



Tomografía Autoguiada con Procesamiento Clásico

Tesis para optar al grado de
Magíster en Ciencias con Mención en Física
Facultad de Ciencias Físicas y Matemáticas

por
Leonardo Hans Isaías Zambrano Valenzuela

Abril, 2018

Director de Tesis: **Dr. Aldo Delgado Hidalgo**

AGRADECIMIENTOS

A mi tutor, Dr. Aldo Delgado, por su ánimo, tiempo y dedicación para conseguir este trabajo. Y por su paciencia durante estos años.

A mis padres, por apoyarme y estar siempre en los momentos mas importantes de mi vida.

A mis amigos, por recorrer juntos este camino.



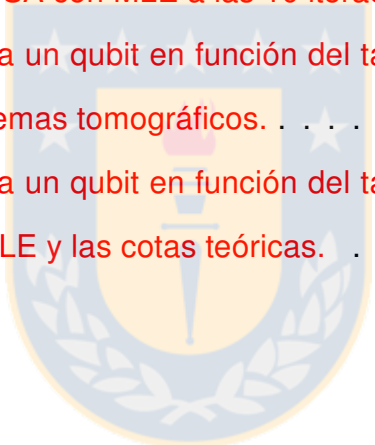
Tabla de Contenido

Índice de Tablas	v
Índice de Ilustraciones	vi
Resumen	viii
1 Introducción	1
2 Conceptos Fundamentales	5
2.1 Probabilidad	5
2.2 Método de Montecarlo	12
2.3 Álgebra Lineal	14
2.4 Proceso de Ortogonalización de Gram-Schmidt	19
2.5 Medida de Haar	20
3 Mecánica Cuántica	21
3.1 Postulados de la Mecánica Cuántica	21
3.2 Matriz Densidad	24
3.3 Descomposición de Schmidt	26
3.4 Medidas de Distancia	27
3.5 Tomografía Cuántica	30

3.5.1 Tomografía Estándar	30
3.5.2 Tomografía SIC-POVM	32
3.5.3 Tomografía MUBs	33
3.6 Estimación por Máxima Verosimilitud	33
3.7 Cota de Gill-Massar	36
4 Gradiente Descendente	41
4.1 Algoritmo Básico de Gradiente Descendente	41
4.2 CSPA-SPSA	43
5 Tomografía Autoguiada	49
6 Tomografía Autoguiada con Procesamiento Clásico	54
6.1 Algoritmo	55
6.2 Resultados	62
7 Detección de Entrelazamiento	75
8 Conclusión	80
Bibliografía	85
9 Apéndice	88

Índice de Tablas

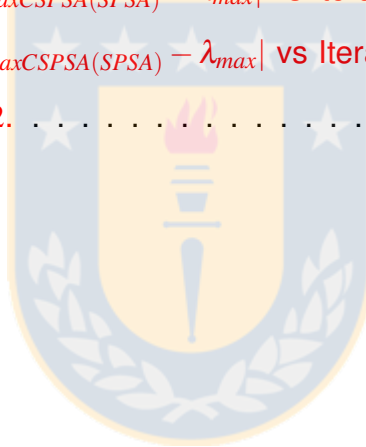
6.1	Infidelidad para un qubit en función del tamaño de ensemble para CSPSA y CSPSA con MLE a las 10 iteraciones.	66
6.2	Infidelidad para un qubit en función del tamaño de ensemble para distintos esquemas tomográficos.	67
6.3	Infidelidad para un qubit en función del tamaño de ensemble para CSPSA con MLE y las cotas teóricas.	68



Índice de Ilustraciones

4.1	Gráfico del método de gradiente descendente para la función $f(x) = x^2$.	42
5.1	Gráfico de Tomografía Autoguiada usando CSPSA en la esfera de Bloch para distintas trayectorias.	51
5.2	Gráfico de la Infidelidad Promedio vs Iteraciones para un qubit usando CSPSA y SPSA.	53
6.1	Diagrama de flujo de CSPSA con MLE.	56
6.2	Gráfico de la Infidelidad Promedio vs Iteraciones para un qubit usando CSPSA con MLE.	62
6.3	Gráfico de la Infidelidad Mediana vs Iteraciones para un qubit usando CSPSA con MLE.	63
6.4	Gráfico de la Distribución de infidelidades para un qubit a la iteración 10 usando CSPSA con MLE sobre 10^4 repeticiones.	64
6.5	Gráfico de la Infidelidad Promedio vs Iteraciones para dimensiones 32, 16, 8, 4 y 2 usando CSPSA con MLE.	65
6.6	Gráfico de la Infidelidad Promedio vs Iteraciones para un qubit usando CSPSA con MLE comparado con la cota de Cramer-Rao.	69

6.7	Gráfico de la Infidelidad Promedio vs Iteraciones para un qubit usando CSPA con MLE comparado con la cota de Gill-Massar. . . .	70
6.8	Gráfico de la Infidelidad Promedio vs Iteraciones para un qubit usando CSPA con MLE comparado con la cota de Gill-Massar de estados puros.	71
6.9	Gráfico de la Infidelidad Promedio vs Iteraciones para un estado cuasi puro con pureza 99.9% usando CSPA con MLE.	73
7.1	Gráfico de $ \lambda_{maxCSPA(SP)} - \lambda_{max} $ vs Iteraciones para dos qubits. . .	76
7.2	Gráfico de $ \lambda_{maxCSPA(SP)} - \lambda_{max} $ vs Iteraciones para dimensiones 32, 16, 8, 4 y 2.	78



RESUMEN

En esta tesis modificamos Complex Simultaneous Perturbation Stochastic Approximation (CSPSA), un método de optimización sobre el campo de los complejos utilizado en tomografía autoguiada de estados cuánticos. La tomografía autoguiada es un protocolo iterativo para estados puros basado en minimizar la infidelidad entre el estado cuántico del sistema y un estimado de este. Nuestra mejora consiste en agregarle procesamiento clásico a CSPSA mediante Maximum Likelihood Estimation (MLE) entre iteraciones. Gracias a esto, observamos una gran reducción en el número de iteraciones y de recursos a utilizar, a expensas de incrementar el costo computacional clásico. Además, mostramos que se puede utilizar el mismo método de optimización CSPSA para determinar una medida en el entrelazamiento de estados para sistemas bipartitos, sin la necesidad de conocer el estado, y, por lo tanto, sin realizar tomografía.

Para probar estas afirmaciones realizamos extensas simulaciones numéricas sobre un gran número de estados desconocidos, en varias dimensiones y considerando distintos tamaños en la muestra a medir.

Capítulo 1

Introducción

Todo sistema cuántico tiene asociado un vector unitario llamado estado, que caracteriza todas las propiedades cuánticas del sistema. Es de gran interés determinar experimentalmente este estado, ya que si lo conocemos, podemos predecir el comportamiento de nuestro sistema bajo cualquier tipo de operación cuántica. A partir de esto surge la estimación de estados cuánticos, definida como el proceso de inferir el estado a partir de mediciones generalizadas, conocidas como Positive Operator-Valued Measurements (POVMs). Es un paso previo de muchos procesos en Mecánica Cuántica, como por ejemplo, Información Cuántica [1], Metrología Cuántica [2] y Computación Cuántica [3], entre otros, ya que en todas estas tareas la habilidad de determinar el estado cuántico es fundamental para completar el proceso.

Existen muchos métodos de tomografía o estimación de estados cuánticos [4], pero una característica en común que tienen todos ellos es que se necesita medir sobre un gran número de partículas idénticamente preparadas para poder conocer con precisión el estado del sistema. Esto proviene del postulado de medición, ya que al colapsar la función de onda a cualquiera de los vectores de la base

en que medimos, es imposible inferir con una sola medición el estado cuántico del sistema.

Como generalmente en el ámbito experimental los recursos disponibles son limitados, es de gran interés desarrollar protocolos de tomografía capaces de encontrar el estado cuántico de la forma mas precisa gastando el mínimo posible de recursos. Uno de estos métodos, llamado tomografía autoguiada [5], formula la estimación como un problema variacional, donde el estado desconocido del sistema es caracterizado como el parámetro de alguna función de distancia, y luego es encontrado usando algún método de optimización. Para el caso de estados puros, podemos escoger esta función como la infidelidad entre estados, $I(\psi, \phi) = 1 - |\langle \psi | \phi \rangle|^2$, donde $|\psi\rangle$ es el estado fijo del sistema y $|\phi\rangle$ es nuestro estimado. Luego, de forma iterativa, y a través de mediciones, el algoritmo va mejorando el estimado $|\phi\rangle$, disminuyendo la distancia $I(\psi, \phi)$, hasta converger al estado del sistema $|\psi\rangle$ con alguna precisión determinada. Este método aprovecha de forma óptima los recursos, alcanzando cotas en la estimación del estado órdenes mejor que algunos métodos estándar [6]. Sin embargo, el inconveniente es que el número de mediciones a realizar para alcanzar el óptimo puede resultar prohibitivo.

En vista de lo anterior, lo que proponemos es mejorar la tasa de convergencia en la tomografía autoguiada utilizando procesamiento estadístico clásico, lo cual no requiere más mediciones en el sistema pero disminuye el número de etapas en el proceso tomografía, e incluso mejora la precisión en la estimación. A diferencia de los métodos tomográficos estándar, donde se usa la estimación por

máxima verosimilitud (MLE) como postprocesamiento, nosotros utilizaremos este mismo método entre iteraciones de la tomografía autoguiada, refinando así de forma clásica el estimado $|\phi\rangle$, antes de continuar con la próxima iteración del algoritmo de optimización.

Además, mostraremos que para determinar entrelazamiento cuántico en sistemas físicos bipartitos no es necesario conocer la totalidad del estado cuántico, y por lo tanto no es necesario realizar tomografía. Esto lo lograremos aplicando un método de optimización sobre la matriz densidad reducida de uno de los dos sistemas bipartitos, lo cual nos entregará un coeficiente que mide el grado de entrelazamiento entre los sistemas, y que coincide con la medida geométrica de entrelazamiento [7].

Esta tesis está organizada como sigue: En el capítulo 2 introduciremos herramientas matemáticas que usaremos en el desarrollo del trabajo, principalmente teoría de probabilidades y álgebra lineal, lo que nos dará el marco sobre el que yace la Teoría Cuántica. Luego, en el capítulo 3 veremos conceptos básicos en Mecánica Cuántica y estimación de estados, lo que nos permitirá poner en contexto nuestro trabajo y comparar los resultados de esta tesis con esquemas previos. En el capítulo 4 mostraremos los algoritmos de optimización usados en tomografía autoguiada, dentro de los que están CSPA y SPSA. El capítulo 5 está específicamente dedicado a la tomografía autoguiada y su relación con CSPA, lo cual es la base de la que parte nuestro estudio, el cual presentamos en el capítulo 6. Aquí encontramos el principal resultado del trabajo, al derivar un protocolo que mejora la tomografía autoguiada de estados cuánticos al incluir proce-

samiento clásico de forma iterativa. Incluimos los algoritmos necesarios para la implementación y pruebas numéricas del método. Por último, en el capítulo 7 desarrollamos un algoritmo capaz de detectar entrelazamiento cuántico. Mostramos también simulaciones numéricas que prueban el desempeño del método.



Capítulo 2

Conceptos Fundamentales

En este capítulo estudiaremos conceptos básicos de la teoría de probabilidad y álgebra lineal, fundamentos matemáticos sobre los que está construida la Mecánica Cuántica. Además, presentaremos el método de Montecarlo, el cual usaremos al realizar todas las simulaciones numéricas en este trabajo, junto a la medida de Haar y el proceso de ortogonalización de Gram-Schmidt.

2.1 Probabilidad

La teoría de probabilidad entrega un modelo matemático para fenómenos aleatorios, esto es, aquellos que involucran incerteza [8]. Como ejemplos de fenómenos aleatorios tenemos lanzar un dado o una moneda: al realizar estos experimentos una y otra vez bajo las mismas condiciones tenemos como resultado posible un conjunto de alternativas determinadas.

Denominamos espacio muestral Ω al conjunto de posibles resultados de un experimento aleatorio. Cada elemento individual $\omega \in \Omega$ es llamado punto muestral. El conjunto de puntos muestrales $A \subseteq \Omega$ para el cual una sentencia dada es

válida es llamado evento. Así, Ω es una colección de posibles resultados de un experimento. Dado esto, podemos identificar una familia $\mathcal{F} = \{A_1, A_2, \dots\}$ de eventos, y una función P sobre \mathcal{F} tal que por cada A en \mathcal{F} , $P(A)$ representa la posibilidad de que un evento A ocurra.

La función medida de probabilidad $P : \mathcal{F} \rightarrow \mathbb{R}$ debe satisfacer las 3 siguientes propiedades, llamadas axiomas de probabilidad:

- $P(A) \geq 0, \forall A \in \mathcal{F}$.
- $P(\Omega) = 1$.
- Si A_1, A_2, \dots son eventos disjuntos, entonces $P(\cup_i A_i) = \sum_i P(A_i)$.

Un ejemplo de los conceptos previos es lanzar un dado: El espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$. Podemos definir distintos espacios de eventos $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. Por ejemplo, $\mathcal{F} = \{A_1 = \{1, 2, 3\}, A_2 = \{4, 5, 6\}\}$, con probabilidades $P(A_1) = P(A_2) = 1/2$.

Decimos que un evento A es imposible si $P(A) = 0$. Algunas propiedades de las probabilidades son:

- Si $A \subseteq B \implies P(A) \leq P(B)$.
- $P(A \cap B) \leq \min\{P(A), P(B)\}$.
- $P(A \cup B) \leq P(A) + P(B)$.
- $P(\Omega \setminus A) = 1 - P(A)$.

- Si A_1, \dots, A_n forman un conjunto de eventos disjuntos tal que $\cup_{i=1}^n A_i = \Omega$, entonces $\sum_{i=1}^n P(A_i) = 1$.

Dado un evento B con $P(B) \neq 0$, definimos la probabilidad condicional de un evento A dado B como

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (2.1)$$

Así, $P(A|B)$ es la probabilidad de medir el evento A después de que haya ocurrido el evento B . Dos eventos son llamados independientes si y solo si $P(A \cap B) = P(A)P(B)$, o equivalentemente $P(A) = P(A|B)$. Luego, la independencia es equivalente a decir que observar B no tiene ningún efecto en la probabilidad de A .

Una variable aleatoria X es una función $X : \Omega \rightarrow \mathbb{R}$, esto es, asigna a cada resultado $\omega \in \Omega$ un número real $X(\omega) = x$. Si el conjunto de todos los posibles valores de X es numerable, decimos que X es discreto. Si no, es continuo.

Por ejemplo, si lanzamos 5 monedas podemos tener como punto muestral $\omega_0 = \{C, S, S, C, S\}$. Sin embargo, usualmente no estamos interesados en una secuencia particular de caras y sellos, sino, por ejemplo, el número de veces que sale cara, $X(\omega_0) = 2$.

Definimos la probabilidad de obtener x como la probabilidad del evento ω tal que $X(\omega) = x$. Si X es discreta, podemos expresar esto como

$$p(X = x) = p(\{\omega : X(\omega) = x\}). \quad (2.2)$$

Si X es continua, tenemos valores dentro de intervalos, de manera que la probabilidad entre los valores a y b es

$$P(a < X \leq b) = p(\{\omega : a < X(\omega) \leq b\}). \quad (2.3)$$

Una forma simple de representar la medida de probabilidad asociada con una variable aleatoria es la distribución de probabilidad. Esta es una función que asigna a cada evento A definido sobre la variable X la probabilidad $P(A)$ de que dicho suceso ocurra.

Cuando la variable aleatoria X es discreta especificamos directamente la probabilidad de cada valor que X puede tomar. Así, la distribución de probabilidad es una función de masa de probabilidad $p_X : \Omega \rightarrow \mathbb{R}$ tal que

$$p_X(x) = P(X = x). \quad (2.4)$$

p_X tiene las siguientes propiedades:

- $0 \leq p_X(x) \leq 1$.
- $\sum_{x \in \text{Val}(X)} p_X(x) = 1$.
- $\sum_{x \in A} p_X(x) = P(X \in A)$.

$\text{Val}(X)$ es el conjunto de todos los valores que X puede tomar.

Para variables aleatorias continuas podemos definir una función densidad de probabilidad f_X , que satisface

- $f_X(x) \geq 0$.

- $\int_{-\infty}^{+\infty} f_X(x)dx = 1.$
- $\int_{x \in A} f_X(x)dx = P(X \in A).$

Una distribución de probabilidad de particular importancia en este trabajo es la binomial,

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}. \quad (2.5)$$

Modela el número de éxitos en una secuencia de n experimentos independientes, cada uno de estos con solo dos resultados son posibles (o sea, dicotómico). Llamamos éxito a uno de estos resultados, con probabilidad p , y al otro fracaso, con probabilidad $1 - p$. En Mecánica Cuántica es importante porque las mediciones sobre qubits obedecen esta distribución cuando medimos de forma ideal, esto es, no consideramos el error asociado a los detectores. Podemos generalizar este resultado cuando tenemos mediciones no dicotómicas en la distribución multinomial,

$$p(x_1, \dots, x_k) = \begin{cases} \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} & \text{cuando } \sum_{i=1}^k x_i = n. \\ 0 & \text{otros casos.} \end{cases} \quad (2.6)$$

Así, esta es la regla que seguimos cuando realizamos mediciones sobre sistemas cuánticos en mayores dimensiones.

También es importante en este trabajo la distribución normal, dada por

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (2.7)$$

con σ la desviación estándar y μ la media. Podemos encontrar esta distribución modelando caracteres tan variados como la estatura de individuos, los efectos de

un fármaco, el cociente intelectual, entre otros.

Existe una asociación entre las distribuciones anteriores. El teorema del límite central afirma que si tenemos X_1, \dots, X_n , un conjunto de variables aleatorias independientes e idénticamente distribuidas de una distribución con media μ y varianza $\sigma^2 \neq 0$, entonces, cuando n es muy grande,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.8)$$

tiene aproximadamente una distribución normal con $\mu_X = \mu$ y $\sigma_X^2 = \sigma^2/n$. Como la distribución binomial no es más que la suma de n resultados obtenidos al efectuar n repeticiones de un experimento aleatorio, o equivalentemente, la suma de n variables independientes, cumple las condiciones del teorema y se aproxima a la distribución normal cuando n es grande.

Dada una variable aleatoria X discreta con masa de probabilidad p_X , y una función $g : \mathbb{R} \rightarrow \mathbb{R}$ arbitraria, definimos el valor esperado, o esperanza de $g(X)$ como

$$E[g(X)] = \sum_{x \in \text{Val}(X)} g(x)p_X(x). \quad (2.9)$$

Si X es continua con densidad de probabilidad f_X , entonces el valor esperado de $g(X)$ es

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx. \quad (2.10)$$

El valor esperado de $g(X)$ puede ser visto como un promedio ponderado de los valores que la función $g(x)$ puede tomar en los diferentes valores de x , donde los pesos están dados por $p_X(x)$ o $f_X(x)$. Tiene las siguientes propiedades:

- Si a es una constante, $E(a) = a$.
- $E(af(X)) = aE(f(X))$ para $a \in \mathbb{R}$.
- $E[f(X) + g(X)] = E[f(X)] + E[g(X)]$.

Como caso especial, llamamos media μ a calcular la esperanza de la variable misma $E[X]$, esto es, $g(X) = x$. Es un valor típico o central para una determinada distribución de probabilidad. Así, para distribuciones discretas, $\mu = E[X] = \sum_{x \in \text{Val}(X)} x p_X(x)$, y para continuas, $\mu = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$. En un experimento podemos estimar este valor usando el promedio, $\tilde{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$.

La varianza es una medida de que tanto se alejan los valores de una variable aleatoria X de su media. Está definida por

$$\text{Var}[X] = E[(X - E[X])^2]. \quad (2.11)$$

Algunas propiedades son:

- $\text{Var}[X] \geq 0$.
- $\text{Var}[af(X) + b] = a^2 \text{Var}[f(X)]$ si $a, b \in \mathbb{R}$. De acá deducimos que la varianza de una constante es cero.
- $\text{Var}[X \pm Y] = V(X) + V(Y) \pm 2\text{Cov}(X, Y)$,

donde denominamos covarianza a $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$. Este valor indica el grado de variación conjunta de dos variables aleatorias respecto a sus medias, y es el test básico para determinar si existe dependencia entre variables. La varianza al ser cuadrática no tiene las mismas unidades que la media, y por lo

tanto también podemos usar su raíz, la desviación estándar $\sigma = \sqrt{\text{Var}(X)}$, como una medida de la dispersión en los datos.

