una misma solución final. Asimismo, el entrenamiento se llevó a cabo empleando distintos tamaños del conjunto de datos, sin que se evidenciaran cambios significativos en el desempeño alcanzado.

La convergencia estable del proceso de optimización puede apreciarse en la Fig. 22, donde se muestra la evolución de la función de costo tanto para el conjunto de entrenamiento como para el de validación. La coincidencia entre ambas curvas y su estabilización en un mismo valor indican que el modelo converge de manera robusta y descartan problemas de sobreajuste o falta de convergencia.

De manera consistente, la Fig. 23 muestra la evolución de las tasas de éxito, error e inconclusión durante el entrenamiento. En ella se observa que estas métricas se estabilizan tras 300 iteraciones, presentando un comportamiento prácticamente idéntico en los conjuntos de entrenamiento y validación.

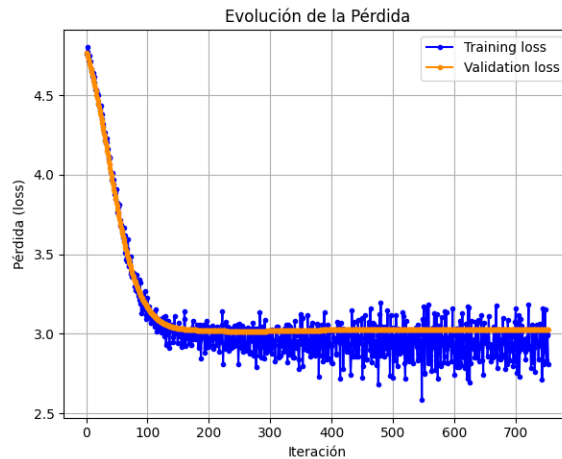


Figura 22: Evolución de la función de costo durante el entrenamiento y la validación para el escenario EJM–Werner ( $N = 1000$ ) con  $(\alpha_{err}, \alpha_{inc}) = (5, 10)$ . Se observa una disminución rápida de la pérdida en las primeras iteraciones, seguida de una estabilización alrededor de un valor casi constante, lo que indica convergencia del proceso de optimización. La cercanía entre las curvas de entrenamiento y validación sugiere ausencia de sobreajuste bajo esta elección de hiperparámetros.

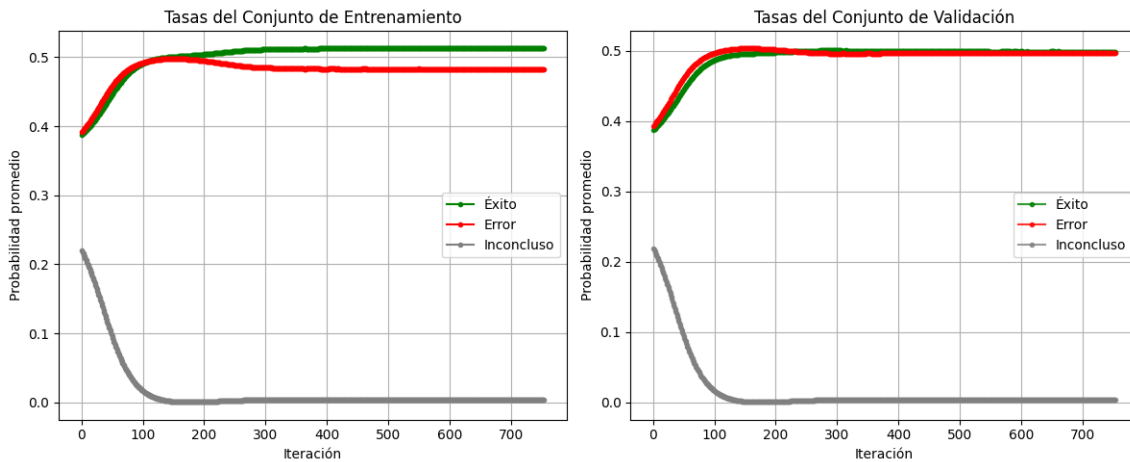


Figura 23: Evolución de las tasas de éxito, error e inconclusión en los conjuntos de entrenamiento (izquierda) y validación (derecha) para el escenario EJM–Werner ( $N = 1000$ ) con  $(\alpha_{err}, \alpha_{inc}) = (5, 10)$ . La tasa de éxito aumenta progresivamente hasta estabilizarse, mientras que la tasa de inconclusión disminuye rápidamente hacia valores cercanos a cero. Este comportamiento es consistente con la penalización dominante de los resultados inconclusos ( $\alpha_{inc} > \alpha_{err}$ ).

En conjunto, estos resultados indican que, si bien el clasificador aprende una solución estable y generalizable que minimiza la función de costo, dicha solución no es óptima, y además, no permite implementar una estrategia de discriminación informativa

entre las familias de estados consideradas. Como consecuencia, el desempeño alcanzado es estadísticamente equivalente al de una clasificación aleatoria bajo la arquitectura del clasificador y la función de costo empleadas.

Debido a esta diferencia con respecto a las tasas teóricas de éxito, y siguiendo la misma estrategia adoptada para las familias de validación, se optó por modificar la configuración del clasificador eliminando explícitamente la posibilidad de obtener resultados inconclusos. Bajo esta nueva configuración, el clasificador opera directamente en un régimen de discriminación de mínimo error. Los resultados obtenidos en estas condiciones se resumen en la Tabla 10.

Cuadro 10: Tasas de éxito  $P_{\text{suc}}$  obtenidas por el clasificador variacional en el régimen de discriminación de mínimo error (MESD), bajo una configuración estrictamente binaria en la que no se permiten resultados inconclusos. Los valores se reportan para los distintos pares de familias de estados y pueden compararse con las probabilidades óptimas  $P_{\text{ME}}^*$  obtenidas mediante programación semidefinida (véase Tabla 7). Se observa una mejora sustancial en los pares EJM–Werner y Popescu–Werner respecto a la configuración anterior.

| <b>Par de familias</b> | <b><math>P_{\text{suc}}</math></b> |
|------------------------|------------------------------------|
| EJM - Popescu          | 0.50                               |
| EJM - Werner           | 0.63                               |
| Popescu - Werner       | 0.61                               |

### *C. Resultados del Clasificador en el Escenario de Tres Familias*

Finalmente, se modificó la configuración del clasificador para permitir la discriminación simultánea entre las tres familias de estados EJM, Popescu y Werner, operando directamente en el régimen de discriminación de mínimo error. Al igual que en el análisis por pares de familias, se eliminó explícitamente la posibilidad de obtener resultados inconclusos, de modo que cada medición condujera a una decisión entre las tres clases.

La probabilidad de éxito promedio obtenida por el clasificador bajo esta configuración se resume en la Tabla 11.

La matriz de confusión mostrada en la Fig. 24 permite analizar en detalle el

Cuadro 11: Tasa de éxito promedio  $P_{\text{suc}}$  obtenida por el clasificador variacional en la discriminación simultánea entre las familias EJM, Popescu y Werner, en el régimen de discriminación de mínimo error (MESD). El valor puede compararse con la probabilidad óptima  $P_{\text{MESD}} = 0.47$  obtenida mediante programación semidefinida (véase Tabla 11).

| <b>Escenario</b>       | <b><math>P_{\text{suc}}</math></b> |
|------------------------|------------------------------------|
| EJM - Popescu - Werner | 0.33                               |

comportamiento del clasificador en el escenario de discriminación simultánea entre las familias EJM, Popescu y Werner.

En particular, el clasificador asigna la mayoría de los estados a la clase Werner, independientemente de si el estado de entrada pertenece a las familias EJM, Popescu o Werner. Las clases Popescu y EJM aparecen únicamente como predicciones secundarias, mientras que la probabilidad de obtener un resultado inconcluso es prácticamente nula.

Desde el punto de vista de la discriminación, este comportamiento indica que el clasificador no logra aprender una regla de decisión efectiva. En su lugar, converge a una solución caracterizada por una distribución de salidas casi fija, que permite minimizar la función de costo sin capturar la estructura del problema de discriminación cuántica.