Otra medida de tendencia central particularmente importante es la mediana, que corresponde al valor que se encuentra en el lugar central de todos los datos de una muestra cuando éstos están ordenados de menor a mayor. Por ejemplo, en el conjunto $\{1, 3, 3, 6, 7, 8, 9\}$ la mediana es 6. La ventaja básica de la mediana en la descripción de datos en comparación con la media es que es más robusta, ya que no está tan sesgada por valores extremadamente grandes o pequeños. Sin embargo, no se presta a cálculos algebraicos tan bien como la media aritmética. Finalmente, otra medida en la dispersión de los datos es el rango intercuartil. Es igual a la diferencia entre el percentil 75 y el 25, y por esto concentra el 50 por ciento de los datos que se agrupan alrededor de la mediana.

2.2 Método de Montecarlo

El método de Montecarlo es un protocolo no determinista o estadístico numérico, usado para aproximar expresiones matemáticas complejas y costosas de evaluar con exactitud, basado en la simulación de variables aleatorias [9]. Gracias a esto, permite la realización de experimentos con muestreos de números pseudoaleatorios en una computadora. El método es aplicable a cualquier tipo de problema, ya sea estocástico o determinista. En su versión más simple, expresamos la cantidad que queremos conocer como el valor esperado de una variable aleatoria Y , $\mu = E(Y)$. Luego generamos valores y_1, \dots, y_n independientes e idénticamente

distribuidos (i.i.d) de Y y tomamos su promedio

$$\tilde{\mu} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (2.12)$$

Usualmente Y es una función real $Y = f(X)$ que depende de otra variable aleatoria $X \in D \subseteq \mathbb{R}^d$, con densidad (o masa de probabilidad) $p(x)$. Luego $\mu = E[f(x)]$. Así, mientras $Y = f(X)$ sea una cantidad que puede ser promediada, podemos aplicar el método de Montecarlo.

La principal justificación del método es a través de la ley de los grandes números. Sea Y una variable aleatoria con $\mu = E(Y)$, y supongamos que y_1, \dots, y_n son datos i.i.d de Y . Luego, bajo la ley débil de los grandes números

$$\lim_{n \rightarrow \infty} P(|\hat{\mu}_n - \mu| < \varepsilon) = 1, \quad (2.13)$$

para todo $\varepsilon > 0$. Esto nos dice que la probabilidad de error mayor que ε al estimar μ tiende a cero. La ley fuerte de los grandes números afirma que

$$P(\lim_{n \rightarrow \infty} |\hat{\mu}_n - \mu| = 0) = 1. \quad (2.14)$$

Esto es, el error cuando $n \rightarrow \infty$ es cero, suponiendo que μ exista. Así, de ambas leyes se desprende que, usando Montecarlo, eventualmente podemos conseguir un error tan pequeño como queramos, siempre y cuando tengamos un n suficientemente grande. Sin embargo, no nos dice de que tamaño tiene que ser n para tener un determinado error.

La magnitud del error puede ser calculada. Supongamos que Y tiene una varianza finita, $Var(Y) = \sigma^2 < \infty$. Para una muestra independiente e idénticamente

distribuida, $\hat{\mu}_n$ es una variable aleatoria y tiene su propio valor esperado y varianza. El valor esperado de $\hat{\mu}_n$ es

$$E(\hat{\mu}_n) = \frac{1}{n} \sum_{i=1}^n E(Y_i) = \mu. \quad (2.15)$$

Como la esperanza de $\hat{\mu}_n$ es igual a μ decimos que el método de Montecarlo es insesgado. La varianza de $\hat{\mu}_n$ es

$$E((\hat{\mu}_n - \mu)^2) = \frac{\sigma^2}{n}. \quad (2.16)$$

de donde encontramos que la desviación estándar es $\sqrt{E((\hat{\mu}_n - \mu)^2)} = \sigma/\sqrt{n}$. Esto significa que el error en la estimación es del orden $O(n^{-1/2})$ cuando $n \rightarrow \infty$, ya que el error cuadrático medio de un estimador es su varianza más su sesgo al cuadrado, siendo este último actualmente cero.

2.3 Álgebra Lineal

El álgebra lineal estudia espacios vectoriales y operaciones lineales en estos espacios. La Mecánica Cuántica está formulada esencialmente sobre esta teoría matemática.

Un espacio vectorial sobre un cuerpo K es un conjunto V no vacío, dotado de dos operaciones para las cuales será cerrado: La suma $+: V \times V \rightarrow V$, que toma dos vectores \mathbf{u} y \mathbf{v} y les asigna un tercer $\mathbf{w} \in V$, y la multiplicación por escalar $\cdot: K \times V \rightarrow V$, que toma un escalar a y un vector \mathbf{u} , entregando otro vector $a\mathbf{u} \in V$.

Para que un conjunto V sea espacio vectorial debe satisfacer los siguientes axiomas [10]:

- Asociatividad de la suma: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
- Conmutatividad de la suma: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
- Elemento neutro en la suma: Existe un elemento $\mathbf{0} \in V$ llamado vector cero, tal que $\mathbf{u} + \mathbf{0} = \mathbf{u}$.
- Elemento opuesto en la suma: Por cada $\mathbf{u} \in V$ existe un elemento $-\mathbf{u} \in V$, llamado inverso aditivo, tal que $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.
- Distributividad del producto con respecto a la suma vectorial: $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.
- Distributividad del producto con respecto a la suma del cuerpo: $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$.
- Asociatividad del producto: $a(b\mathbf{u}) = (ab)\mathbf{u}$.
- Elemento neutro en la multiplicación: $1 * \mathbf{u} = \mathbf{u}$, donde 1 denota la identidad multiplicativa en K .

En este trabajo, el espacio vectorial de mayor interés es \mathbb{C}^n , el espacio de n -tuplas de números complejos (z_1, \dots, z_n) , y en lo sucesivo nos referiremos a este espacio. Es común representar los vectores como kets, $|\psi\rangle$ [11].

Una función de particular importancia es el producto interior $\langle \phi | \psi \rangle$, que toma dos elementos $|\psi\rangle$ y $|\phi\rangle$ del espacio vectorial y los lleva al cuerpo. Cualquier función $\langle \cdot | \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C}$ es un producto interior si satisface las siguientes propiedades:

- $\langle \phi | \psi \rangle = (a\langle \alpha | + b\langle \beta |) | \psi \rangle = a\langle \alpha | \psi \rangle + b\langle \beta | \psi \rangle$.

- $\langle \phi | \psi \rangle = \langle \psi | \phi \rangle^*$.
- $\langle \psi | \psi \rangle \geq 0$, con igualdad si y solo si $|\psi\rangle = 0$.

Si dos vectores tienen producto interior cero, esto es, $\langle \psi | \phi \rangle = 0$ se dice que son ortogonales. A partir del producto interior podemos definir la norma de un vector $|\psi\rangle$ como

$$\| |\psi\rangle \| = \sqrt{\langle \psi | \psi \rangle}, \quad (2.17)$$

siendo un vector unitario o normalizado al tener norma 1.

Sobre el espacio vectorial \mathbb{C}^n podemos definir un conjunto generador, esto es, elegir una familia de vectores $B = \{|\psi_1\rangle, \dots, |\psi_n\rangle\}$ en \mathbb{C}^n tal que podamos construir cualquier vector arbitrario en el espacio de la forma

$$|\psi\rangle = \sum_{i=1}^n a_i |\psi_i\rangle. \quad (2.18)$$

Si B es linealmente independiente, esto es, ninguno de los vectores en B puede ser escrito con una combinación lineal de los restantes, decimos que B forma una base del espacio. Además, si todos los vectores de la base son unitarios y ortogonales entre si, decimos que forman una base ortonormal.

Un operador lineal entre espacios vectoriales V y W es definido como una función $A : V \rightarrow W$ de la forma

$$A \left(\sum_i a_i |\psi_i\rangle \right) = \sum_i a_i A(|\psi_i\rangle) = \sum_i a_i A|\psi_i\rangle. \quad (2.19)$$

Cuando decimos que A es definido en V , significa que A es un operador lineal de V a V . Un operador importante es la identidad, \mathbb{I} , definida por $\mathbb{I}|\psi\rangle = |\psi\rangle$ para cualquier $|\psi\rangle \in V$. En general, un operador actuando sobre un ket por la izquierda, $A|\psi\rangle$, nos entrega otro ket.

Dos operadores X e Y son iguales, $X = Y$, si $X|\psi\rangle = Y|\psi\rangle$ para cualquier vector en el espacio. Los operadores además se pueden sumar de forma conmutativa y asociativa:

$$X + Y = Y + X. \quad (2.20)$$

$$X + (Y + Z) = (X + Y) + Z. \quad (2.21)$$

La multiplicación de operadores es asociativa, pero no es conmutativa, esto es, $XY \neq YX$. Los observables como la posición, el espín o el momentum, pueden ser representados por un operador lineal A en el espacio vectorial en cuestión, con la condición de que debe ser hermítico, es decir, A debe ser igual a su transpuesto conjugado, $A = A^\dagger$.

Existen vectores de particular importancia dado un operador A , conocidos como autovectores de A , con la propiedad

$$A|a_1\rangle = a_1|a_1\rangle, \quad A|a_2\rangle = a_2|a_2\rangle, \dots \quad (2.22)$$

donde a_1, a_2, \dots son escalares, llamados autovalores. Una consecuencia de la hermiticidad de los observables es que sus autovalores deben ser necesariamente reales. Además, todo operador hermítico A en un espacio vectorial V tiene una descomposición espectral, esto es, es diagonal con respecto a la base ortonormal

B formada por sus autovectores,

$$A = \sum_i a_i |a_i\rangle \langle a_i|, \quad (2.23)$$

donde a_i son autovalores asociados a sus respectivos autovectores. Los $P_i = |i\rangle \langle i|$ son llamados proyectores y satisfacen $P = P^2$. Cuando un operador tiene solo autovalores positivos o cero decimos que es semidefinido positivo.

Usualmente usamos la representación matricial de los operadores lineales, esto es, A es simplemente una matriz compleja de m por n componentes A_{ij} que envía vectores del espacio \mathbb{C}^n al \mathbb{C}^m bajo la multiplicación de la matriz A por el vector en \mathbb{C}^n ,

$$A \left(\sum_i a_i |\psi_i\rangle \right) = \sum_i a_i A |\psi_i\rangle. \quad (2.24)$$

Decimos que un espacio vectorial H es de Hilbert si es completo en la norma inducida por un producto interior. Que un espacio sea completo significa que todas las sucesiones convergen dentro del espacio. Así, cada vector $|\phi\rangle \in H$ puede ser aproximado por una serie $|\phi_n\rangle$,

$$\lim_{n \rightarrow \infty} \| |\phi\rangle - |\phi_n\rangle \| = 0. \quad (2.25)$$

Luego la serie $|\phi_n\rangle$ tiene un único valor límite $|\phi\rangle$.

2.4 Proceso de Ortogonalización de Gram-Schmidt

El proceso de ortonormalización de Gram–Schmidt es un algoritmo para construir, a partir de un conjunto de vectores linealmente independientes de un subespacio vectorial, otro conjunto ortonormal de vectores que genere el mismo subespacio vectorial [12]. Así, es usado generalmente para hallar bases en espacios euclidianos. Si tenemos un conjunto de vectores $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ linealmente independientes, y definimos $proy_{\mathbf{u}}(\mathbf{v}) = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}$, el algoritmo es el siguiente:

$$\begin{aligned}
 \mathbf{u}_1 &= \mathbf{v}_1 & \mathbf{e}_1 &= \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} \\
 \mathbf{u}_2 &= \mathbf{v}_2 - proy_{\mathbf{u}_1}(\mathbf{v}_2) & \mathbf{e}_2 &= \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} \\
 \mathbf{u}_3 &= \mathbf{v}_3 - proy_{\mathbf{u}_1}(\mathbf{v}_3) - proy_{\mathbf{u}_2}(\mathbf{v}_3) & \mathbf{e}_3 &= \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} \\
 \vdots & & \vdots & \\
 \mathbf{u}_k &= \mathbf{v}_k - \sum_{j=1}^{k-1} proy_{\mathbf{u}_j}(\mathbf{v}_k) & \mathbf{e}_k &= \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}.
 \end{aligned} \tag{2.26}$$

Gracias a este proceso los vectores \mathbf{e}_i forman una base ortonormal de un espacio vectorial. Esto es útil en las mediciones de la Mecánica Cuántica, ya que comúnmente realizamos mediciones proyectivas sobre observables. Así, cada uno de los detectores en un experimento, y por lo tanto cada resultado, debe estar asociado a un vector unitario en un espacio de Hilbert H , formando todos estos vectores una base ortonormal.

En el caso de este trabajo estamos interesados en medir sobre direcciones arbitrarias $|\phi\rangle$, para lo cual debemos construir una base ortonormal en H sobre la que medir. Entonces, elegimos al azar $d - 1$ vectores linealmente independientes que junto a $|\phi\rangle$ generen H , y aplicamos el método de Gram-Schmidt para con-

struir una base ortonormal. Es importante que el vector $|\phi\rangle$ permanezca intacto, y por lo tanto lo elegimos como primer vector en el método.

2.5 Medida de Haar

Las matrices unitarias de dimensión $N \times N$, $U(N)$, forman un grupo compacto de Lie, y tienen una única medida integral invariante por traslación izquierda y derecha $d\mu$, conocida como medida de Haar. En otras palabras, si denotamos la medida de Haar sobre $U(N)$ por $d\mu(U)$, tenemos

$$d\mu(VU) = d\mu(UW) = d\mu(U) \quad (2.27)$$

para cualquier $V, W \in U(N)$.

Así, en Mecánica Cuántica, esta medida sobre el grupo $U(N)$ induce un conjunto estándar de estados cuánticos puros aleatorios de tamaño N en un espacio de Hilbert N -dimensional. Esta medida nos entrega una distribución de probabilidades para $U(N)$ que pondera de igual forma diferentes regiones del espacio de las matrices unitarias, y por lo tanto luce como una distribución uniforme. En este caso los coeficientes de los estados cuánticos están dados por una distribución normal [13]. Es importante en este trabajo para que la muestra de los estados escogidos como semilla en el proceso de optimización no tenga zonas privilegiadas sobre el espacio de estados. Así, cuando promediamos en el método de Montecarlo lo hacemos de manera uniforme, no dándole más importancia a ninguna zona del espacio de Hilbert por sobre otra.

Capítulo 3

Mecánica Cuántica

Revisaremos conceptos básicos de Mecánica Cuántica, lo que nos entregará el modelo físico sobre el que se desarrolla la tomografía. También analizaremos este último tema junto a ciertas medidas de distancia en el espacio de estados cuánticos. Además, estudiaremos en detalle Maximum Likelihood Estimation, una técnica de postprocesamiento estadístico usada en tomografía. Finalmente, existen varias cotas fundamentales que nos entregan un límite de precisión en la tomografía de estados. Así, en este capítulo veremos la cota de Cramer-Rao cuántica y la cota de Gill-Massar.

3.1 Postulados de la Mecánica Cuántica

La Mecánica Cuántica es la rama de la física que explica la naturaleza a pequeñas escalas y niveles de energía. Está gobernada por 4 postulados, derivados en un largo proceso de ensayo y error [14]. El primer postulado nos indica como describir matemáticamente las propiedades cuánticas de un sistema físico:

- Todo sistema físico tiene asociado un espacio de Hilbert separable y complejo H con producto interior, conocido como espacio de estados. El sistema es completamente descrito por su estado cuántico, el cual es un vector unitario en el espacio de estados del sistema.

Dada una base ortonormal $\{|u_n\rangle\}_{n=1}^N$ podemos representar el estado en forma de ket:

$$|\psi\rangle = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} \langle c_1 | \Psi \rangle \\ \langle c_2 | \Psi \rangle \\ \vdots \end{bmatrix} \quad (3.1)$$

donde los coeficientes c_n son números complejos. También podemos representar el estado en forma de bra, $\langle\psi| = |\psi\rangle^\dagger$. Los espacios de kets y bras forman espacios vectoriales duales uno del otro. Como $|\psi\rangle$ debe ser normalizado, tenemos que $\langle\psi|\psi\rangle = 1$, o equivalentemente, $|c_1|^2 + |c_2|^2 + \dots = 1$. Si la dimensión del espacio es 2, llamamos a $|\psi\rangle$ un qubit.

El segundo postulado describe como cambian los estados cuánticos con el tiempo.

- La evolución de un sistema cuántico cerrado es descrita por una transformación unitaria U (donde $UU^\dagger = I$). Esto es, el estado $|\psi_1\rangle$ del sistema a tiempo t_1 está relacionado con el estado $|\psi_2\rangle$ del sistema a tiempo t_2 por un operador unitario U de tal forma que

$$|\psi_2\rangle = U(t_2, t_1)|\psi_1\rangle. \quad (3.2)$$

El tercer postulado ilustra nuestra interacción con el sistema al obtener información de este en forma de mediciones:

- Las mediciones cuánticas son descritas por una colección $\{M_m\}$ de operadores de medición que satisfacen $\sum_m M_m^\dagger M_m = I$, y que actúan en el espacio de estados del sistema a medir. El índice m etiqueta las distintas salidas del experimento. Dado un estado $|\psi\rangle$ del sistema cuántico antes de la medición, la probabilidad de obtener el resultado etiquetado por m es

$$p_m = \langle \psi | M_m^\dagger M_m | \psi \rangle, \quad (3.3)$$

y el estado inmediatamente después de que el sistema es medido es

$$|\psi\rangle_m = \frac{M |\psi\rangle}{\sqrt{\langle \psi | M_m^\dagger M_m | \psi \rangle}}. \quad (3.4)$$

Esto es un resultado completamente distinto a lo visto clásicamente, ya que al medir podemos cambiar completamente el estado del sistema.

Por último, el postulado 4 se refiere a los sistemas cuánticos compuestos de dos o más subsistemas:

- El espacio de estados de un sistema físico compuesto es el producto tensorial del espacio de estados de los subsistemas. Si tenemos subsistemas etiquetados del 1 al n , y el subsistema i es preparado en el estado $|\psi\rangle_i$, el estado total del sistema es $|\psi\rangle_1 \otimes |\psi\rangle_2 \otimes \dots \otimes |\psi\rangle_n$.

Existen subsistemas en los cuales el estado cuántico de dos o más objetos se debe describir mediante un estado único que involucra a todos los objetos del

sistema, aun cuando estos estén separados espacialmente. A esto le llamamos entrelazamiento cuántico, y lleva a correlaciones entre las propiedades físicas observables. Por ejemplo, consideremos dos electrones en el estado singlete:

$$|\psi\rangle = \left(\frac{1}{\sqrt{2}}\right)(|+-\rangle - |-+\rangle). \quad (3.5)$$

Supongamos que medimos la componente de espín de uno de los electrones. Si el resultado nos da espín +, el otro electrón necesariamente debe estar en espín -, y viceversa. Estas correlaciones pueden persistir incluso si las dos partículas están separadas distancias macroscópicas y sin interactuar, haciendo que las medidas realizadas sobre un sistema parezcan estar influyendo instantáneamente en otros sistemas que están enlazados con él.

3.2 Matriz Densidad

Una formulación alternativa de la Mecánica Cuántica está dada por la matriz densidad [14]. Este lenguaje nos entrega una forma conveniente de describir sistemas cuánticos cuyo estado no es completamente conocido. Dado un sistema cuántico con probabilidad p_i de encontrarse en uno de los estados del conjunto $\{|\psi_i\rangle\}$, llamamos matriz densidad u operador densidad a

$$\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|. \quad (3.6)$$

A partir de la definición podemos encontrar el resto de los postulados en esta representación. Dado el conjunto $\{p_i, |\psi_i(t_0)\rangle\}$, llamado ensemble, la evolución

del operador densidad desde un tiempo t_0 a t está dada por

$$\rho(t) = \sum_i p_i U(t, t_0) |\psi_i(t_0)\rangle \langle \psi_i(t_0)| U^\dagger(t, t_0) = U(t, t_0) \rho(t_0) U^\dagger(t, t_0). \quad (3.7)$$

Si medimos el conjunto $\{M_m\}$ de operadores de medición sobre el ensemble $\{p_i, |\psi_i\rangle\}$, la probabilidad de obtener el resultado m es

$$\begin{aligned} p(m) &= \sum_i p(m|i) p_i \\ &= \sum_i p_i \langle \psi_i | M_m^\dagger M_m | \psi_i \rangle \\ &= \sum_i p_i \text{tr}(M_m^\dagger M_m | \psi_i \rangle \langle \psi_i |) \\ &= \text{tr}(M_m^\dagger M_m \rho), \end{aligned} \quad (3.8)$$

y el estado tras la medición es

$$\rho_m = \frac{M_m \rho M_m^\dagger}{\text{tr}(M_m^\dagger M_m \rho)}. \quad (3.9)$$

Matemáticamente, las matrices densidades están caracterizadas por el siguiente teorema:

Theorem 1. *Un operador ρ es una matriz densidad asociada a algún ensemble $\{p_i, |\psi_i\rangle\}$ si y solo si satisface las condiciones:*

- ρ tiene traza igual a uno.
- ρ es un operador semidefinido positivo.