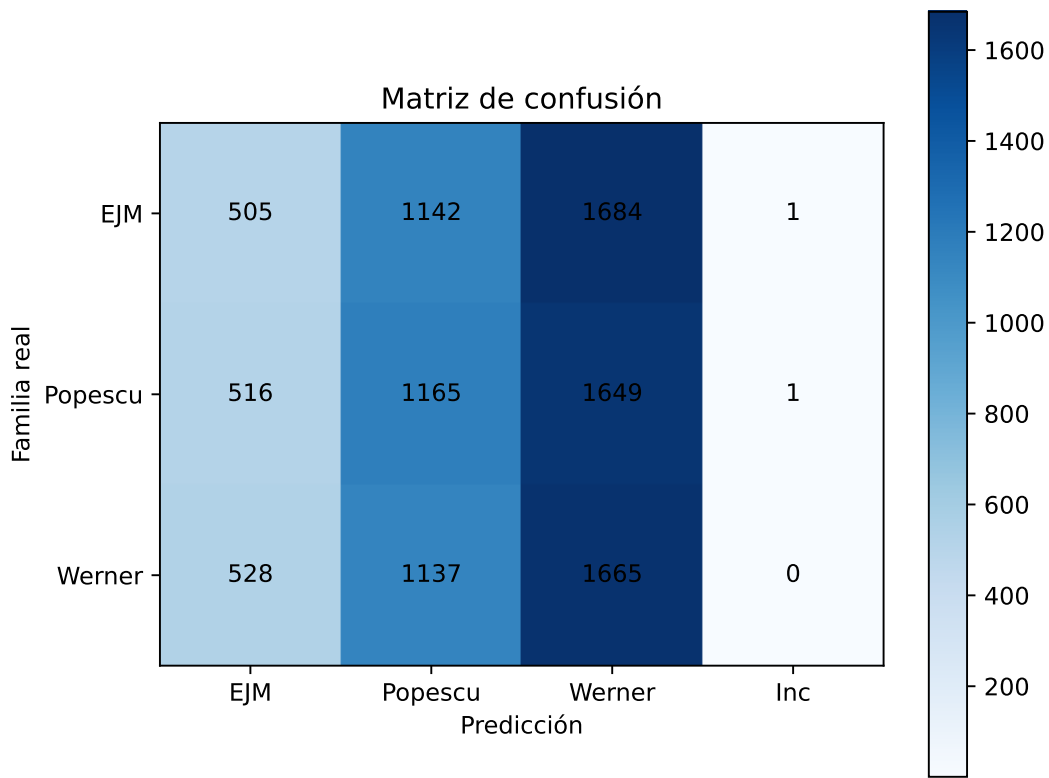


Figura 24: Matriz de confusión del escenario EJM–Popescu–Werner ( $N = 600$ ) con  $(\alpha_{err}, \alpha_{inc}) = (5, 20)$ . Se observa un fuerte sesgo hacia la clase Werner, que concentra la mayoría de las predicciones independientemente de la familia real. Las clasificaciones correctas (elementos diagonales) son comparativamente menores.

En este escenario, la estrategia empleada para las familias de validación, consistente en eliminar explícitamente los resultados inconclusos y forzar al clasificador a operar en el régimen de discriminación de mínimo error, no resulta suficiente para alcanzar las cotas teóricas óptimas. De hecho, aun cuando esta misma configuración se aplica a la discriminación por pares de familias, los resultados obtenidos, resumidos en la Tabla 9, presentan discrepancias apreciables con respecto a las cotas teóricas. En particular, se observa una diferencia del orden del 6 % en la probabilidad de éxito para el par EJM-Werner y del 8 % para el par Popescu-Werner.

Estos resultados sugieren que la limitación observada no está asociada a problemas de optimización ni a una convergencia inadecuada del entrenamiento, sino que podría estar relacionada con ciertas características estructurales del problema considerado. En particular, se observa una posible falta de correspondencia entre la geometría de las

familias de estados cuánticos que se desean discriminar y la forma en que el clasificador asigna sus salidas. Tal como se discutió previamente para las familias de validación, esta situación puede interpretarse desde un punto de vista geométrico, en términos del rango y del soporte de los estados cuánticos involucrados y de los operadores de medición implementados.

## CAPÍTULO VII: CONCLUSIONES Y PERSPECTIVAS

A partir del análisis realizado para las familias de validación, se puede concluir que el clasificador cuántico variacional implementado presenta un desempeño robusto en el régimen de discriminación inequívoca (USD). En este esquema, el modelo logra aproximarse de manera consistente a las cotas teóricas obtenidas mediante programación semidefinida, alcanzando resultados comparables con los reportados en [17]. Asimismo, se observó que la convergencia es estable frente a distintas inicializaciones y tamaños de conjunto de entrenamiento, lo que evidencia la capacidad del *ansatz* para implementar estrategias cercanas a las óptimas cuando la arquitectura es compatible con la estructura del problema.