La primera condición viene del hecho que $\rho = \sum_i p_i |\psi\rangle\langle\psi|$. Entonces,

$$\text{tr}(\rho) = \sum_i p_i \text{tr}(|\psi\rangle\langle\psi|) = \sum_i p_i = 1. \quad (3.10)$$

Para demostrar la segunda, suponemos un vector $|\phi\rangle$ en el espacio de estados. Luego,

$$\langle\phi|\rho|\phi\rangle = \sum_i p_i \langle\phi|\psi\rangle\langle\psi|\phi\rangle = \sum_i p_i |\langle\phi|\psi\rangle|^2 \geq 0. \quad (3.11)$$

Un sistema cuántico cuyo estado es $|\psi\rangle$, esto es, su matriz densidad está dada por $\rho = |\psi\rangle\langle\psi|$, es llamado estado puro. De otra forma, el estado es llamado mezclado, ya que es una mezcla de distintos estados puros en el ensemble. Las matrices densidad cumplen $\rho^2 \leq \rho$, donde la igualdad se da si y solo si el estado es puro.

3.3 Descomposición de Schmidt

El entrelazamiento es una propiedad que se da en sistemas compuestos de dos o más partículas. Para sistemas bipartitos, el estado puede ser representado mediante la descomposición de Schmidt [14]. Esto permite a su vez una fácil caracterización de los estados entrelazados.

Sea $|\psi\rangle$ un estado puro de un sistema compuesto AB . Entonces, existen estados ortonormales $|i_A\rangle$ y $|i_B\rangle$ para los sistemas A y B tal que

$$|\psi\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle, \quad (3.12)$$

donde λ_i son números reales no negativos que satisfacen $\sum_i \lambda_i^2 = 1$, llamados coeficientes de Schmidt.

Gracias a la descomposición de Schmidt podemos encontrar matrices densidad reducidas ρ_A y ρ_B para un estado puro $|\psi\rangle$ en el sistema AB , trazando parcialmente en los sistemas. Para el sistema A tenemos $\rho_A = \sum_i \lambda_i^2 |i_A\rangle\langle i_A|$ y para B , $\rho_B = \sum_i \lambda_i^2 |i_B\rangle\langle i_B|$.

La importancia de los coeficientes de Schmidt λ_i es que en cierto sentido cuantifican la 'cantidad' de entrelazamiento entre los sistemas A y B . Así, si solo existe un coeficiente de Schmidt distinto de cero, $\lambda = 1$, no existe entrelazamiento, y podemos descomponer el estado del sistema en $|\psi\rangle = |\phi_A\rangle|\phi_B\rangle$.

3.4 Medidas de Distancia

Las medidas de distancia en Mecánica Cuántica nos dan una respuesta cuantitativa a la pregunta de que tan cerca están dos estados cuánticos. En esto radica su importancia en tomografía, pues nos ayudan a cuantificar el error en la estimación.

La medida más comúnmente usada es la fidelidad [14]. Para dos estados ρ y σ está dada por

$$F(\rho, \sigma) = \left(\text{tr} \sqrt{\rho^{1/2} \sigma \rho^{1/2}} \right)^2. \quad (3.13)$$

Algunas propiedades de esta medida son:

- Es preservada por evoluciones unitarias U , esto es, $F(\rho, \sigma) = F(U\rho U^\dagger, U\sigma U^\dagger)$.
- Es simétrica en sus argumentos, $F(\rho, \sigma) = F(\sigma, \rho)$
- $F(\rho, \sigma) \in [0, 1]$.
- $F(\rho, \sigma) = 1$ si y solo si $\rho = \sigma$.

Utilizando la fidelidad podemos definir una cantidad llamada distancia de Bures, que para ρ y σ es

$$D(\rho, \sigma)^2 = 2 \left(1 - \sqrt{F(\rho, \sigma)} \right). \quad (3.14)$$

Esta definición satisface las propiedades que debe tener una métrica, es decir, no negativa, simétrica, satisface desigualdad triangular, y es cero entre dos vectores si y solo si estos son el mismo.

La versión infinitesimal de esta distancia viene dada por la métrica de Bures [15]

$$d(\rho, \rho + d\rho) = \frac{1}{2} \text{tr}(d\rho L), \quad (3.15)$$

donde L es un operador hermítico definido implícitamente por $\rho L + L\rho = d\rho$.

Si uno de los estados es puro, la fidelidad se simplifica a

$$F(|\psi\rangle, \sigma) = \langle \psi | \sigma | \psi \rangle. \quad (3.16)$$

En este último caso la función puede ser estimada experimentalmente midiendo un sistema que esté en el estado σ , usando una base que contenga a $|\psi\rangle$, de tal

forma que

$$F(|\psi\rangle, \sigma) \approx n_\psi/N, \quad (3.17)$$

donde n_ψ es el número de detecciones en el detector asociado al estado $|\psi\rangle$, y N es el número total de detecciones.

Otra medida de distinguibilidad simple es la trace distance, definida para dos operadores densidad ρ y σ como

$$D(\rho, \sigma) = \frac{1}{2} \text{tr}|\rho - \sigma|, \quad (3.18)$$

donde $|A| = \sqrt{A^\dagger A}$. Está relacionada con la fidelidad por $1 - \sqrt{F(\rho, \sigma)} \leq D(\rho, \sigma) \leq \sqrt{1 - F(\rho, \sigma)}$. La trace distance cuantifica la probabilidad de error al distinguir dos estados cuánticos en principio igualmente probables con una sola medición. Como en la tomografía cuántica usualmente trabajamos con un amplio número de mediciones, esta cantidad es de poca importancia en el área.

Otras dos cantidades son la distancia de Hilbert Schmidt, $d_{HS}(\rho, \sigma) = \text{Tr}[(\rho - \sigma)^2]$, y la entropía relativa o divergencia KL, $S(\rho||\sigma) = \text{Tr}[\rho(\log\rho - \log\sigma)]$.

Por último, la Medida Geométrica de Entrelazamiento cuantifica para un estado puro en un sistema compuesto, su grado de entrelazamiento [7]. Si $|\phi\rangle$ es un estado separable y $|\psi\rangle$ el estado del sistema, está dada por

$$\lambda_{max} = \max_\phi |\langle \phi || \psi \rangle|. \quad (3.19)$$

Esto significa que buscamos el estado separable más próximo al estado del sistema.

3.5 Tomografía Cuántica

La tomografía cuántica relaciona el modelo matemático de estado cuántico con cantidades experimentalmente medibles [6]. Para el caso de los estados, es el proceso de determinar experimentalmente el estado cuántico de un sistema físico a través de mediciones. Como vimos en el postulado 3, las mediciones sobre el sistema nos dan resultados escogidos aleatoriamente, y además el estado es transformado después de la medición. Así, a diferencia de la caracterización de un objeto clásico, que involucra una serie de mediciones sobre el mismo objeto, la forma de reconstruir el estado es a partir de medir probabilidades en un ensemble de partículas idénticamente preparadas sobre un conjunto convenientemente elegido de bases de medición. Con los datos obtenidos a partir de mediciones se estiman funciones, tales como valores de expectación o fidelidades, que con cierto procesamiento matemático entregan un estado estimado del sistema. Como las funciones son solo aproximadas a través de estas mediciones, en la práctica nunca la tomografía cuántica entrega resultados exactos, ya que se requeriría medir infinitas partículas para eliminar el error estadístico asociado al experimento.

3.5.1 Tomografía Estándar

El método más utilizado en Mecánica Cuántica es llamado Tomografía Estándar. La matriz densidad de un sistema de dimensión d , llamada qudit, al ser hermítica tiene un total de d^2 coeficientes reales linealmente independientes, ligados por la condición de $\text{tr}(\rho) = 1$. Por lo tanto, necesitamos medir un conjunto de

$d^2 - 1$ operadores más la identidad para poder reconstruir el estado del sistema, siempre y cuando estos operadores elegidos generen el espacio de las matrices hermíticas. Con esto nos aseguramos que la solución al problema sea única. Así, podemos elegir estas matrices como las generadoras del grupo unitario especial de dimensión d , $SU(d)$. Estas matrices cumplen el álgebra de conmutadores $[\sigma_i, \sigma_j] = 2i\varepsilon_{ijk}\sigma_k$. Definiendo $\sigma_0 = \mathbb{I}$, y $\{\sigma_j\}_{j=1}^{d^2-1}$ como el conjunto $su(d)$, el estado del sistema puede escribirse como

$$\rho = \frac{1}{d} \sum_{j=0}^{d^2-1} r_j \sigma_j, \quad (3.20)$$

donde $r_j \in \mathbb{R}$. De las propiedades de las matrices de $su(d)$ podemos demostrar que $r_j = \frac{d}{2} \text{tr}(\rho \sigma_j)$, o sea, es el valor de expectación de la respectiva matriz de Pauli. Esta cantidad puede ser reconstruida experimentalmente con $d^2 - 1$ mediciones.

El caso $d = 2$ es de particular importancia. Los generadores de $SU(2)$ son

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad (3.21)$$

$$\sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad (3.22)$$

$$\sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (3.23)$$

Luego $r_k = \text{tr}(\rho \sigma_k)$ (donde $k = 1, 2, 3$) es el valor de expectación de la correspondiente matriz de Pauli sobre el sistema, cantidad que estimamos experimentalmente. Por lo tanto, ρ es

$$\rho = \frac{1}{2} \begin{bmatrix} 1 + \text{tr}(\rho \sigma_3) & \text{tr}(\rho \sigma_1) - i \cdot \text{tr}(\rho \sigma_2) \\ \text{tr}(\rho \sigma_1) + i \cdot \text{tr}(\rho \sigma_2) & 1 - \text{tr}(\rho \sigma_3) \end{bmatrix}. \quad (3.24)$$

3.5.2 Tomografía SIC-POVM

Un POVM sobre un espacio de Hilbert H de dimensión d es un conjunto de m operadores semidefinidos positivos $\{F_i\}$ en H que suman la identidad,

$$\sum_{i=1}^m F_i = I. \quad (3.25)$$

Un POVM informacionalmente completo (IC-POVM) es un POVM que contiene al menos d^2 operadores que generan H .

Si tenemos un conjunto de d^2 proyectores de rango 1, $\{\Pi_i\}_{i=1}^{d^2}$, que satisfacen la condición

$$tr(\Pi_i \Pi_j) = \frac{d\delta_{ij} + 1}{d + 1}, \quad (3.26)$$

esto define un conjunto $\{F_i\}_{i=1}^{d^2} = \{\frac{1}{d}\Pi_i\}_{i=1}^{d^2}$ llamado SIC-POVM.

Como este conjunto contiene d^2 elementos podemos reconstruir la matriz densidad ρ de la forma

$$\rho = \sum_{i=1}^{d^2} c_i \Pi_i, \quad (3.27)$$

de donde podemos demostrar que $c_i = (d + 1)tr(\rho F_i) - 1/d$. Así, podemos reconstruir el estado a partir de mediciones experimentales de $tr(\rho F_i)$,

$$\rho = \sum_{i=1}^{d^2} ((d + 1)tr(\rho F_i) - 1/d) \Pi_i. \quad (3.28)$$

3.5.3 Tomografía MUBs

Llamamos MUBs a dos o mas bases ortonormales $\{|e_1\rangle, \dots, |e_d\rangle\}$ y $\{|f_1\rangle, \dots, |f_d\rangle\}$ en un espacio de Hilbert H de dimensión d tal que satisfacen

$$|\langle e_i | f_j \rangle|^2 = \frac{1}{d}, \quad (3.29)$$

para todo $j, k \in \{1, \dots, d\}$.

Cualquier estado cuántico ρ actuando en H puede ser escrito como una combinación lineal de proyectores de rango 1 asociados a un conjunto completo de $d + 1$ MUBs. Así, podemos realizar tomografía usando MUBs [16], y reconstruimos la matriz densidad de la forma

$$\rho = \sum_{i=0}^d \sum_{j=0}^{d-1} \left(p_j^i - \frac{1}{d+1} \right) \Pi_j^i, \quad (3.30)$$

donde $p_j^i = \text{tr}(\rho \Pi_j^i)$ puede ser determinado experimentalmente. El índice j etiqueta cada base, y el i etiqueta los elementos de la base dada.

3.6 Estimación por Máxima Verosimilitud

Maximum Likelihood Estimation (MLE), o estimación por máxima verosimilitud es un método en inferencia estadística usado para estimar algún parámetro desconocido de una población a partir de datos observados [17]. En Mecánica Cuántica se utiliza como técnica de postprocesamiento que se aplica a los datos obtenidos por algún protocolo tomográfico.

La medición de los estados $\{|\phi_1\rangle, \dots, |\phi_n\rangle\}$ sobre el estado desconocido ρ del sistema genera un conjunto de n resultados $\{y_1, \dots, y_n\}$, dados por una distribución de probabilidad empírica $p_{datos}(y|\phi)$. Sea $p_{modelo}(y|\phi; \rho)$ una familia paramétrica de distribuciones de probabilidad sobre el mismo espacio que p_{datos} . El estimador de máxima verosimilitud es entonces definido como

$$\rho_{ML} = \arg \max_{\rho} \prod_{i=1}^n p_{modelo}(y_i|\phi_i; \rho). \quad (3.31)$$

Así, usamos la información disponible en la muestra para elegir el valor del parámetro ρ para el cual es más probable haber observado ese resultado.

El producto sobre muchas probabilidades tiene muchos inconvenientes a la hora de aplicarlo numéricamente. Para solucionar esta situación aplicamos el logaritmo natural, obteniendo así un problema de optimización equivalente pero más sencillo, al transformar el producto en una sumatoria:

$$\rho_{ML} = \arg \max_{\rho} \sum_{i=1}^n \ln p_{modelo}(y_i|\phi_i; \rho). \quad (3.32)$$

Como el resultado de la optimización no cambia cuando reescalamos la función, podemos dividir por n para obtener

$$\rho_{ML} = \arg \max_{\rho} E_{y \sim p_{datos}} \ln p_{modelo}(y|\phi; \rho). \quad (3.33)$$

Una forma de interpretar MLE es verlo como una minimización en la disimilitud entre la distribución empírica p_{datos} definida por las mediciones, y el modelo que le queremos asociar a esta [18]. Este grado de disimilitud es cuantificado por

la divergencia de Kullback-Leibler (KL), dada por

$$D_{KL}(p_{datos}, p_{modelo}) = E_{y \sim p_{datos}} [\ln p_{datos}(y_i) - \ln p_{modelo}(y_i | \phi_i; \rho)]. \quad (3.34)$$

La divergencia KL es una medida no simétrica de la similitud o diferencia entre dos funciones de distribución de probabilidad. Como el primer término es función solo de la data, cuando minimizamos solo necesitamos optimizar

$$-E_{y \sim p_{datos}} \ln p_{modelo}(y_i | \phi_i; \rho). \quad (3.35)$$

lo que coincide con maximizar la ecuación 3.33.

En el caso de la tomografía la medición de los estados $\phi = \{|\phi_1\rangle, \dots, |\phi_n\rangle\}$ sobre el sistema nos entrega un conjunto de infidelidades para cada estado, $\mathbf{y} = \{y_1, \dots, y_n\}$. Considerando que la distribución de probabilidad de los datos p_{datos} es normal, definimos nuestro p_{modelo} como

$$p_{modelo}(\mathbf{y} | \phi; \rho) = \mathcal{N}(\mathbf{y}; \hat{\mathbf{y}}(\phi; \rho), \sigma^2). \quad (3.36)$$

La función $\hat{\mathbf{y}}(\phi; \rho) = \langle \phi_i | \rho | \phi_i \rangle$ es la predicción de la infidelidad dada por nuestro modelo, calculada con la estimación del estado ρ del sistema y los estados sobre los que medimos, y nos entrega un estimador del promedio de la Gaussiana. σ^2 es la varianza, fija en alguna constante. Luego, al calcular el logaritmo del estimador de verosimilitud tenemos

$$\sum_{i=1}^n \ln p_{modelo}(y_i | \phi_i; \rho) = -n \ln \sigma - \frac{n}{2} \ln 2\pi - \sum_{i=1}^n \frac{||\langle \phi_i | \rho | \phi_i \rangle - y_i||^2}{2\sigma^2}. \quad (3.37)$$

Los dos primeros términos son constantes, por lo que al optimizar el único término relevante es el tercero. Este, por su parte, tiene bastante similitud con el error cuadrático medio,

$$ECM = \frac{1}{n} \sum_{i=1}^n ||\langle \phi_i | \rho | \phi_i \rangle - y_i||^2, \quad (3.38)$$

de donde vemos inmediatamente que maximizar la verosimilitud y minimizar el error cuadrático medio nos entrega un mismo punto óptimo $\tilde{\rho}$.

Bajo condiciones apropiadas, MLE tiene la propiedad de consistencia, esto es, a medida que el número de mediciones n aumenta, el parámetro estimado por máxima verosimilitud converge a su valor verdadero. Estas condiciones son:

- La distribución verdadera p_{data} debe vivir en el mismo espacio que $p_{modelo}(\cdot; \rho)$.
- La distribución p_{data} debe tener solo un valor de ρ . De otra forma, MLE puede recobrar la p_{data} correcta, pero no sera capaz de determinar cual valor de ρ fue el que genero la data.

3.7 Cota de Gill-Massar

Una forma de medir la cantidad de información que una variable aleatoria X entrega sobre un parámetro desconocido θ de la distribución que modela X es la información de Fisher [19],

$$I_{jk} = E \left[\left(\frac{\partial \ln p(\xi|\theta)}{\partial \theta_j} \right) \left(\frac{\partial \ln p(\xi|\theta)}{\partial \theta_k} \right) \right]. \quad (3.39)$$

Esto representa la sensibilidad promedio del logaritmo de la función de verosimilitud $p(\xi|\theta)$ con respecto a la variación del parámetro θ . Mientras mayor sea la

información de Fisher, mejor podremos estimar el valor de θ . La información de Fisher es aditiva. Así, la matriz I de N mediciones idénticas e independientes es N veces una de estas mediciones.

En el caso de los estados cuánticos, es factible caracterizar $\rho(\theta)$ por un conjunto de parámetros $\theta_1, \dots, \theta_n$. Para determinar el valor de estos parámetros podemos realizar mediciones y construir un estimador $\hat{\theta}$ basado en estas. De esta forma, elegimos un esquema tomográfico, y para estimar $\hat{\theta}$ medimos probabilidades siguiendo la regla de Born, $p(\xi|\theta) = \text{tr}(\rho\Pi_\xi)$, con Π_ξ operadores de medición asociados al resultado ξ . Con estas probabilidades podemos calcular también la información de Fisher asociada al protocolo tomográfico.

Basándonos en la información de Fisher, podemos encontrar una cota llamada de Cramer-Rao,

$$C(\theta) \geq I^{-1}(\theta). \quad (3.40)$$

Esto establece que la matriz de covarianza $C_{jk} = E[(\hat{\theta}_j - \theta_k)(\hat{\theta}_k - \theta_j)]$ de un estimador insesgado es acotada inferiormente por el inverso de la matriz de información de Fisher. Así, cuantifica el error mínimo que podemos tener al inferir los parámetros de nuestra distribución. De acuerdo a la aditividad, la matriz de covarianza de un estimador basado en N mediciones satisface $C^N(\theta) \geq NI^{-1}(\theta)$.

Una alternativa general a C es la matriz de covarianza ponderada $W(\theta)C(\theta)$, donde $W(\theta)$ es una matriz semidefinida positiva que puede depender de θ . Como en la práctica es más conveniente usar un solo número en vez de una matriz para

cuantificar el error, usamos $tr(W(\theta)C(\theta))$. Este término tiene la ventaja de que no depende de la parametrización que estemos usando, y para cierto caso especial de W converge a la medida de Bures. El límite de Cramer-Rao ahora implica que $tr(W(\theta)C(\theta)) \geq tr(W(\theta)I^{-1}(\theta))$.

Podemos definir una información de Fisher Cuántica [15] de la forma

$$J(\theta) = tr(\rho(\theta)L(\theta)^2), \quad (3.41)$$

donde $L(\theta)$ es un operador hermítico que satisface

$$\rho'(\theta) = \frac{1}{2}[\rho(\theta)L(\theta) + L(\theta)\rho(\theta)] = d\rho(\theta)/d\theta. \quad (3.42)$$

J se relaciona con la distancia infinitesimal de Bures,

$$D_B^2 = (\rho(\theta), \rho(\theta + d\theta)) = \frac{1}{4}J(\theta)d\theta^2. \quad (3.43)$$

La información de Fisher cuántica es una cota superior para la clásica, esto es,

$$I(\theta) \leq J(\theta). \quad (3.44)$$

Dada una matriz ponderada W , esta desigualdad configura una cota inferior para $tr(WC)$ para cualquier estimador insesgado,

$$tr(WC) \geq tr(WI^{-1}) \geq tr(WJ^{-1}). \quad (3.45)$$

Esta desigualdad es independiente de la parametrización, lo que nos permite elegir W del tal forma que $tr(WC)$ coincida con algunas figuras de mérito. Por

ejemplo, si $W = J/4$ [20], tenemos una cota para la medida de Bures dada por

$$\varepsilon_{CR}(\rho) = \frac{1}{4} \text{tr}(JJ^{-1}) = \frac{1}{4} \text{tr}(\mathbb{I}) = \frac{d^2 - 1}{4}, \quad (3.46)$$

válida para matrices densidad de $d^2 - 1$ parámetros. Así, si el tamaño de la muestra es N , por la aditividad de la información de Fisher esta cota se convierte en $\frac{d^2 - 1}{4N}$, la cota de Cramer-Rao cuántica.

Otro límite en la estimación es lo derivado por Gill y Massar, lo cual es aplicable a mediciones separables.

La traza de Gill-Massar es definida por la traza del producto entre la matriz de Fisher y la matriz cuántica de Fisher, esto es, $\text{tr}(J^{-1}(\theta)I(\theta))$. Gill y Massar en [21] derivaron una relación entre este término y la dimensión del espacio de estados, denominada cota de Gill-Massar. Para un estado cuántico $\rho(\theta)$ en un espacio de Hilbert H de dimensión d ,

$$\text{tr}(J^{-1}I^{(N)}) \leq N(d - 1), \quad (3.47)$$

valido para mediciones separables. $I^{(N)}$ es la información de Fisher en $H^{\otimes N}$. Esta desigualdad impone un límite de precisión y eficiencia fundamental en la tomografía de estados cuánticos basada en mediciones individuales.

El teorema anterior y la cota de Cramer-Rao clásica implican que

$$\text{tr}(J^{-1}(\theta)C^{(N)-1}(\theta)) \leq N(d - 1). \quad (3.48)$$

Esta desigualdad entrega una cota inferior para $\text{tr}(WC)$ [20],

$$N \text{tr}(WC^N) \geq \varepsilon_{GM} = \frac{\left(\text{tr} \sqrt{J^{-1/2} W J^{-1/2}} \right)^2}{d - 1}. \quad (3.49)$$

Si $W = J/4$ encontramos una cota para la distancia de Bures,

$$\begin{aligned}
 \varepsilon_{GM} &= \frac{1}{4(d-1)} (\text{tr} \sqrt{JJ^{-1}})^2 \\
 &= \frac{1}{4(d-1)} (\text{tr}(\mathbb{I}))^2 \\
 &= \frac{1}{4N} (d+1)^2 (d-1), \tag{3.50}
 \end{aligned}$$

la cual es independiente del estado del sistema. Este resultado es valido para mediciones separables en $H^{\otimes N}$ sobre matrices densidad.

Para métodos que estiman estados puros la cota de Gill-Massar para la distancia de Bures es diferente. Esto es porque ahora el número de parámetros a estimar es $2d - 2$, a diferencia de las matrices densidad, donde es $d^2 - 1$. Luego $\text{tr}(\mathbb{I}) = 2d - 2$, y entonces,

$$\begin{aligned}
 \varepsilon(\rho) &= \frac{1}{4N(d-1)} (\text{tr}(\mathbb{I}))^2 \\
 &= \frac{1}{4N(d-1)} 4(d-1)^2 \\
 &= \frac{d-1}{N}. \tag{3.51}
 \end{aligned}$$

Esto nos dice que podemos obtener un menor error en la estimación si sabemos que el estado desconocido del sistema es puro, al disminuir el número de parámetros a estimar, y por lo tanto, el número de observables a medir.

Capítulo 4

Gradiente Descendente

Los resultados de este trabajo provienen de optimizar ciertas funciones de mérito en Mecánica Cuántica, y por lo tanto es importante el método que elegimos para realizar esta tarea. En este capítulo presentamos un esquema básico de algoritmo de gradiente descendente, para luego ir a los métodos específicos que usaremos en este trabajo, Complex Simultaneous Perturbation Stochastic Approximation (CSPSA) y Simultaneous Perturbation Stochastic Approximation (SPSA).

4.1 Algoritmo Básico de Gradiente Descendente

El algoritmo de gradiente descendente busca los puntos $\tilde{\mathbf{x}} \in D$ donde las funciones del tipo $f : D \subseteq \mathbb{R}^m \rightarrow \mathbb{R}$ alcanzan un mínimo. La idea proviene de la aproximación a primer orden de $f(\mathbf{x})$ en serie de Taylor alrededor de un punto \mathbf{x}^* :

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + O(2). \quad (4.1)$$

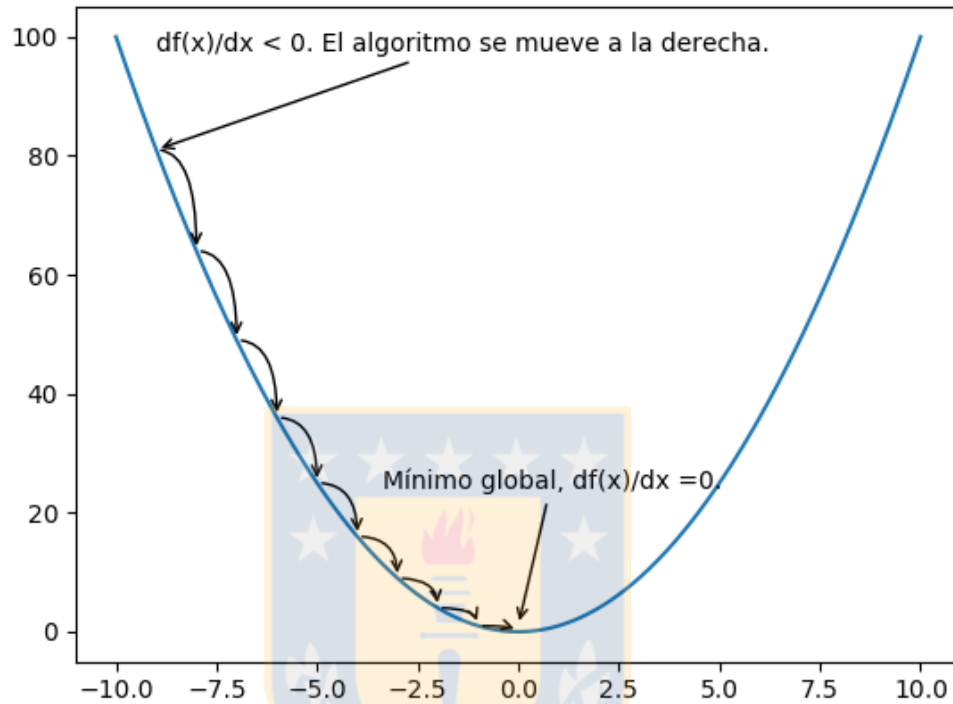


Figura 4.1: Gráfico del método de gradiente descendente para la función $f(x) = x^2$. Un punto inicial cualquiera va actualizándose iterativamente hasta converger al mínimo cuando el gradiente es cero. Fuente: Elaboración propia.

Consideremos un movimiento desde \mathbf{x}^* una pequeña cantidad h en una dirección unitaria \mathbf{u} . Esperamos encontrar el \mathbf{u} que minimice $f(\mathbf{x}^* + h\mathbf{u})$. De la expansión en serie de Taylor de $f(\mathbf{x}^* + h\mathbf{u})$ alrededor de \mathbf{x}^* encontramos que

$$f(\mathbf{x}^* + h\mathbf{u}) - f(\mathbf{x}^*) = h\nabla f(\mathbf{x}^*)\mathbf{u} + O(2). \quad (4.2)$$

De acá vemos para que $f(\mathbf{x}^* + h\mathbf{u}) - f(\mathbf{x}^*) < 0$ decrezca más rápido debemos minimizar $h\nabla f(\mathbf{x}^*)\mathbf{u}$. Así, el vector unitario que minimiza $h\nabla f(\mathbf{x}^*)\mathbf{u}$ es $\mathbf{u} = -\nabla f(\mathbf{x}^*) / \|\nabla f(\mathbf{x}^*)\|$. Luego, el algoritmo aprovecha esto para converger de forma iterativa.

Para llegar al punto $\tilde{\mathbf{x}}$ donde encontramos el mínimo de una función $f : D \subseteq \mathbb{R}^m \rightarrow \mathbb{R}$, construimos una sucesión de puntos $\{\mathbf{x}_n\}$ tal que \mathbf{x}_n converge a $\tilde{\mathbf{x}}$. Para esto, suponiendo que en la iteración n estamos en un punto cualquiera \mathbf{x}_n , nos desplazamos una cantidad $-h_n \nabla f(\mathbf{x}_n)$ para encontrar un punto \mathbf{x}_{n+1} más cercano a $\tilde{\mathbf{x}}$, como vemos en la figura 4.1. Luego,

$$\mathbf{x}_{n+1} = \mathbf{x}_n - h_n \nabla f(\mathbf{x}_n). \quad (4.3)$$

h_n se usa para controlar la convergencia del algoritmo, o cuanto avanza en cada iteración, y su valor depende del método a utilizar.

4.2 CSPA-SPSA

Un problema recurrente del algoritmo de gradiente descendente es que puede resultar computacionalmente demandante cuando trabajamos en altas dimensiones. Esto es porque debemos calcular las derivadas parciales de la función objetivo sobre todos los ejes del espacio, para así obtener el gradiente utilizado en cada paso de la optimización. Luego el número de términos a evaluar crece con la dimensión. Además, requiere conocer la forma analítica de la función a optimizar, lo cual dependiendo del problema a tratar podría no estar disponible.

Complex Simultaneous Perturbation Stochastic Approximation (CSPA) es un algoritmo usado en la optimización de sistemas con múltiples parámetros desconocidos [6] que soluciona el problema anterior. Pertenece a la familia de métodos de optimización estocásticos de gradiente descendente, y tiene la caracterís-

tica de estar formulado directamente para funciones $f : \mathbb{C}^n \rightarrow \mathbb{R}$, evitándonos parametrizar problemas de variable compleja sobre partes real e imaginaria. A diferencia de otros protocolos (como gradiente descendente) que requieren mediciones directas del gradiente de la función a optimizar, CSPSA solo requiere una estimación del gradiente, usando dos mediciones de la función objetivo por iteración, independiente de la dimensión en la que estemos trabajando.

Para optimizar una función $f(\hat{\mathbf{z}}) : \mathbb{C}^n \rightarrow \mathbb{R}$, debemos encontrar el punto $\tilde{\mathbf{z}}$ donde la función f alcanza un extremo. Así, los pasos de cada iteración del método son los siguientes:

- Dado $\hat{\mathbf{z}}_k$, generamos una dirección aleatoria a buscar, definiendo un vector complejo Δ_k de la misma dimensión que el espacio.
- Calculamos las perturbaciones simultaneas de \mathbf{z}_k , $\mathbf{z}_{k+} = \mathbf{z}_k + c_k \Delta_k$ y $\mathbf{z}_{k-} = \mathbf{z}_k - c_k \Delta_k$ y medimos la función objetivo con ruido en estas perturbaciones, $h_{k+} = h(\mathbf{z}_{k+}, \xi)$ y $h_{k-} = h(\mathbf{z}_{k-}, \xi)$.
- Estimamos el gradiente en la dirección Δ_k ,

$$\mathbf{g}_k = \frac{h_{k+} - h_{k-}}{2c_k \Delta_k^*}. \quad (4.4)$$

- Calculamos la próxima iteración vía

$$\mathbf{z}_{k+1} = \mathbf{z}_k - a_k \mathbf{g}_k. \quad (4.5)$$

Las funciones a y c controlan la convergencia, y son de la forma

$$a_k = \frac{a}{(k+1+A)^s} \quad c_k = \frac{b}{(k+1)^t}. \quad (4.6)$$

Las componentes del vector $\Delta_k \in \mathbb{C}$ son generadas aleatoria e independientemente. Como vemos en el método, CSPSA no requiere la evaluación del gradiente complejo. En vez de esto, usa un estimado insesgado del gradiente para generar una secuencia de estimados \mathbf{z}_k que convergen al punto donde la función alcanza el mínimo. Esto es una ventaja cuando la evaluación del gradiente complejo no es posible.

La convergencia del algoritmo está garantizada por los siguientes teoremas:

Theorem 2. Consideremos $k \geq K$ para $K < \infty$. $\mathbb{E}[\Delta_{k,i}|\omega] = 0$ para todo k y $\omega \in \Omega$, el espacio muestral generador de las secuencias $\hat{z}_1, \dots, \hat{z}_n$ de estimaciones. Las componentes $\Delta_{k,i}$ son independientes e idénticamente distribuidas casi seguramente. $\mathbb{E}[\Delta_{k,i}] = 0 = \mathbb{E}[e^{2i\phi_{k,i}}]$, donde $\phi_{k,i}$ es la fase de $\Delta_{k,i}$ en su descomposición polar, y $\mathbb{E}[\Delta_{k,i}^{-1}] \leq \alpha_1$. Para casi todo \hat{z}_k y $\forall k \geq K$ existe una vecindad abierta, que no es función de k o ω , donde $f \in \mathbb{C}^n \times \mathbb{C}^n$ y $|\partial_{z_i}\partial_{z_j}f| \leq \alpha_2$. Entonces, para casi todo $\omega \in \Omega$ y

$$b_k(\hat{z}_k) = \mathbb{E}[\hat{g}_k(\hat{z}_k) - \hat{g}(\hat{z}_k)|\hat{z}_k] \quad (4.7)$$

tenemos que

$$b_k(\hat{z}_k) = O(c_k). \quad (4.8)$$

Asumiendo que $O(c_k) \rightarrow 0$ tenemos que $b_k(\hat{z}_k) \rightarrow 0$ y consecuentemente $\hat{g}_k(\hat{z}_k) \rightarrow g_k(\hat{z}_k)$, esto es, \hat{g}_k es asintóticamente un estimado insesgado del gradiente g de la función f a ser minimizada.

En síntesis, este teorema nos entrega las condiciones que debe cumplir el gradiente para que sea insesgado, hipótesis que necesitaremos en el futuro.

El siguiente teorema establece la convergencia del algoritmo.

Theorem 3. *Asumiendo las condiciones del teorema 1 y las siguientes condiciones:*

- $a_k, c_k > 0 \quad \forall k, \lim_{k \rightarrow \infty} |a_k| + |c_k| = 0, \sum_k a_k = \infty, \text{ y } \sum_k \left(\frac{a_k}{c_k}\right)^2 < \infty.$
- *Existen β_0, β_1 y β_2 tal que $\mathbb{E}(\epsilon_{k,\pm})^2 \leq \beta_0, \mathbb{E}[(f(\hat{z} \pm c_k \Delta_k))^2] \leq \beta_1$ y $\mathbb{E}[\Delta_{k,l}^{-2}] \leq \beta_2.$*
- $\|\hat{z}\| < \infty$ casi siempre $\forall k.$
- \tilde{z} es una solución asintóticamente estable del sistema de ecuaciones $\dot{z}(t) = \partial_{z^*} f(z).$
- *Dado $G(\tilde{z}) = \{z_0 : \lim_{t \rightarrow \infty} z(t|z_0) = \tilde{z}\}$ existe un compacto $S \subseteq G(\mathbf{z}^*)$ tal que $\hat{z}_k \in S$ infinitamente frecuente por casi todos los puntos muestrales.*

Entonces,

$$\lim_{k \rightarrow \infty} \hat{z}_k = \tilde{z}, \quad (4.9)$$

esto es, CSPSA converge a la solución del problema de optimización.

Las ganancias a_k y c_k anteriormente elegidas coinciden con las hipótesis de los teoremas 1 y 2. Los valores de a, A, s, b y r deben ser ajustados para optimizar la tasa de convergencia y dependen de la función objetivo. Para satisfacer los teoremas, las componentes de Δ_k deben ser generadas independientemente seleccionando en cada iteración con igual probabilidad valores en el conjunto $\{e^{2iv_p}\}$ con $p = 0, \dots, 3$. Una elección puede ser $v_p = \{0, \pi/2, \pi, 3\pi/2\}$. Con esto las componentes $\Delta_{k,i}$ toman valores del conjunto $\{1, -1, i, -i\}$.

El algoritmo SPSA [22] puede ser visto como un caso particular de CSPSA para funciones reales, $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Al igual que en CSPSA, usa una estimación del gradiente, en este caso real, para encontrar mínimos (o máximos) de una función $f(\theta)$, con θ un vector en los reales. Así, este método iterativo genera una secuencia de estimados θ_k del vector $\tilde{\theta}$ que minimiza nuestra función. El algoritmo 1 es el mismo que en CSPSA, con la particularidad que la función objetivo debe tener parámetros reales, y que el vector Δ debe tener componentes elegidas independientes unas de otras del conjunto $\{1, -1\}$, distribuidas de acuerdo a la distribución de Bernoulli simétrica.

Algorithm 1 CSPSA-SPSA

Require:
 K , un número entero;

 ϕ ($\in \mathbb{C}^n$ para CSPSA, y $\in \mathbb{R}^n$ para SPSA), un vector semilla inicial;

 $h(\phi, \xi)$, una medición con ruido de la función a optimizar;

 $c(k)$ y $a(k)$, ganancias que cumplan las hipótesis de convergencia;

Ensure: $\tilde{\phi} = \phi(K)$;

procedure CSPSA-SPSA(ψ, ϕ_0, K, N)

for $k \leftarrow 1, K$ **do**
 $\Delta \leftarrow$ con componentes en $\{+1, -1, +i, -i\}$ en CSPSA y $\{+1, -1\}$ en

SPSA

 $Y(k)_+ \leftarrow h(\phi_0 + c_k \Delta, \xi_+(k))$
 $Y(k)_- \leftarrow h(\phi_0 - c_k \Delta, \xi_-(k))$
 $\phi(k+1) \leftarrow \phi(k) - a_k \frac{Y_+ - Y_-}{2c_k} \Delta$
end for
 $\tilde{\phi} = \phi(K)$
end procedure

SPSA también puede ser usado para optimizar funciones complejas mediante alguna parametrización real de los números complejos.



Capítulo 5

Tomografía Autoguiada

En este capítulo introduciremos la Tomografía Autoguiada, método en el que se fundamenta este trabajo. Mostramos resultados de esta aplicando los métodos de optimización CSPSA y SPSA.

La tomografía autoguiada (SGQT) usa un algoritmo iterativo para minimizar una medida de distancia entre el estado del sistema físico $|\psi\rangle$, y un estado estimado de este, $|\phi\rangle$. La distancia en principio puede ser cualquiera, con la condición de que pueda ser medida experimentalmente. En el caso de SGQT buscamos minimizar la infidelidad entre estados puros, definida por $I(\psi, \phi) = 1 - F(\psi, \phi) = 1 - |\langle\psi|\phi\rangle|^2$, estimada experimentalmente como $1 - n_\phi/N$, con n_ϕ el número de veces que el estado $|\phi\rangle$ es detectado, y N el número total de detecciones.

Tomando como guía la técnica de optimización estocástica CSPSA, usamos dos estados propuestos $|\phi_\pm\rangle$ y dos estimaciones de la infidelidad $h(\psi, \phi_\pm, \xi)$, con ξ el ruido en la medición, para obtener una aproximación o estimación del gradiente, que a su vez nos entrega la dirección del estado del sistema $|\psi\rangle$.