Sin embargo, la configuración originalmente propuesta del clasificador no alcanza de manera óptima las tasas teóricas en el régimen de discriminación de mínimo error (MESD). Si bien el modelo converge correctamente hacia soluciones de mínimo error, es decir, sin generar resultados inconclusos, las probabilidades de éxito alcanzadas presentan diferencias significativas respecto al límite de Helstrom. Este comportamiento no se atribuye a fallas del proceso de entrenamiento ni a problemas de optimización, sino a una restricción estructural en la clase de mediciones realizables por la arquitectura considerada.

De lo anterior, es posible concluir que la asignación de salidas del clasificador no puede realizarse de manera arbitraria. Inicialmente podría suponerse, como se sugiere en [19], que dicha asignación puede elegirse de forma aleatoria sin afectar el desempeño del clasificador. Sin embargo, los resultados obtenidos indican que esta decisión debe fundamentarse en la estructura geométrica de las familias de estados a discriminar, particularmente en el rango y el soporte de sus operadores densidad.

En el caso estudiado, la familia  $\rho_1(a)$  está compuesta por estados puros de rango uno, mientras que  $\rho_2(b)$  corresponde a estados mixtos de rango dos. La asignación

---

original, que otorgaba un mayor número de salidas a  $\rho_1$ , introducía una asimetría que limitaba la capacidad del modelo para capturar completamente el soporte de los estados mixtos. En el régimen USD esta restricción no afectó los resultados, ya que la presencia de resultados inconclusos permite absorber la ambigüedad asociada a las componentes del estado que no pueden identificarse sin error. No obstante, en el régimen MESD, esta limitación implica un cuello de botella geométrico que impide alcanzar la cota teórica óptima.

Al modificar la asignación de salidas de modo que el operador de medición asociado a  $\rho_2$  tuviera un rango efectivo mayor, se amplió el conjunto de POVMs realizables por la arquitectura. Como resultado, el clasificador logró aproximarse significativamente a las cotas teóricas de mínimo error.

En cuanto a la discriminación entre las familias EJM, Popescu y Werner, los resultados muestran un escenario más restrictivo. Tanto el análisis teórico como los experimentos numéricos indican que la discriminación inequívoca es imposible en estos casos. En el régimen de mínimo error, aunque el clasificador converge de manera estable, las tasas de éxito obtenidas no alcanzan las cotas óptimas y, en algunos escenarios, son básicamente una clasificación aleatoria. La estabilidad del proceso de entrenamiento sugiere que la limitación observada podría estar asociada a la geometría intrínseca de las familias consideradas y a la capacidad de la arquitectura para representar las mediciones óptimas correspondientes.

Se concluye que el diseño de clasificadores cuánticos variacionales para tareas de discriminación de estados no depende únicamente de la elección del ansatz o de la función de costo, sino que requiere un análisis cuidadoso de la correspondencia entre la estructura geométrica de los estados y la clase de mediciones que la arquitectura puede implementar. La compatibilidad entre rango, soporte y efectos de medición es un criterio necesario para alcanzar las cotas teóricas de discriminación.

Como perspectivas futuras, resulta natural explorar arquitecturas que permitan una implementación más general de POVMs, por ejemplo mediante extensiones con ancillas adicionales. Asimismo, sería interesante analizar la relación entre el rango de

los estados a discriminar y el número mínimo de salidas necesarias para implementar mediciones óptimas dentro de un esquema variacional.

Asimismo, la estructura geométrica abre la posibilidad de extender el enfoque hacia esquemas de aprendizaje no supervisado. En este contexto, la organización de los estados en el espacio de Hilbert podría surgir de manera natural a partir de criterios variacionales basados en distancias o fidelidades, en lugar de depender de etiquetas previamente definidas.

Por otra parte, una línea de investigación especialmente relevante consiste en evaluar el desempeño de estas arquitecturas en dispositivos cuánticos reales. Analizar su comportamiento frente a ruido, errores de compuerta y procesos de decoherencia permitiría determinar hasta qué punto las restricciones geométricas observadas se intensifican o, por el contrario, se atenúan en el régimen NISQ. La implementación experimental de estos clasificadores no solo permitiría examinar su viabilidad práctica, sino también delimitar con mayor precisión los alcances y limitaciones de los métodos variacionales en tareas de discriminación cuántica.

## REFERENCIAS

- [1] A. Chefles, “Quantum state discrimination,” *Contemp. Phys.*, vol. 41, no. 6, pp. 401–424, 2000.
- [2] S. M. Barnett and S. Croke, “Quantum state discrimination,” *Adv. Opt. Photon.*, vol. 1, 2009.
- [3] Y. Lu, N. Coish, R. Kaltenbaek, D. R. Hamel, S. Croke, and K. J. Resch, “Minimum-error discrimination of entangled quantum states,” *Phys. Rev. A*, vol. 82, no. 4, 2010.
- [4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. Cambridge University Press, 2010.
- [5] V. Giovannetti, S. Lloyd, and L. Maccone, “Advances in quantum metrology,” *Nature Photonics*, vol. 5, no. 4, pp. 222–229, 2011.
- [6] P. Raynal, N. Lütkenhaus, and S. J. van Enk, “Reduction theorems for optimal unambiguous state discrimination of density matrices,” *Phys. Rev. A*, vol. 68, no. 2, 2003.
- [7] J. J. Sakurai and J. Napolitano, *Modern Quantum Mechanics*, 2nd ed. Cambridge University Press, 2017.
- [8] S. M. Barnett and E. Riis, “Experimental demonstration of polarization discrimination at the Helstrom bound,” *J. Mod. Opt.*, vol. 44, no. 6, pp. 1061–1064, 1997.
- [9] C. W. Helstrom, “Quantum Detection and Estimation Theory,” *J. Stat. Phys.*, vol. 1, no. 2, 1969.
- [10] A. S. Holevo, “Statistical decision theory for quantum systems,” *J. Multivariate Anal.*, vol. 3, no. 4, pp. 337–394, 1973.

- 
- [11] I. D. Ivanovic, “How to differentiate between non-orthogonal states,” *Phys. Lett. A*, vol. 123, no. 6, pp. 257–259, 1987.
- [12] D. Dieks, “Overlap and distinguishability of quantum states,” *Phys. Lett. A*, vol. 126, no. 5, 1988.
- [13] A. Peres, “How to differentiate between non-orthogonal states,” *Phys. Lett. A*, vol. 128, no. 1, 1988.
- [14] R. B. M. Clarke, V. M. Kendon, A. Chefles, S. M. Barnett, E. Riis, and M. Sasaki, “Experimental realization of optimal detection strategies for overcomplete states,” *Phys. Rev. A*, vol. 64, no. 1, 2001.
- [15] S. Gómez, E. S. Gómez, O. Jiménez, A. Delgado, S. P. Walborn, and G. Lima, “Experimental quantum state discrimination using the optimal fixed rate of inconclusive outcomes strategy,” *Sci. Rep.*, vol. 12, no. 1, 2022.
- [16] R. P. Feynman, “Simulating Physics with Computers,” *Int. J. Theor. Phys.*, vol. 21, no. 6/7, pp. 467–488, 1982, based on the keynote at the Physics of Computation Conference, 1981.
- [17] H. Chen, L. Wossnig, S. Severini, H. Neven, and M. Mohseni, “Universal discriminative quantum neural networks,” *Quantum Mach. Intell.*, vol. 3, no. 1, 2021.
- [18] J. Preskill, “Quantum Computing in the NISQ era and beyond,” *Quantum*, vol. 2, p. 79, 2018.
- [19] A. Patterson, H. Kelly, S. Gowrishankar, S. Guha, and P. Narang, “Quantum state discrimination using noisy quantum neural networks,” *Phys. Rev. Research*, vol. 3, no. 1, 2021.
- [20] Y. S. Kim, J. C. Lee, O. Kwon, and Y. H. Kim, “Protecting entanglement from decoherence using weak measurement and quantum measurement reversal,” *Nature Physics*, vol. 8, no. 2, pp. 117–120, 2012.