Supongamos que estamos en una iteración k , con $|\phi_k\rangle$ nuestro candidato a estado del sistema. Primero, debemos generar una dirección aleatoria a buscar $|\Delta_k\rangle$, para luego perturbar $|\phi_k\rangle$ en la dirección de $|\Delta_k\rangle$. Así, obtenemos dos estados $|\phi\rangle_{k\pm} = |\phi\rangle_k \pm c_k \Delta_k$ sobre los que estimamos la infidelidad $h(\psi, \phi_{k\pm}, \xi)$, tras lo cual debemos calcular el gradiente estimado en la dirección Δ_k ,

$$|g_k\rangle = \frac{h(\psi, \phi_{k+}, \xi) - h(\psi, \phi_{k-}, \xi)}{2c_k} |\Delta_k\rangle \quad (5.1)$$

$$= \frac{h(\psi, \phi_k + c_k \Delta_k, \xi) - h(\psi, \phi_k - c_k \Delta_k, \xi)}{2c_k} |\Delta_k\rangle. \quad (5.2)$$

Luego, calculamos la siguiente iteración vía

$$|\phi_{k+1}\rangle = |\phi_k\rangle + a_k |g_k\rangle. \quad (5.3)$$

Así, debemos utilizar el algoritmo 1, considerando como función objetivo la infidelidad, y preocupándonos de normalizar los vectores en los que se evalúa esta. Además, un punto importante es que cuando medimos experimentalmente la infidelidad entre un estado cualquiera $|\phi\rangle$ y el estado del sistema, realizamos mediciones proyectivas, por lo cual necesitamos construir una base que contenga a $|\phi\rangle$ y a $d - 1$ estados ortonormales. Esto lo realizamos construyendo un conjunto de $d - 1$ vectores aleatorios y linealmente independientes, que junto a $|\phi\rangle$ generan el espacio de estados del sistema. Luego, aplicamos el método de ortogonalización de Gram-Schmidt para construir una base ortonormal, conservando en esta el primer estado $|\phi\rangle$.

Una buena elección para las constantes en el problema de tomografía es $s = 1$, $t = 1/6$, $a = 3$, $A = 0$ y $c = 0.1$, cuyos valores han sido obtenidos mediante

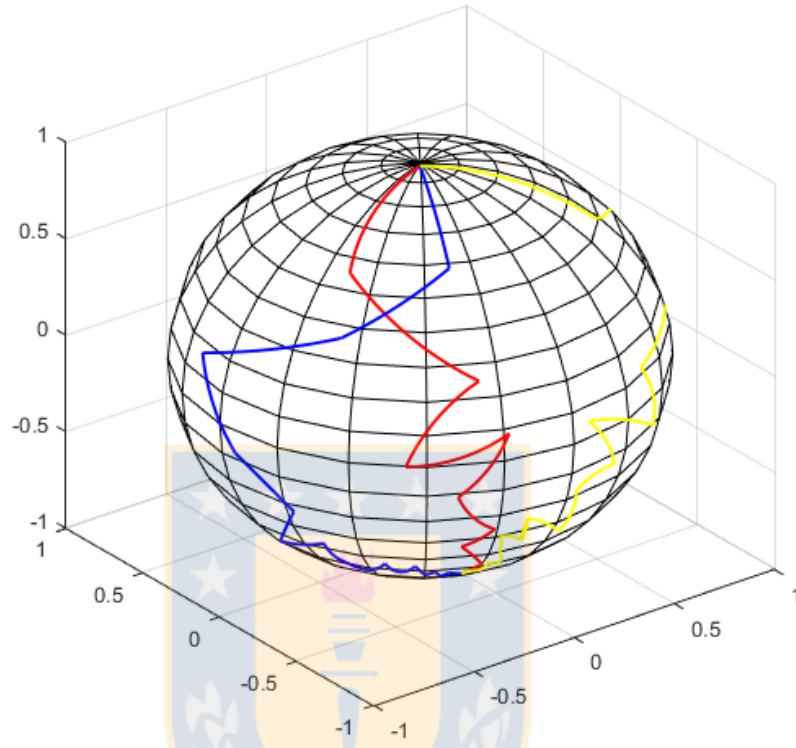


Figura 5.1: Gráfico de Tomografía Autoguiada usando CSPA en la esfera de Bloch para distintas trayectorias. El estado semilla va actualizándose de forma iterativa, pasando por estados que están cada vez más cerca del óptimo, hasta converger al estado del sistema. Fuente: Elaboración propia.

ensayo y error. El vector $|\Delta_k\rangle$ debe tener la misma dimensión que la del espacio de estados, y todas sus componentes pueden ser elegidas aleatoria e independientemente del conjunto $\{\pm i, \pm 1\}$. En total, el método usa un número total de partículas a medir $N_{tot} = 2Nk$. Esto viene de que por iteración k necesitamos medir 2 veces la función, utilizando en cada evaluación N partículas.

También podemos realizar tomografía autoguiada utilizando el método en los reales SPSA. El procedimiento es similar, pero ahora debemos parametrizar el

estado cuántico sobre la esfera de Bloch, y optimizar sobre los ángulos. Entonces, el vector $|\Delta_k\rangle$ ahora debe tener como dimensión el número de parámetros del estado cuántico, y sus componentes son elegidas aleatoria e independientemente con probabilidad 0.5 del conjunto $\{1, -1\}$. Las ganancias apropiadas son las asintóticas, $s = 1$, $t = 1/6$, $A = 3$, $B = 0$ y $C = 0.1$.

Así, el algoritmo en CSPA y SPSA va recorriendo distintos estados sobre la esfera de Bloch, cada uno generalmente con una menor infidelidad que el anterior, hasta llegar a la aproximación deseada, como vemos en la figura 5.1.

En 5.2 vemos el desempeño del método en dimensión 2. Graficamos la infidelidad promedio en el espacio de Hilbert sobre 10^4 estados desconocidos como función de las iteraciones k . Las líneas continuas representan el desempeño de CSPA, y las discontinuas de SPSA. Desde arriba hacia abajo en CSPA estimamos la infidelidad con distintos tamaños de ensemble, 10^1 , 10^2 , 10^3 y 10^4 , y lo mismo en SPSA. Podemos observar un menor error para todas las curvas de CSPA comparadas con SPSA, mostrando claramente una mejora en el rendimiento para un mismo número de partículas.

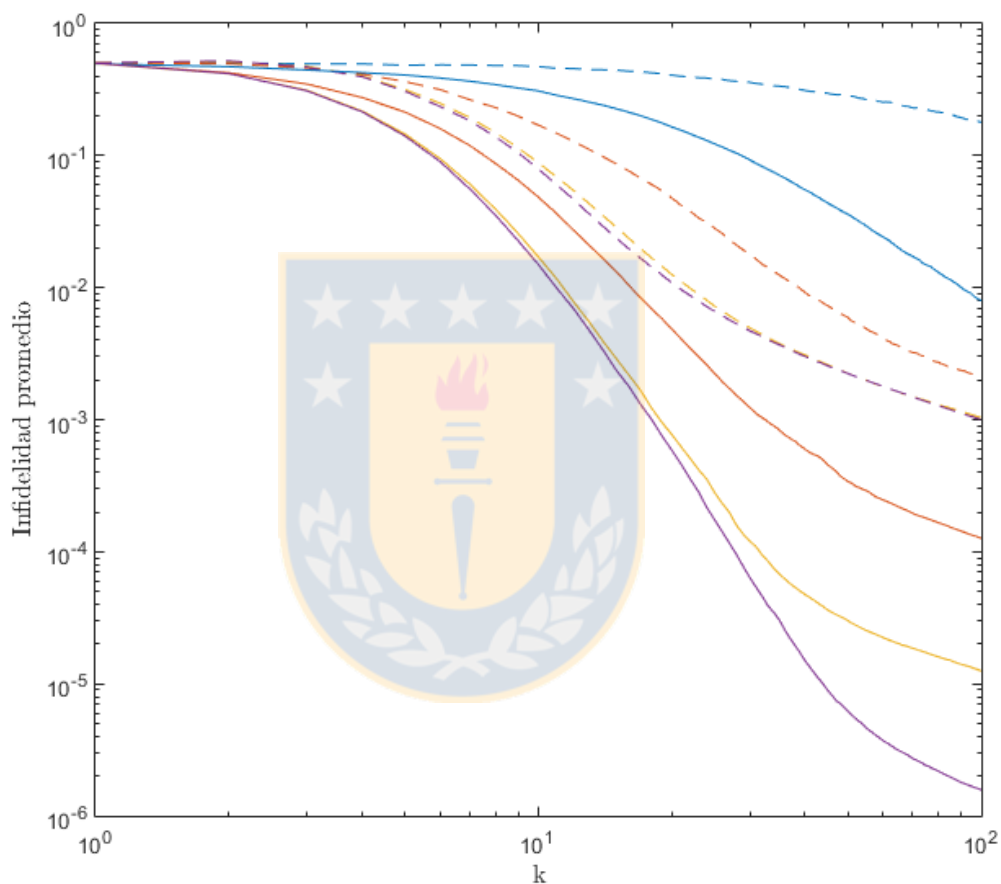
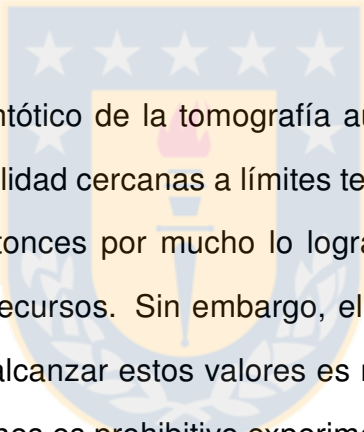


Figura 5.2: Gráfico de la Infidelidad Promedio vs Iteraciones para un qubit usando CSPA (curvas continuas) y SPSA (curvas discontinuas), con un distinto número de experimentos por evaluación de la función, 10^1 , 10^2 , 10^3 y 10^4 desde arriba a abajo para cada método. El promedio en cada curva es sobre 10^4 estados. Fuente: Elaboración propia.

Capítulo 6

Tomografía Autoguiada con Procesamiento Clásico



El desempeño asintótico de la tomografía autoguiada de estados puros alcanza cotas en la infidelidad cercanas a límites teóricos como Cramer-Rao o Gill-Massar, mejorando entonces por mucho lo logrado con tomografía estándar a un mismo número de recursos. Sin embargo, el número de iteraciones y mediciones a realizar para alcanzar estos valores es necesariamente muy grande, lo cual en algunas ocasiones es prohibitivo experimentalmente, y puede desmotivar su aplicación práctica. En cada iteración k de CSPA debemos necesariamente evaluar la infidelidad $I(\psi, \phi) = 1 - F(\psi, \phi_k)$ dos veces, de manera de aproximar un gradiente y mejorar el estado guess $|\phi\rangle_k$, acercándonos iterativamente al óptimo. Todas estas mediciones generan un gran número de datos que se usan en una iteración, y luego son simplemente descartados.

Así surge la motivación para desarrollar un protocolo complementario a CSPA, que aproveche el gran número de mediciones del método, y que gracias a esto sea capaz de aumentar la tasa de convergencia, para de esta forma utilizar menos recursos. Podemos lograr esto incorporando procesamiento clásico de datos en

el algoritmo. MLE es uno de los protocolos más usados en tomografía estándar como postprocesamiento, de manera de asegurar la positividad de la matriz densidad y aumentar la 'coincidencia' entre el estado del sistema y el obtenido en la tomografía. En la tomografía adaptativa podemos hacer lo mismo, incorporando MLE al protocolo de forma iterativa, sin requerir más mediciones, pero aumentando enormemente la tasa de convergencia.

La gran diferencia de aplicar MLE a tomografía estándar y tomografía adaptativa es que en nuestro caso lo hacemos de forma iterativa: En tomografía estándar MLE es aplicado como postprocesamiento, pues es una vez completado el proceso es cuando tenemos un estado estimado del sistema. En cambio, al ser CSPA una tomografía que en cada iteración entrega un nuevo estimado del sistema, podemos aplicar MLE en cada paso del algoritmo.

6.1 Algoritmo

En cada iteración de CSPA tenemos un estado guess $|\phi\rangle_k$, el cual es un estimado del real estado del sistema $|\psi\rangle$. A partir de esto, debemos medir la infidelidad en bases $B_{k\pm}$ que contengan perturbaciones del guess, $|\phi\rangle_{k\pm} = |\phi\rangle_k \pm c_k \Delta_k$. Cada una de las bases $B_{k\pm}$ contiene d estados, uno de los cuales es $|\phi\rangle_{k\pm}$ y el resto son estados ortogonales a él. Para construir las bases a medir elegimos $d - 1$ vectores aleatorios y linealmente independientes, que junto a $|\phi\rangle_{k\pm}$ generen el espacio de estados, y luego utilizamos el método de Gram-Schmidt para ortonormalizar, preocupándonos de conservar en la base el estado $|\phi\rangle_{k\pm}$ sin modificar, esto es, $B_{k+} = \{|\phi\rangle_{k+}, |\phi_2\rangle_{k+}, \dots, |\phi_d\rangle_{k+}\}$, y lo mismo para B_{k-} . Al medir

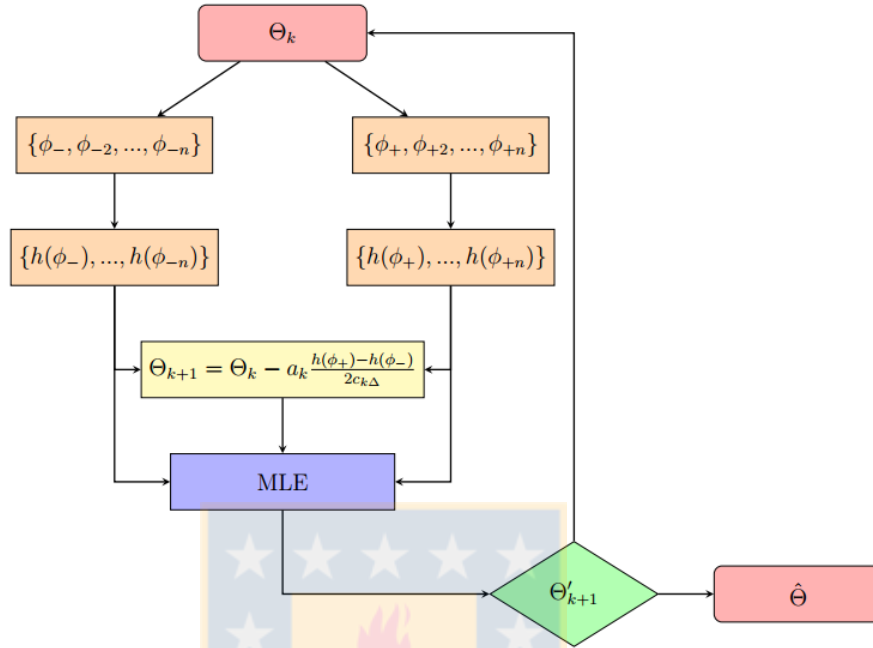


Figura 6.1: Diagrama de flujo de CSPA con MLE. Suponiendo que estamos en la iteración k , estimamos la infidelidad en dos bases, actualizamos el guess mediante CSPA y con estos datos optimizamos la verosimilitud para obtener un guess más cercano al óptimo, repitiendo luego el proceso hasta converger a $\hat{\Theta}$. Fuente: Elaboración propia.

sobre estas bases obtenemos dos conjuntos $F_{k\pm}$ de d estimaciones con ruido ξ de la infidelidad, $F_{k\pm} = \{h(|\phi\rangle_{k\pm}, \xi), \dots, h(|\phi_d\rangle_{k\pm}, \xi)\}$, cada uno de estos asociados a un vector de la base $B_{k\pm}$. Estos resultados los almacenamos, y usando el gradiente aproximado

$$|g\rangle_{k,i} = \frac{h_{k+} - h_{k-}}{2\beta_k \Delta_{k,i}^*}. \quad (6.1)$$

calculamos un nuevo estimado del sistema

$$|\phi\rangle_{k+1} = |\phi\rangle_k - \alpha_k |g\rangle_k. \quad (6.2)$$

Este vector normalizado es el estado inicial en MLE. Como coeficientes de la función de verosimilitud usamos los vectores del conjunto $\{B_{k+}, B_{k-}\}_{k=1}^n$ y $\{F_{k+}, F_{k-}\}_{k=1}^n$

donde n es la iteración actual. Esto significa que para ajustar la verosimilitud no solo usamos los datos obtenidos en esta iteración, sino que además agregamos los resultados de todas las mediciones anteriores. Luego minimizamos

$$\min_{\rho} F(\theta) = \sum_i^d \sum_j^n \frac{(f_{ij\pm} - \langle \phi_i |_{j\pm} \rho | \phi_i \rangle_{j\pm})^2}{2 \langle \phi_i |_{j\pm} \rho | \phi_i \rangle_{j\pm}}. \quad (6.3)$$

El índice j corre para las bases sobre las que hemos medido en las distintas iteraciones, mientras que el i etiqueta cada elemento de la base. La verosimilitud nos entrega un mejor estimado del sistema, $\rho = |\phi'\rangle_{k+1} \langle \phi'|_{k+1}$, el cual usamos luego como guess en la próxima iteración de CSPSA, $|\phi\rangle_{k+1} = |\phi'\rangle_{k+1}$, para nuevamente repetir el proceso. Es importante parametrizar el estado ρ en coordenadas n-esféricas, ya que así nos aseguramos de que el estado permanezca siendo puro a medida que iteramos en la optimización de la verosimilitud.

Más ilustrativo para comprender el funcionamiento del método puede resultar un diagrama de flujo, 6.1 .

En resumen, el procedimiento es el siguiente:

- Dado $|\phi\rangle_k$, generamos un Δ_k en \mathbb{C}^n .
- Calculamos $|\phi\rangle_{k\pm} = |\phi\rangle_k \pm \beta_k \Delta_k$, construimos las bases $B_{k\pm}$, medimos las infidelidades $F_{k\pm}$ asociadas a cada vector de la base, y guardamos los resultados.
- Estimamos el gradiente \hat{g}_k ,

$$|g\rangle_{k,i} = \frac{h_{k+} - h_{k-}}{2\beta_k \Delta_{k,i}^*}. \quad (6.4)$$

- Calculamos el guess a usar en el procesamiento clásico,

$$|\phi\rangle_{k+1} = |\phi\rangle_k - \alpha_k |g\rangle_k. \quad (6.5)$$

- Minimizamos el logaritmo de la función de Likelihood,

$$\min_{\rho} F(\theta) = \sum_i^d \sum_j^n \frac{(f_{ij\pm} - \langle \phi_i |_{j\pm} \rho | \phi_i \rangle_{j\pm})^2}{2 \langle \phi_i |_{j\pm} \rho | \phi_i \rangle_{j\pm}}. \quad (6.6)$$

con $\rho = |\phi'\rangle_{k+1} \langle \phi'|_{k+1}$.

- Por último, obtenemos el guess a usar en la próxima iteración de CSPSA a partir de $|\phi\rangle_{k+1} = |\phi'\rangle_{k+1}$.

Siguiendo este algoritmo hemos realizado experimentos numéricos para estimar su desempeño, utilizando el método de Montecarlo. Nuestra variable aleatoria es la infidelidad entre el estado del sistema $|\psi\rangle$ y nuestro estimado $|\phi\rangle$ a un dado número de iteraciones K . Esta función a su vez depende de las cuentas en los detectores del experimento, que siguen una distribución multinomial. Podemos simular la infidelidad para cada guess $|\phi\rangle_k$ a una iteración k como si hubiese sido medida en el laboratorio mediante la función 4, $I(k) = 1 - n_{\phi_k}/N$, para luego calcular su promedio sobre todos los guess $\{|\phi_1\rangle_k, \dots, |\phi_n\rangle_k\}$, obteniendo de esta forma una infidelidad $\bar{I}(k)$ estimada promedio. Dependiendo del número N de detecciones tenemos una mejor o peor estimación de la función, como nos indican los teoremas de convergencia del método de Montecarlo. Los n estimados iniciales $|\phi\rangle_1$ del algoritmo son elegidos según la medida de Haar, esto es, con coeficientes provenientes al azar de la distribución normal. Esto nos permite escoger estados de manera uniforme sobre la esfera de Bloch, y al promediar

Algorithm 2 CSPSA con MLE

Require:

- K , un número entero;
- $\phi \in \mathbb{C}^n$, un vector semilla inicial;
- $inf(\phi, \xi)$, una medición con ruido de la función infidelidad;
- $c(k)$ y $a(k)$, ganancias que cumplan las hipótesis de convergencia;

Ensure: $\tilde{\phi} = \phi(K)$;**procedure** CSPSA(ψ, ϕ_0, K, N)**for** $k \leftarrow 1, K$ **do**

$\Delta \leftarrow$ vector cuyas componentes son elegidas aleatoriamente de $\{+1, -1, +i, -i\}$

$$\phi_{\pm}(k) \leftarrow \phi(k) \pm c_k \Delta$$

$$\phi_{\pm}(k) \leftarrow \phi_{\pm}(k) / \|\phi_{\pm}(k)\|$$

$$inf_{\pm}(k) \leftarrow h(\phi_{\pm}(k), \xi_{\pm}(k))$$

$$\phi(k+1) \leftarrow \phi(k) - a_k \frac{inf_+ - inf_-}{2c_k} \Delta$$

$$\phi(k+1) \leftarrow \phi(k+1) / \|\phi(k+1)\|$$

$$\phi(k+1) \leftarrow \operatorname{argmin}[LF(\phi(k+1), \phi_{\pm}(1:k), inf_{\pm}(1:k))]$$

end for

$$\tilde{\phi} = \phi(K)$$

end procedure

ponderamos todas las zonas del espacio de estados de la misma forma.