- 
- [21] M. Schlosshauer, *Decoherence and the Quantum-to-Classical Transition*. Springer, 2007.
- [22] S. Fortin, “Hacia una mejor comprensión de la decoherencia desde una perspectiva general,” *Rev. Colomb. Filos. Cienc.*, vol. 12, no. 24, pp. 65–82, 2012.
- [23] IBM Quantum, “Variational Algorithms,” [Online]. Available: <https://quantum.cloud.ibm.com/learning/es/courses/variational-algorithm-design/variational-algorithms>, 2023, accessed: Apr. 14, 2025.
- [24] H. Chen, L. Wossnig, S. Severini, H. Neven, and M. Mohseni, “Universal discriminative quantum neural networks,” *Quantum Mach. Intell.*, vol. 3, p. 1, 2021.
- [25] D. S. Acosta, “Propiedades del espectro y rango numérico de un operador,” 2024, dirigido por T. J. Bermúdez de León.
- [26] S. Halder and R. Augusiak, “Local unambiguous unidentifiability, entanglement generation, and Hilbert space splitting,” *arXiv preprint arXiv:2503.05673*, 2025. [Online]. Available: <https://arxiv.org/abs/2503.05673>
- [27] N. Johnston and C.-K. Li, “Approximating quantum states by states of low rank,” *arXiv preprint arXiv:2510.08463*, 2026. [Online]. Available: <https://arxiv.org/abs/2510.08463>
- [28] J. Watrous, *The Theory of Quantum Information*. Cambridge University Press, 2018.
- [29] M. M. Wilde, *Quantum Information Theory*, 2nd ed. Cambridge University Press, 2017.
- [30] R. Iten, R. Colbeck, I. Kukuljan, J. Home, and M. Christandl, “Quantum circuits for isometries,” *Phys. Rev. A*, vol. 93, no. 3, 2016.
- [31] D. Lee, K. Baek, J. Huh, and D. K. Park, “Variational quantum state discriminator for supervised machine learning,” *Quantum Sci. Technol.*, vol. 8, no. 4, p. 045023, 2023.

- 
- [32] Y. S. Yordanov and C. H. W. Barnes, “Implementation of a general single-qubit positive operator-valued measure on a circuit-based quantum computer,” *Phys. Rev. A*, vol. 100, no. 6, p. 062317, 2019.
- [33] P. Skrzypczyk and D. Cavalcanti, *Semidefinite Programming in Quantum Information Science*. IOP Publishing, 2023.
- [34] M. Sasaki, “Two examples of measurement processes illustrating Helstrom’s optimum decision bound,” Tech. Rep., 1996.
- [35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [36] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.
- [37] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, “Variational Quantum Algorithms,” *Nat. Rev. Phys.*, vol. 3, pp. 625–654, 2021.
- [38] Xanadu Quantum Technologies Inc., “Parametrized Quantum Circuits - PennyLane Codebook,” [Online]. Available: <https://pennylane.ai/codebook/variational-quantum-algorithms/parametrized-quantum-circuits>, 2025, accessed: Oct. 7, 2025.
- [39] E. H. Houssein, Z. Abohashima, M. Elhoseny, and W. M. Mohamed, “Machine learning in the quantum realm: The state-of-the-art, challenges, and future vision,” *Expert Syst. Appl.*, vol. 194, p. 116512, 2022.
- [40] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.
- [41] R. Iten, R. Colbeck, and M. Christandl, “Quantum circuits for quantum channels,” *Phys. Rev. A*, vol. 95, no. 5, p. 052316, 2017.
- [42] N. Gisin, “Entanglement 25 years after Quantum Teleportation: testing joint measurements in quantum networks,” *Entropy*, vol. 21, no. 3, p. 325, 2019.

- 
- [43] A. Tavakoli, N. Gisin, and C. Branciard, “Bilocal Bell Inequalities Violated by the Quantum Elegant Joint Measurement,” *Phys. Rev. Lett.*, vol. 126, no. 22, 2021.
- [44] J. Pauwels and N. Gisin, “The Multiqubit Elegant Joint Measurement,” *arXiv preprint arXiv:2509.01852*, 2025. [Online]. Available: <http://arxiv.org/abs/2509.01852>
- [45] C.-X. Huang, X.-M. Hu, Y. Guo, C. Zhang, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, N. Gisin, C. Branciard, and A. Tavakoli, “Entanglement swapping and quantum correlations via Elegant Joint Measurements,” *Phys. Rev. Lett.*, vol. 129, no. 3, 2022.
- [46] N. Gisin, “The Elegant Joint Quantum Measurement and some conjectures about N-locality in the Triangle and other Configurations,” arXiv:1708.05556, 2017. [Online]. Available: <https://arxiv.org/abs/1708.05556>
- [47] R. F. Werner, “Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model,” *Phys. Rev. A*, vol. 40, no. 8, pp. 4277–4281, 1989.
- [48] Xanadu, “PennyLane Documentation,” [Online]. Available: <https://docs.pennylane.ai/en/stable/>, 2024, accessed: May 20, 2024.
- [49] G. Sagnol and M. Stahlberg, “PICOS: A Python interface to conic optimization solvers,” *J. Open Source Softw.*, vol. 7, no. 70, 2022.
- [50] M. Mohseni, A. M. Steinberg, and J. A. Bergou, “Optical realization of optimal unambiguous discrimination for pure and mixed quantum states,” *Phys. Rev. Lett.*, vol. 93, no. 20, 2004.