Otra característica de particular importancia en el algoritmo son las ganancias de CSPSA,

$$a_k = \frac{a}{(k+1+A)^s} \quad c_k = \frac{b}{(k+1)^t}. \quad (6.7)$$

Dependiendo del valor que asociemos a ellas podemos tener distintos niveles de convergencia. Empíricamente, probamos que para la mayoría de las dimensiones y aproximaciones de la infidelidad las ganancias apropiadas eran las estándar, esto es, $s = 0.606$, $t = 1/6$, $a = 3$, $A = 0$, mientras que b varia de acuerdo al error que tengamos en la infidelidad. Así, mientras mejor sea la estimación de la fun-

Algorithm 3 Función de Verosimilitud

Require:

$\phi \in \mathbb{C}^n$, un vector semilla próximo al óptimo;
 ϕ_{\pm} , matriz cuyas columnas son vectores de las bases en que medimos;
 $inf(\phi_{\pm}, \xi)$, un vector de mediciones de la infidelidad asociadas a cada vector en ϕ_{\pm} ;

Ensure: LF , función de verosimilitud;

```

procedure LF( $\phi, \phi_{\pm}, inf_{\pm}$ )
   $inf_t \leftarrow 1 - \phi'_{\pm} \phi \phi_{\pm}$ 
   $inf_t \leftarrow \text{diag}(inf_t)$ 
   $LF \leftarrow \sqrt{1/2inf_t}(inf_t - inf)$ 
   $LF \leftarrow ||LF||^2$ 
end procedure
  
```

ción, menor debe ser el valor de b . Si calculamos la función con 10 partículas, $b = 0.3$, con 10^2 , $b = 0.18$, con 10^3 $b = 0.07$, con 10^4 $b = 0.05$ y por último con 10^5 $b = 0.02$. Otras combinaciones en las ganancias también pueden llevarnos a valores de convergencia similares, pero mediante ensayo y error son estas las que funcionan en la mayor parte de los casos.

El algoritmo que calcula la función de verosimilitud es **3**. Para encontrar el mínimo de la verosimilitud usamos la función *fminsearch* de Matlab, la cual está basada en el método de optimización Nelder-Mead.

Como no hemos hecho el experimento, la infidelidad debe ser simulada por el algoritmo **4**. Las variables de entrada son el estado del sistema ψ , el estado guess ϕ , el número de cuentas en cada estimación de la infidelidad N_{exp} , y una matriz con estados aleatorios M . El algoritmo nos entrega una matriz B cuyas columnas son los d vectores de la base de medición, y que contiene a ϕ , y las infidelidades

Algorithm 4 Infidelidad Simulada

Require:

$\psi \in \mathbb{C}^n$, el estado del sistema;
 $\phi \in \mathbb{C}^n$, estado sobre el que queremos calcular la infidelidad;
 $M \in \mathbb{C}^n \times \mathbb{C}^{n-1}$ matriz con componentes aleatorias en \mathbb{C} ;
 N_{exp} , tamaño del ensemble que queremos medir;

Ensure:

B , base que contiene a ϕ sobre la que medimos;
 h , infidelidad simulada para cada elemento de la base B ;

procedure $h(\psi, \phi, N_{exp})$

$B \leftarrow \{\phi, M\}$

$B \leftarrow \text{Gram-Schmidt}(B)$

for $k \leftarrow 1, n$ **do**

$h(k) \leftarrow \text{abs}(\text{dot}(\psi, B(k, :)))^2$

end for

$h(k) \leftarrow 1 - \text{multinomial_random_number}(N_{exp}, h(k)) / N_{exp}$

end procedure

estimadas asociadas a cada uno de estos vectores, h . Para lograr esto generamos una matriz B cuyas columnas forman un conjunto de d vectores aleatorios linealmente independientes que contiene a ϕ , y luego aplicamos el método de ortogonalización de Gram-Schmidt, conservando ϕ . Luego, calculamos las probabilidades de transición teóricas entre los vectores de la base y el estado del sistema, para después generar números aleatorios con distribución multinomial asociados a cada una de las probabilidades teóricas, los que normalizados por el número total de experimentos nos entregan una fidelidad simulada.

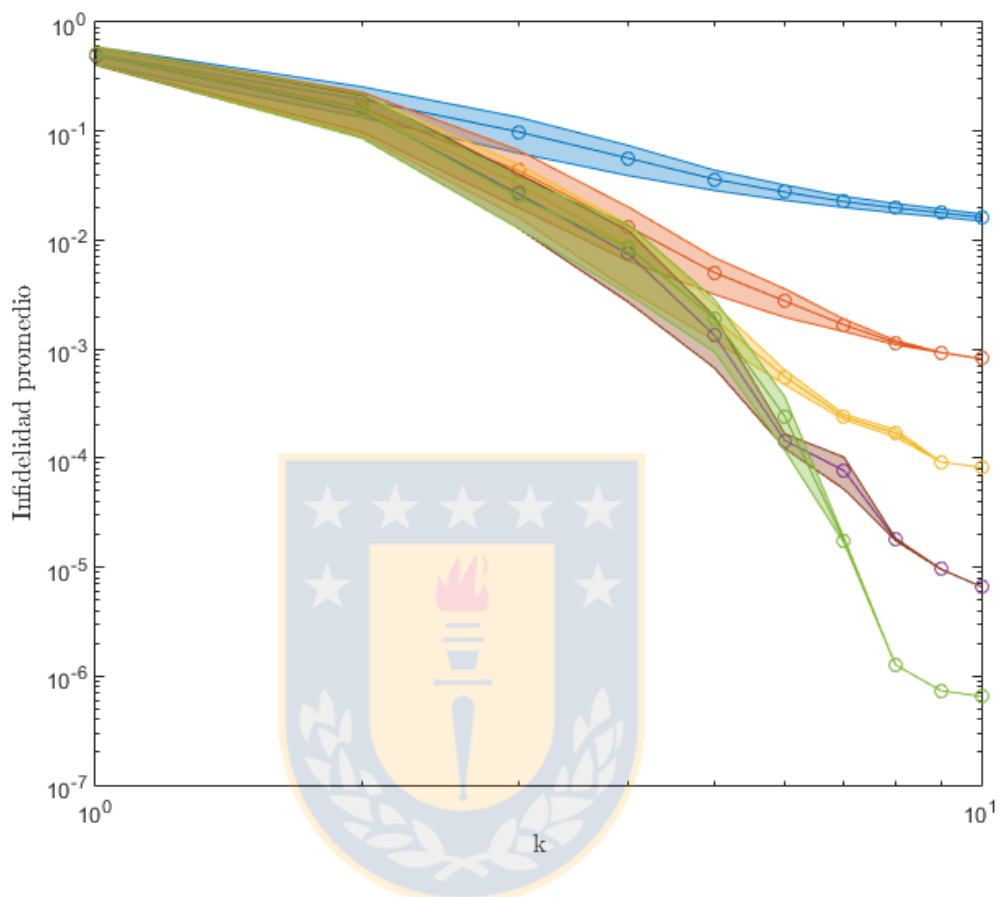


Figura 6.2: Gráfico de la Infidelidad Promedio vs Iteraciones para un qubit, usando CSPA con MLE, para un distinto número de experimentos por evaluación de la función, 10^1 , 10^2 , 10^3 , 10^4 y 10^5 desde arriba a abajo. La infidelidad está promediada sobre 10^4 estados desconocidos, y las zonas sombreadas muestran la varianza en los datos. Fuente: Elaboración propia.

6.2 Resultados

En la figura 6.2 vemos la infidelidad promedio en función del número de iteraciones para dimensión 2, calculada usando CSPA con MLE. Desde arriba hacia abajo observamos curvas con distintas aproximaciones en la infidelidad, y por lo tanto, distintos ensembles de partículas. En la primera curva (azul) aproximamos

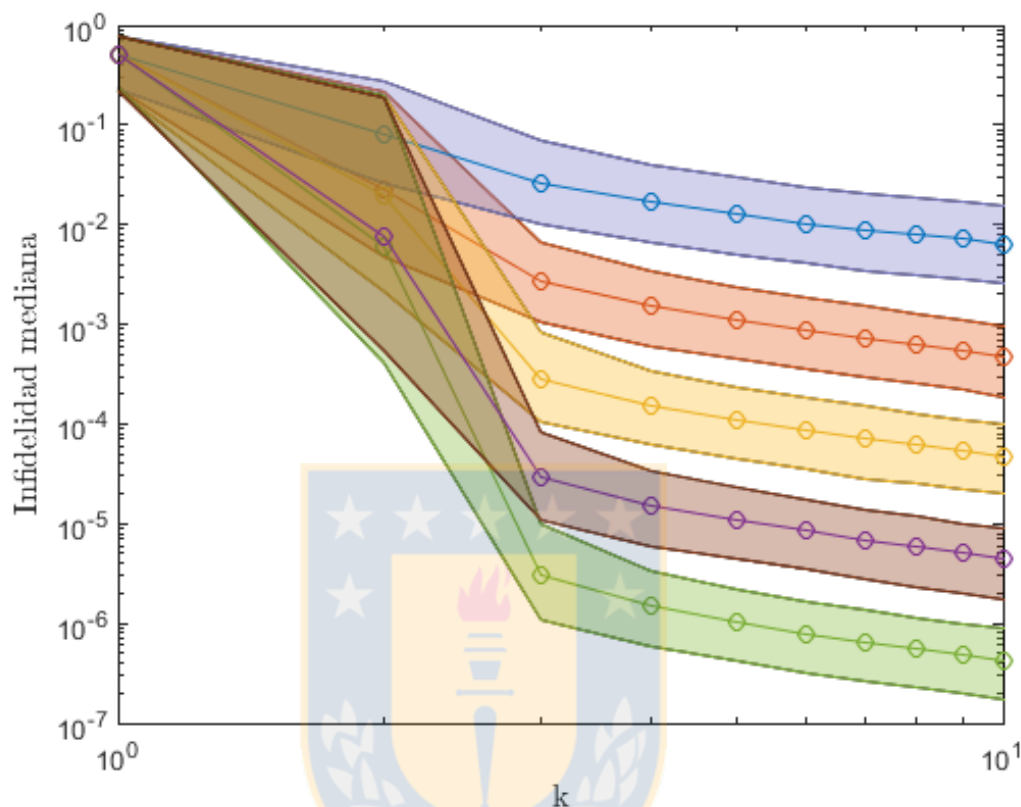


Figura 6.3: Gráfico de la Infidelidad Mediana vs Iteraciones para un qubit, usando CSPSA con MLE, para un distinto número de experimentos por evaluación de la función, 10^1 , 10^2 , 10^3 , 10^4 y 10^5 desde arriba a abajo. La infidelidad mediana está calculada sobre 10^4 estados desconocidos, y las zonas sombreadas muestran los rangos intercuantiles de los datos. Fuente: Elaboración propia.

la infidelidad con 10 partículas por medición, en la segunda 10^2 , la tercera 10^3 , la cuarta 10^4 y la última con 10^5 detecciones por evaluación. Todas las simulaciones están hechas para 10 iteraciones, y las zonas sombreadas corresponden a la varianza en los datos.

Observamos una mayor convergencia en la infidelidad a medida que el número de iteraciones aumenta, llegando a valores en la décima iteración que son del

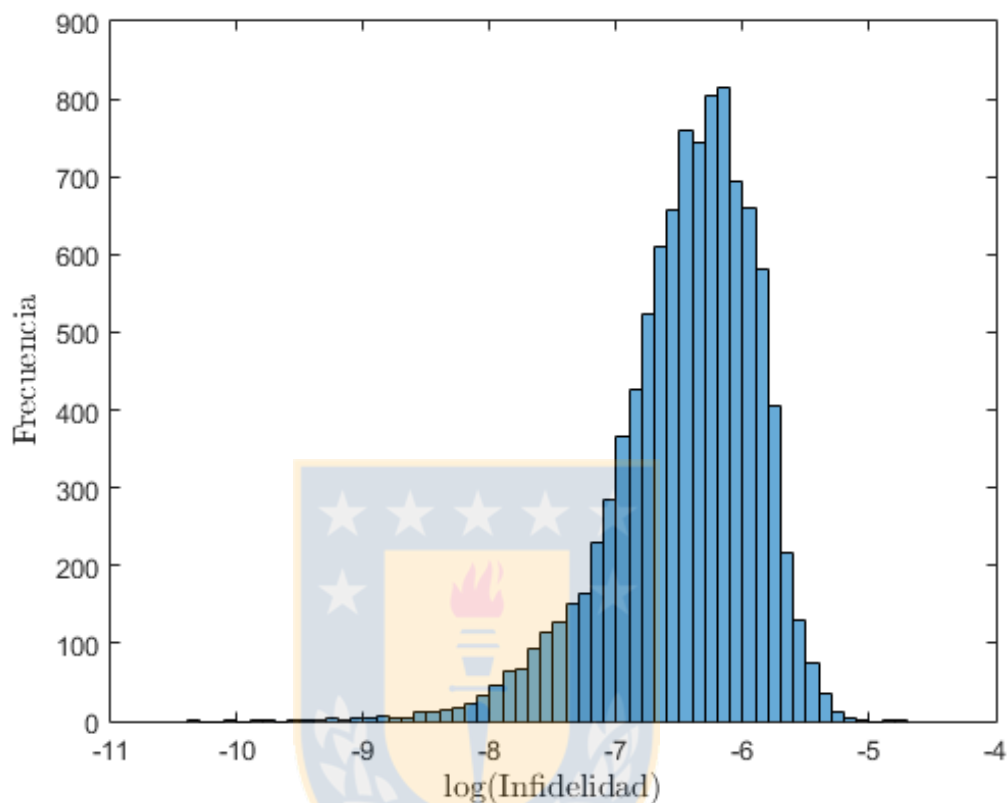


Figura 6.4: Gráfico de la Distribución de infidelidades para un qubit a la iteración 10 usando CSPA con MLE sobre 10^4 repeticiones. La infidelidad es estimada con 10^5 detecciones. Fuente: Elaboración propia.

orden de $1/N_{tot}$, donde N_{tot} es el número total de partículas medidas en el experimento. También observamos que a medida que medimos un mayor número de partículas en la infidelidad la tasa de convergencia aumenta. Esto se explica porque al tener una función estimada con más partículas, las fluctuaciones debidas a la estadística son menores, y la precisión en aproximar el gradiente y en aplicar MLE aumenta. Además, la varianza de los datos va haciéndose más estrecha con las iteraciones, lo cual es una medida de que los datos se van agrupando alrededor del promedio a medida que el algoritmo avanza. Esto también

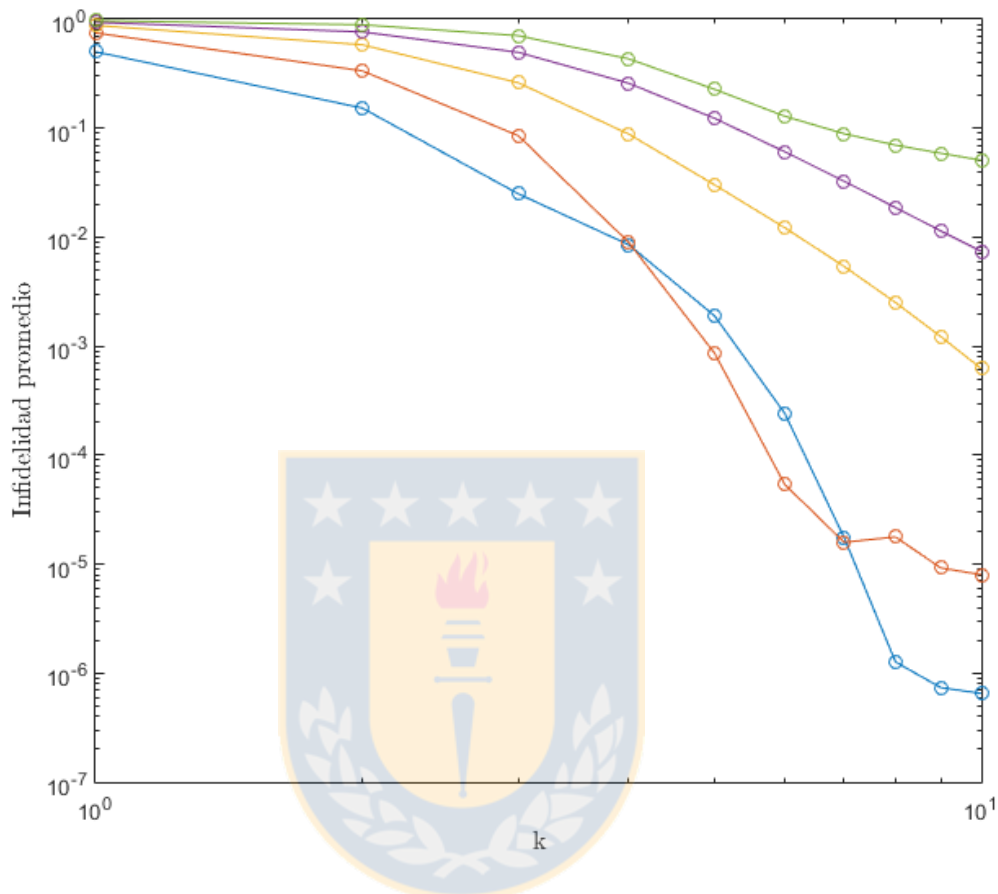


Figura 6.5: Gráfico de la Infidelidad Promedio vs Iteraciones para dimensiones 32, 16, 8, 4 y 2 usando CSPA con MLE, y estimando la infidelidad con 10^4 partículas. El error crece a medida que aumenta la dimensión. Fuente: Elaboración propia.

es un signo de que el estado inicial en la aproximación es de poca importancia en la convergencia del algoritmo, ya que pasado un número de iteraciones todos convergen a aproximadamente el mismo punto.

En la figura 6.3 observamos la mediana de la infidelidad en función del número de iteraciones para dimensión 2, con el área sombreada dada por el rango intercuartil. Se observan las mismas tendencias en los datos que al usar la media,

N_{total}	CSPSA MLE	CSPSA
$2 * 10^2$	$3.02 * 10^{-2}$	$3.04 * 10^{-1}$
$2 * 10^3$	$8.94 * 10^{-4}$	$4.78 * 10^{-2}$
$2 * 10^4$	$7.92 * 10^{-5}$	$1.85 * 10^{-2}$
$2 * 10^5$	$6.86 * 10^{-6}$	$1.54 * 10^{-2}$
$2 * 10^6$	$6.73 * 10^{-7}$	$1.51 * 10^{-2}$

Tabla 6.1: Infidelidad para un qubit en función del tamaño de ensemble para CSPSA y CSPSA con MLE a las 10 iteraciones. Fuente: Elaboración propia.

obteniendo resultados del mismo orden para los distintos ensembles de partículas medidas. Esto nos dice que las distribuciones de los datos están relativamente centradas, y que el método tiene poco sesgo, como vemos en la figura 6.4, donde apreciamos una distribución en los datos relativamente simétrica. Además, el ancho del rango intercuartil es de menos de un orden, lo que nos asegura que al menos el 75% de las estimaciones son de un nivel similar al de la mediana.

La figura 6.5 representa el comportamiento del método para mayores dimensiones. Desde arriba hacia abajo tenemos, respectivamente, dimensión 32, 16, 8, 4 y 2. En todos estos experimentos usamos el mismo ensemble, $N = 10^5$ partículas por evaluación de la función, dándonos un $N_{tot} = 2Nk = 2 * 10^6$. A medida que la dimensión aumenta obtenemos infidelidades mayores, por lo que la estimación del estado del sistema empeora. Esto puede explicarse porque al ser la dimensión mayor la calidad de la aproximación en la infidelidad empeora, al tener que repartir el recurso en más salidas del experimento.

Como podemos apreciar en de la tabla 6.1, bajo las mismas condiciones, esto es, mismo numero de iteraciones y de recurso por estimación de la función,

N_{total}	Tomografía Estándar	CSPSA MLE	SPSA MLE	CSPSA
$2 * 10^2$	$3.48 * 10^{-2}$	$3.02 * 10^{-2}$	$1.49 * 10^{-1}$	
$2 * 10^3$	$8.44 * 10^{-3}$	$8.94 * 10^{-4}$	$4.78 * 10^{-2}$	$9.90 * 10^{-3}$
$2 * 10^4$	$2.40 * 10^{-3}$	$7.92 * 10^{-5}$	$1.72 * 10^{-2}$	$1.24 * 10^{-4}$
$2 * 10^5$	$7.61 * 10^{-4}$	$6.86 * 10^{-6}$	$1.32 * 10^{-2}$	$1.22 * 10^{-5}$
$2 * 10^6$	$2.37 * 10^{-4}$	$6.73 * 10^{-7}$	$1.12 * 10^{-2}$	$1.73 * 10^{-6}$

Tabla 6.2: Infidelidad para un qubit en función del tamaño de ensemble para distintos esquemas tomográficos. Fuente: Elaboración propia.

CSPSA con MLE obtiene resultados ordenes mejor que CSPSA en la tomografía sobre qubits. Esta diferencia aumenta aún más cuando tenemos un ensemble mayor, debido a que la convergencia para las primeras iteraciones de CSPSA no depende mucho de la precisión de la función estimada, siendo esto relevante en las iteraciones finales, como podemos ver en la figura 5.2. De la gráfica observamos que, para al menos las primeras iteraciones, el error en la estimación es del mismo orden, independientemente del recurso invertido en obtenerlas. Así, es claro el aumento en la tasa de convergencia del algoritmo al aplicar MLE.

En la tabla 6.2 realizamos la comparativa de 4 diferentes métodos de tomografía. Comparamos las infidelidades promedio en función del número total de detecciones para dimensión 2. CSPSA con MLE presenta un mejor desempeño con respecto a tomografía estándar y SPSA con MLE para todo número de detecciones, obteniendo resultados de al menos un orden mejor que la mejor configuración encontrada de CSPSA. Este último método tiene la ventaja de no tener postprocesamiento. Sin embargo, las simulaciones con CSPSA fueron hechas con 100 iteraciones, lo que implica medir en al menos 200 bases distintas, mien-

N_{total}	CSPSA MLE	Gill-Massar Puro	Gill-Massar	Cramer-Rao
$2 * 10^2$	$3.02 * 10^{-2}$	$5 * 10^{-3}$	$1.13 * 10^{-2}$	$3.75 * 10^{-3}$
$2 * 10^3$	$8.94 * 10^{-4}$	$5 * 10^{-4}$	$1.13 * 10^{-3}$	$3.75 * 10^{-4}$
$2 * 10^4$	$7.92 * 10^{-5}$	$5 * 10^{-5}$	$1.13 * 10^{-4}$	$3.75 * 10^{-5}$
$2 * 10^5$	$6.86 * 10^{-6}$	$5 * 10^{-6}$	$1.13 * 10^{-5}$	$3.75 * 10^{-6}$
$2 * 10^6$	$6.73 * 10^{-7}$	$5 * 10^{-7}$	$1.13 * 10^{-6}$	$3.75 * 10^{-7}$

Tabla 6.3: Infidelidad para un qubit en función del tamaño de ensemble para CSPSA con MLE y las cotas teóricas. Fuente: Elaboración propia.

tras que las de CSPSA con MLE solo necesitaron 10 iteraciones.

En nuestro algoritmo debemos optimizar la función de verosimilitud iteración tras iteración, lo cual computacionalmente puede parecer, a priori, más demandante que otros métodos, que solo deben aplicar MLE una vez como postprocesamiento. El número de términos a evaluar en la verosimilitud de CSPSA con MLE para una dada iteración es lineal en la dimensión, ya que va como $2dk$, con d la dimensión espacio de estados y k la iteración. Mientras, en tomografía estándar, la verosimilitud tiene al menos un total de d^2 términos. Así, al aumentar la dimensión en la que trabajamos la complejidad de la verosimilitud crece más rápido en tomografía estándar que en nuestro método, siendo por lo tanto su gasto computacional mayor. Por ejemplo, simular CSPSA con MLE para un solo estado en dimensión 32 nos tomó 3 minutos, mucho menos tiempo que el gastado en tomografía estándar para encontrar el mismo estado. Aparte de la complejidad computacional, en tomografía estándar debemos medir al menos $d^2 - 1$ observables, comparado con $2dk$ de CSPSA con MLE. Luego en altas dimensiones la complejidad experimental puede llegar a ser mayor, al tener que medir en un

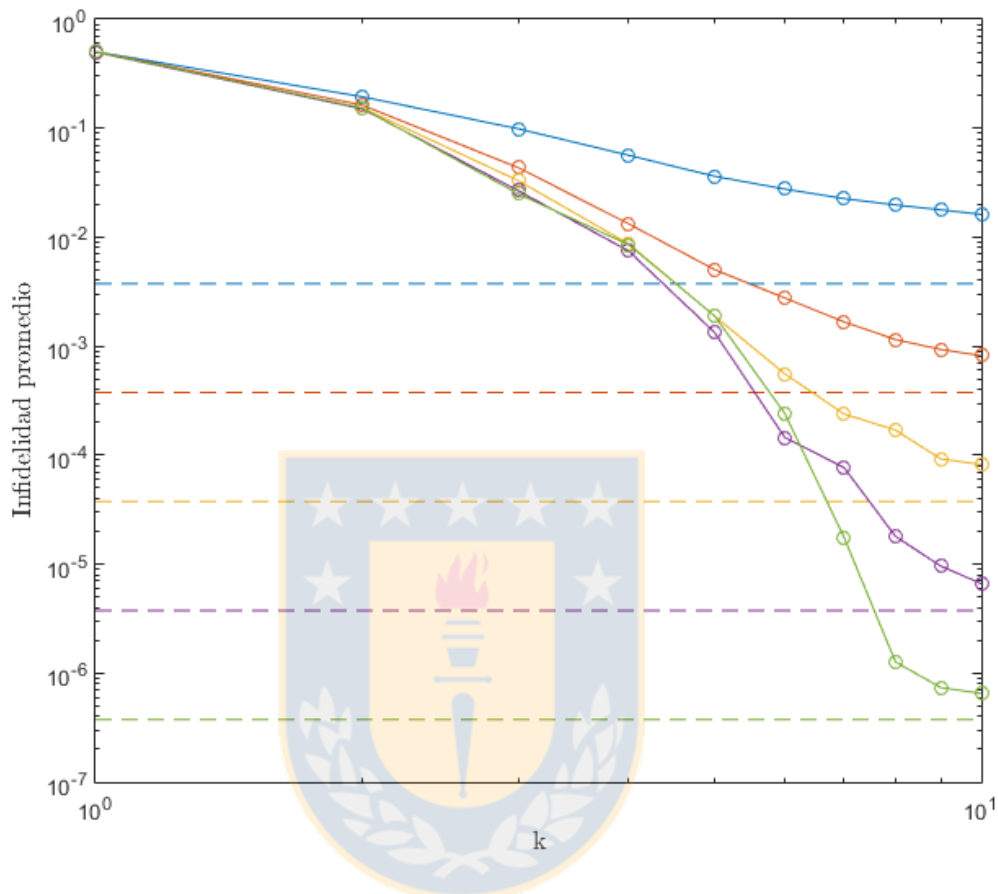


Figura 6.6: Gráfico de la Infidelidad Promedio vs Iteraciones para un qubit usando CSPA con MLE comparado con la cota de Cramer-Rao cuántica para un mismo tamaño de ensemble. El promedio es sobre 10^4 estados desconocidos, y la infidelidad para cada curva es calculada con 10^1 , 10^2 , 10^3 , 10^4 y 10^5 partículas, desde arriba a abajo, por cada evaluación de la función. Fuente: Elaboración propia.

mayor número de bases.

De los capítulos anteriores vimos que la cota de Cramer-Rao cuántica para el error en estimar matrices densidad es del orden de $\frac{d^2-1}{4N_{tot}}$, lo que para dimensión $d = 2$ equivale a $\frac{3}{4N_{tot}}$. En la figura 6.6 vemos esta cota comparada con los valores

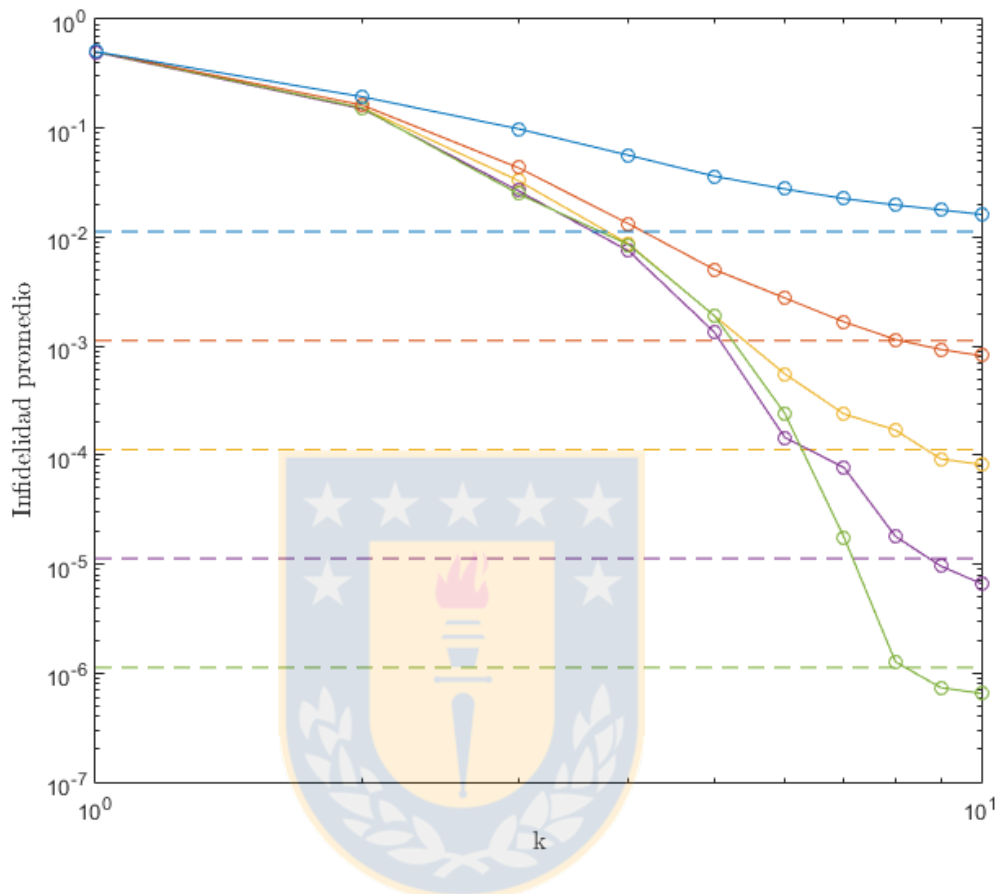


Figura 6.7: Gráfico de la Infidelidad Promedio vs Iteraciones para un qubit usando CSPSA con MLE comparado con la cota de Gill-Massar para un mismo tamaño de ensemble. El promedio es sobre 10^4 estados desconocidos, y la infidelidad para cada curva es calculada con 10^1 , 10^2 , 10^3 , 10^4 y 10^5 partículas, desde arriba a abajo, por cada evaluación de la función. Fuente: Elaboración propia.

promedio de infidelidad alcanzados por el método para los mismos tamaños de ensemble N_{tot} en dimensión 2. Podemos ver que nuestros resultados alcanzan valores del mismo orden que lo impuesto por esta cota teórica en la infidelidad.

Mientras tanto, la cota de Gill-Massar es $\frac{1}{4N_{tot}}(d+1)^2(d-1)$, también calculada considerando matrices densidad. Para $d = 2$ equivale a $\frac{9}{4N_{tot}}$, con N_{tot} el

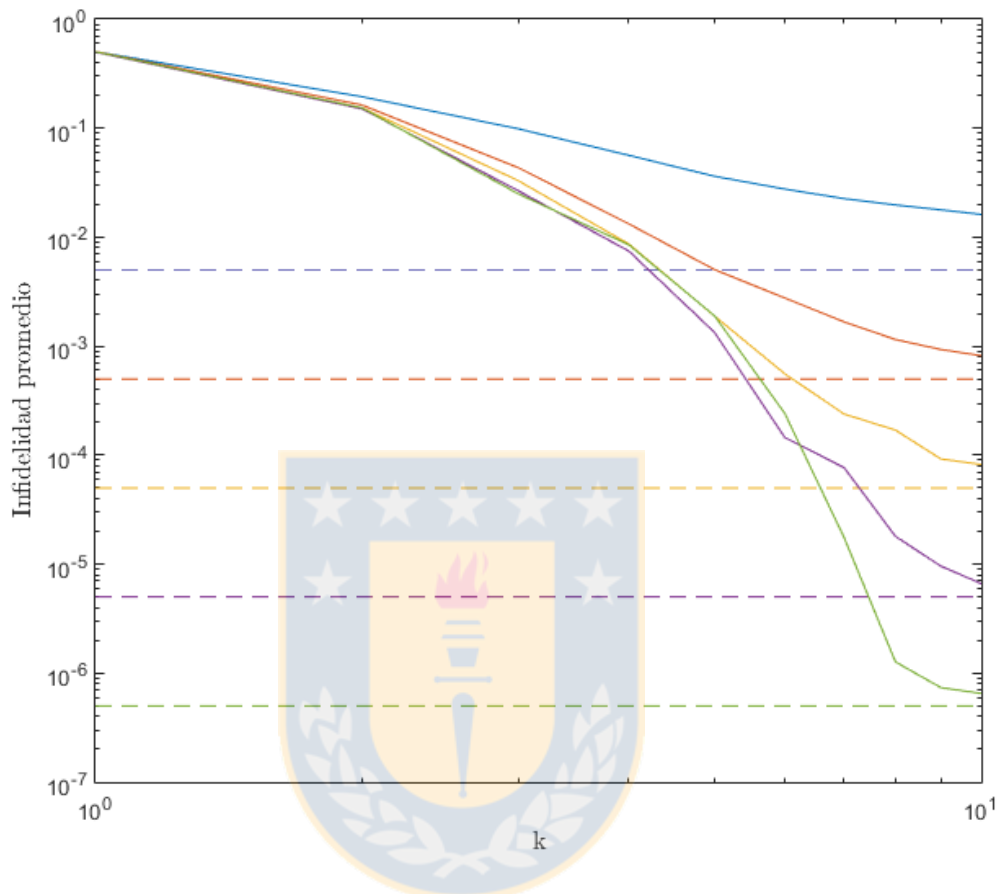


Figura 6.8: Gráfico de la Infidelidad Promedio vs Iteraciones para un qubit usando CSPSA con MLE comparado con la cota de Gill-Massar de estados puros para un mismo tamaño de ensemble. El promedio es sobre 10^4 estados desconocidos, y la infidelidad para cada curva es calculada con 10^1 , 10^2 , 10^3 , 10^4 y 10^5 partículas, desde arriba a abajo, por cada evaluación de la función. Fuente: Elaboración propia.

tamaño total del ensemble. En la figura 6.7 observamos esta cota comparada con los resultados de las simulaciones del método en dimensión 2. Vemos que, a excepción de la curva azul, el ensemble más pequeño, logramos superar este límite teórico en todos los demás casos.

Un aspecto importante acerca de esto es que nuestro método está formulado solo para estados puros, con $2d - 2$ parámetros reales a estimar, mientras que estas cotas teóricas están calculadas pensando en protocolos tomográficos estándar, diseñadas para reconstruir matrices densidad con $d^2 - 1$ parámetros. Esto puede ayudar a que CSPA con MLE obtenga mejores resultados, ya que tenemos a priori la información de que el estado del sistema es puro.

Usando la información de que el estado es puro, la cota de Gill-Massar está dada por

$$\varepsilon = \frac{d - 1}{N_{tot}}, \quad (6.8)$$

lo cual en dimensión $d = 2$ equivale a $\varepsilon = 1/N_{tot}$. En la figura 6.8 podemos ver la cota y los resultados de CSPA para el mismo tamaño N_{tot} de detecciones totales. Así, el valor de esta cota se hace más pequeño, permitiendo un menor error en métodos pensados para estimar estados puros, y por lo tanto alejándose de los datos obtenidos con CSPA y MLE, a pesar de estar en el mismo orden. Sin embargo, esta mejora en las cotas teóricas es válida solo para métodos que estiman los $2d - 2$ parámetros de estados puros, no pudiendo por lo tanto ser alcanzada por métodos tomográficos estándar.

Finalmente, en la figura 6.9 vemos el desempeño del método para estados puros con ruido blanco. Construimos operadores densidad aleatorios de la forma

$$\rho = \lambda |\psi\rangle\langle\psi| + \frac{1 - \lambda}{d} \mathbb{I}, \quad (6.9)$$

y usamos el método para encontrar el estado $|\psi\rangle$ del sistema. Con $\lambda = 0.99$ los estados tienen una pureza del 99.9%, y como vemos en el gráfico la convergencia

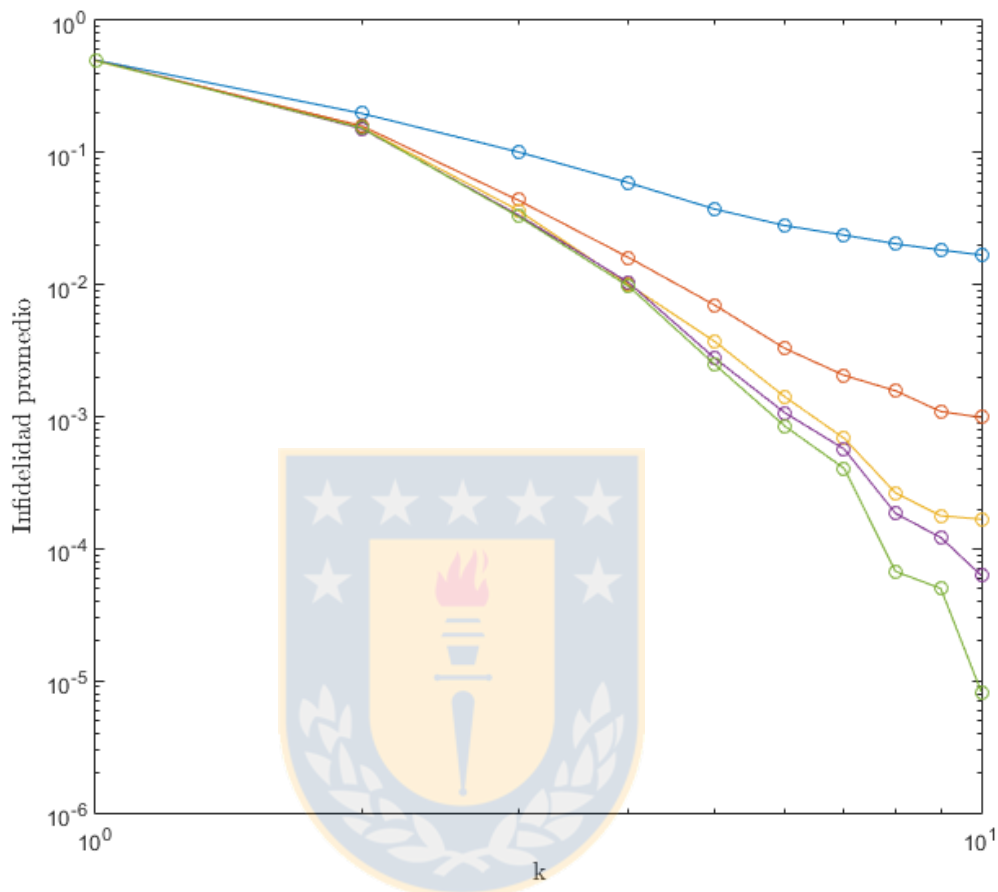


Figura 6.9: Gráfico de la Infidelidad Promedio vs Iteraciones para un estado cuasi puro con pureza 99.9% usando CSPA con MLE en dimensión 2. El promedio es sobre 10^4 estados desconocidos, y la infidelidad para cada curva es calculada con 10^1 , 10^2 , 10^3 , 10^4 y 10^5 partículas, desde arriba a abajo, por cada evaluación de la función. Fuente: Elaboración propia.

del algoritmo es menor a lo obtenido anteriormente, llegando a resultados en la infidelidad de alrededor un orden mas altos. Esto se explica porque al ser CSPA con MLE un método que estima estados puros, nos quedamos con el autovector asociado mayor autovalor de la matriz densidad, omitiendo los demás, lo cual puede aumentar el error en la estimación. Como hay mayor error es adecuado

una ganancia b mas grande, reemplazando los valores anteriores en el algoritmo por $b = 0.2$. A pesar de todo, los resultados son mejores que los obtenidos con tomografía estándar, como podemos apreciar al comparar con la tabla 6.1.



Capítulo 7

Detección de Entrelazamiento

Usando los métodos de optimización CSPA y SPSA podemos determinar el entrelazamiento de un estado desconocido en un sistema compuesto AB , sin la necesidad de conocer el estado, y por lo tanto, evitando realizar tomografía. Para esto, encontramos el mayor autovalor (coeficiente de Schmidt) de la matriz densidad reducida de uno de los sistemas, $\rho_A = \sum_i \lambda_i^2 |i_A\rangle\langle i_A|$, usando como función objetivo en el método de optimización la infidelidad $I(\phi, \rho) = \langle \phi | \rho | \phi \rangle$ sobre la matriz densidad reducida de ρ_A (o también ρ_B) y maximizando. Luego, si el resultado es 1 el estado es separable, y si no, entrelazado.

La función $I(\phi)$ puede ser aproximada midiendo en el laboratorio ensembles con N copias del mismo estado ρ , sobre una base que contenga el estado $|\phi\rangle$, esto es, $I = 1 - n_\phi/N$. Así, obtenemos una estimación de la infidelidad $h(\phi, \xi)$, con ξ el error asociado a las mediciones.

El procedimiento a seguir es el mismo que en la tomografía autoguiada, y por lo tanto usamos el algoritmo 1. La diferencia es que los datos usados en el protocolo ahora no provienen de estimar infidelidades sobre estados puros, sino

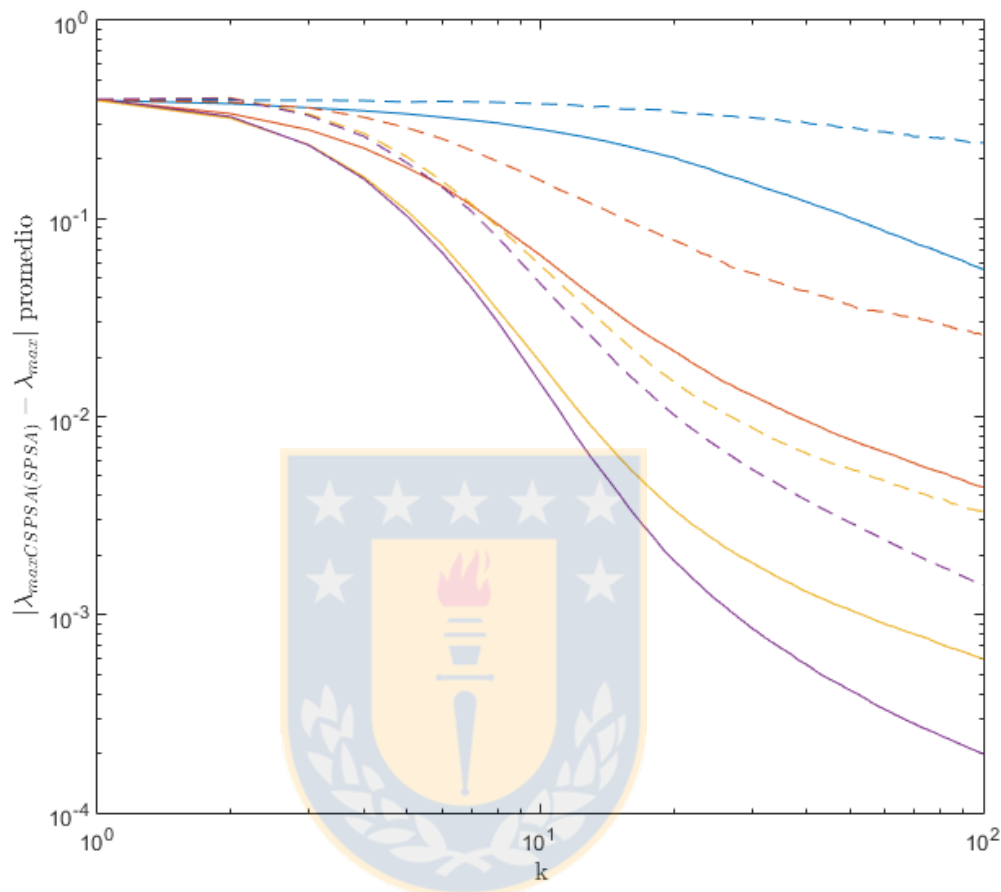


Figura 7.1: Gráfico de $|\lambda_{maxCSPA(SPSA)} - \lambda_{max}|$ vs Iteraciones para dos qubits. Las líneas continuas ilustran la diferencia entre el valor propio máximo teórico y el calculado mediante CSPA para un distinto número de experimentos por evaluación de la función, 10^1 , 10^2 , 10^3 y 10^4 desde arriba a abajo. Las líneas punteadas muestran lo mismo considerando el valor propio máximo calculado con SPSA. Fuente: Elaboración propia.

de matrices densidad reducidas en un sistema A de un sistema bipartito AB .

En 7.1 vemos para dimensión 2 (2 qubits) la diferencia entre los coeficientes de Schmidt teóricos y los calculados con CSPA y SPSA. La infidelidad está

promediada sobre un conjunto de 10^4 estados. Los estados semilla $|\phi\rangle_1$ con los que comenzamos la optimización son elegidos al azar según la distribución de Haar, lo que nos asegura una distribución 'uniforme' sobre las matrices densidad. Esto significa que, al promediar en el método de Montecarlo, no otorgamos mas importancia a ciertas zonas del espacio de Hilbert por sobre otras. La función infidelidad $I(\phi)$ es aproximada con 10, 100, 1000 y 10000 partículas idénticas por iteración.

A medida que aumenta el número de iteraciones, vemos en la figura 7.1 que la precisión de la estimación del autovalor más grande mejora, obteniendo resultados al menos un orden mejor con CSPSA. También vemos una mejora apreciable, de entre uno y dos órdenes de magnitud, al mejorar la estimación de la función objetivo I al aumentar el tamaño del ensemble al estimar esta.

En la figura 7.1 vemos lo mismo para mayores dimensiones, 2, 4, 8, 16 y 32. En esta simulación en todas las curvas tenemos 10^4 detecciones por evaluación de la función infidelidad, y el promedio es sobre 10^4 estados iniciales. Notamos que el error depende de la dimensión en la que trabajamos, siendo $|\lambda_{maxCSPSA} - \lambda_{max}|$ de peor calidad a medida que aumenta la dimensión.

Estos resultados coinciden con los de la Medida Geométrica de Entrelazamiento para 2 qudits, que está dada por

$$\lambda_{max} = \max_{\phi} |\langle \phi | \psi \rangle|. \quad (7.1)$$

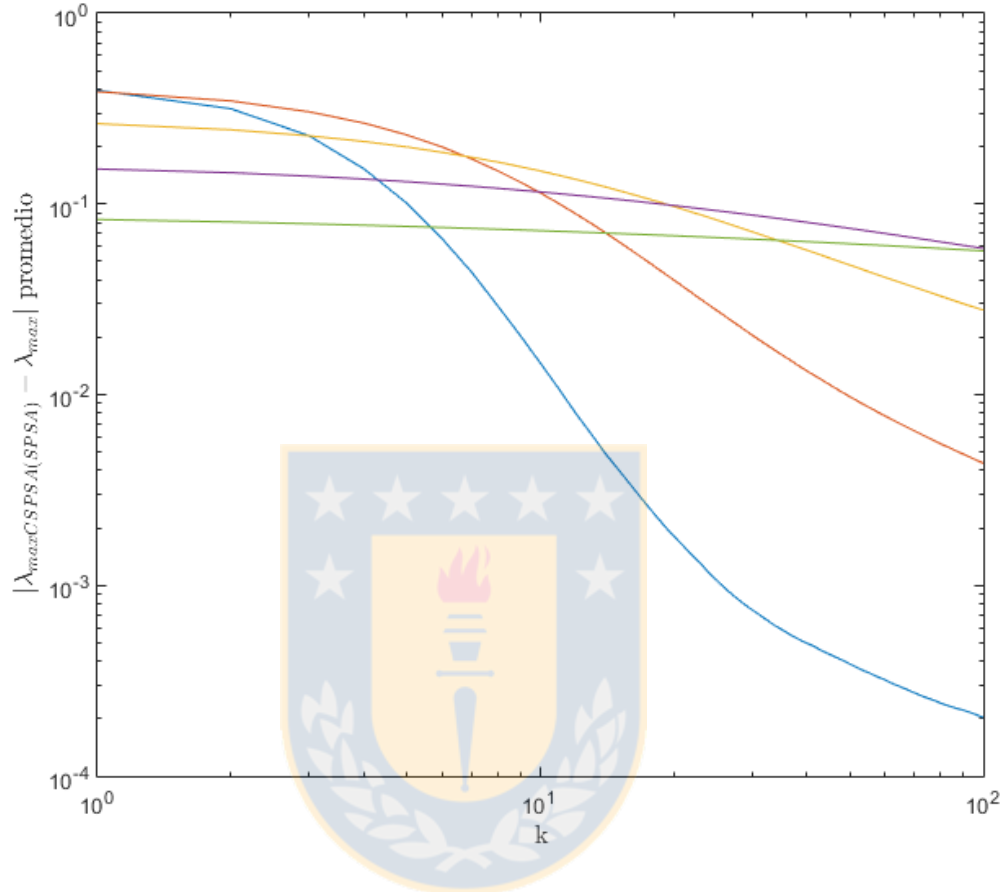


Figura 7.2: Gráfico de $|\lambda_{maxCSPSA(SP SA)} - \lambda_{max}|$ vs Iteraciones para dimensiones 32, 16, 8 , 4 y 2, estimando la infidelidad con 10^4 partículas. El error crece a medida que aumenta la dimensión. Fuente: Elaboración del autor.

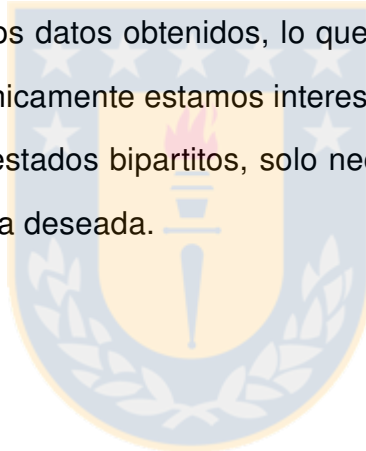
donde $\langle \phi |$ es un estado separable y $|\psi\rangle$ el estado del sistema. Esto significa que buscamos el estado separable más próximo al estado del sistema. Este es un problema de autovalores, que para sistemas bipartitos es lineal, y se reduce a encontrar la descomposición de Schmidt

$$|\psi\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle \quad (7.2)$$

del estado. Más aun, el autovalor de entrelazamiento λ_{max} es igual al coeficiente de Schmidt más grande. Así, encontrar el mayor autovalor de la matriz densi-

dad reducida para un estado bipartito es equivalente a encontrar la Medida Geométrica de Entrelazamiento, ya que usando CSPA el estado separable $|\phi\rangle$ converge al estado $|i_A\rangle|i_B\rangle$ asociado al mayor coeficiente de Schmidt, con el mérito de no necesitar conocer el estado del sistema. Por lo tanto, el método, además de ser un test de entrelazamiento nos entrega una medida de este.

La ventaja de este algoritmo con respecto a realizar tomografía para estimar el estado, y luego calcular el mayor autovalor, es que no necesitamos realizar postprocesamiento a los datos obtenidos, lo que ahorra tiempo y recursos computacionales. Así, si únicamente estamos interesados en obtener una medida de entrelazamiento para estados bipartitos, solo necesitamos iterar para obtener el resultado con la certeza deseada.



Capítulo 8

Conclusión

El desarrollo de nuevos métodos de optimización es hoy un área intensiva de investigación. Éstos han encontrado recientemente en la Mecánica Cuántica un área fértil de aplicación. El problema de la determinación de estados cuánticos desconocidos juega un rol fundamental en la Mecánica Cuántica, dada la necesidad de contar con métodos confiables para asignar estados a sistemas cuánticos. Además, con el advenimiento de la Teoría Cuántica de la Información, la estimación de estados cuánticos desconocidos cobra un interés práctico, puesto que la implementación de algoritmos cuánticos, los protocolos de comunicaciones cuánticas, y los dispositivos cuánticos requieren de la determinación precisa de estados cuánticos para un adecuado benchmarking.

En trabajos previos, la estimación de estados cuánticos desconocidos ha sido formulada como un problema de optimización, en la llamada tomografía autoguiada. Para el caso específico de estados puros desconocidos, estos pueden ser caracterizados como soluciones de la minimización de la infidelidad $1 - |\langle \psi | \phi \rangle|^2$, donde $|\psi\rangle$ es el estado del sistema y $|\phi\rangle$ un estimado de este. La infidelidad coincide en su límite infinitesimal con la distancia de Bures. No es un problema

de optimización trivial, puesto que la función a minimizar contiene parámetros desconocidos, por lo que su gradiente no puede ser evaluado. Esto impide el uso de los tradicionales métodos de descenso. No obstante, es posible recurrir a métodos de optimización estocástica, los cuales al solo requerir un estimado del gradiente si pueden ser empleados.

En esta tesis hemos presentado una modificación del método de optimización Complex Simultaneous Perturbation Stochastic Approximation (CSPSA), destinada a aumentar su tasa de convergencia. Este método, diseñado para funcionar dentro del campo de los complejos, requiere de dos evaluaciones de la función a optimizar por iteración. En el caso de la determinación de estados cuánticos desconocidos, la función objetivo es la infidelidad, la cual puede ser estimada para cada iteración por medio de proyectar el sistema sobre dos bases de medición distintas. Se obtienen así dos distribuciones de las cuales sólo dos probabilidades son empleadas. En el caso de un sistema de dimensión d , este hecho implica que $2d - 2$ probabilidades son descartadas por el algoritmo. Nuestra propuesta es simple: en lugar de descartar dichas probabilidades, podemos aprovechar la información que contienen aplicando en cada iteración k del algoritmo Estimación via Máxima Verosimilitud (MLE) sobre el estimado $|\phi\rangle_k$. Este proceso nos entrega un estado $|\phi'\rangle_k$ mas cercano a la solución del problema de optimización, es decir, del estado cuántico desconocido. Luego, repetimos el proceso aplicando Estimación via Máxima Verosimilitud de forma acumulativa, es decir, usando todo el conjunto de probabilidades obtenidas en las iteraciones previas.

Por medio de simulaciones numéricas de Monte Carlo mostramos que, cuando aplicamos el algoritmo modificado CSPA/MLM al problema de la determinación de estados cuánticos puros la tasa de convergencia es aumentada notablemente. Por ejemplo, en dimensión $d = 2$ y para un ensemble de tamaño $N = 2 \times 10^6$ obtenemos después de sólo 10 iteraciones una infidelidad, promediada sobre todo el espacio de Hilbert correspondiente, del orden 6.5×10^{-7} . Sin el uso de Estimación por Máxima Verosimilitud se obtiene una infidelidad del orden de 1.5×10^{-2} . Un algoritmo clásico alcanzaría con los mismo recursos una infidelidad del orden de 10^{-2} . La ventaja comparativa de CSPA con MLM se mantiene en dimensiones mayores. Además, CSPA con MLM conduce a histogramas altamente simétricos de modo que media y mediana no difieren significativamente.

También hemos analizado el método desde el punto de vista de las cotas óptimas relevantes para la estimación cuántica en el caso de $d = 2$. La cota de Gill-Massar describe el óptimo teórico alcanzable en la estimación de la infidelidad para un estado mixto, por medio de un estimador que no presente sesgo. Nuestro algoritmo entrega una infidelidad ligeramente menor que la cota. Esto se debe a que el algoritmo ha sido desarrollado específicamente para estados puros, los cuales están caracterizados por $2d - 2$ parámetros, lejos de los $d^2 - 1$ que caracterizan un estado mixto. En el caso de la cota de Gill-Massar para estados puros, CSPA con MLM entrega infidelidades en el mismo orden pero superiores. Este punto es interesante, puesto que aún es posible intentar menores infidelidades vía CSPA/MLM por medio de optimizar con respecto a las ganancias del algoritmo.

En este trabajo también mostramos otra aplicación del método de optimización CSPA, relacionada con el entrelazamiento cuántico. Esta es una de las propiedades más fascinantes de la Mecánica Cuántica. Es un fenómeno sin equivalente clásico, que involucra correlaciones entre las propiedades físicas observables de sistemas compuestos, aun cuando los objetos estén separados espacialmente. Esto puede ser visto como un recurso, y usado en procesos como teleportación cuántica y criptografía cuántica, entre otros.

Utilizando CSPA podemos obtener una medida de la cantidad de entrelazamiento para estados puros $|\psi\rangle$ en sistemas bipartitos AB , haciendo uso de la descomposición de Schmidt. Esta última nos permite escribir el estado cuántico como una suma de estados puros $|i_A\rangle|i_B\rangle$, ortogonales entre sí, multiplicados por respectivos coeficientes reales λ_i , llamados coeficientes de Schmidt. El valor del λ_i más grande nos entrega una medida en el entrelazamiento. Con CSPA podemos encontrar el valor de este parámetro maximizando sobre la matriz densidad reducida de cualquiera de los dos sistemas. En sí, el protocolo a seguir es el mismo que en tomografía autoguiada, pero ahora estamos estimando la infidelidad a partir de datos provenientes de un estado mixto, y buscando el estado puro más cercano a este. Es importante destacar que el método no requiere conocer el estado del sistema, lo que puede ahorrarnos recursos al no tener que realizar tomografía para conocer este.

Las simulaciones numéricas han mostrado la convergencia del método, al disminuir la distancia entre el coeficiente de Schmidt teórico y el calculado medi-

ante CSPSA. Así, con el suficiente numero de mediciones e iteraciones podemos disminuir el error hasta hacerlo arbitrariamente pequeño.



Bibliografía

- [1] A. I. Lvovsky and M. G. Raymert, *Rev. Mod. Phys.* **81** (2009). <https://doi.org/10.1103/RevModPhys.81.299>.
- [2] V. Giovannetti, S. Lloyd and L. Maccone, *Nature Photonics* **5** (2011). <https://doi.org/10.1038/nphoton.2011.35>.
- [3] T. D. Ladd, *Nature* **464** (2010). <https://doi.org/10.1038/nature08812>.
- [4] J. e. a. Matteo G.A. Paris, *Quantum State Estimation* (Springer-Verlag Berlin Heidelberg, 2004).
- [5] C. Ferrie, *Phys. Rev. Lett.* **113**, 190404 (2014). <http://dx.doi.org/10.1103/PhysRevLett.113.190404>.
- [6] A. Utreras, *Una mejora a la tomografía cuántica autoguiada.*, Tesis Doctoral, Universidad de Concepción (2017).
- [7] P. M. Wei, Tzu-Chieh; Goldbart, *Physical Review A* **68**, 042307 (2003). <http://dx.doi.org/10.1103/physreva.68.042307>.
- [8] S. Ross, *A First Course in Probability* (Prentice Hall PTR, 1997), quinta edición.
- [9] A. B. Owen, *Monte Carlo theory, methods and examples* (2013).

- [10] G. Strang, *Linear Algebra and Its Applications* (Brooks Cole, 2005), cuarta edición.
- [11] W. Greiner, *Quantum mechanics : an Introduction* (Springer, 2001), cuarta edición.
- [12] Ward Cheney and David R. Kincaid, *Linear Algebra: Theory and Applications* (Jones and Bartlett, 2012), segunda edición.
- [13] F. Mezzadri, How to generate random matrices from the classical compact groups, arXiv:math-ph/0609050 (2007). <https://arxiv.org/abs/math-ph/0609050>.
- [14] M. A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2010).
- [15] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Edizioni della Normale, 2011).
- [16] S. N. Filippov and V. I. Man'ko, *Physica Scripta* **2011**, T143 (2011). <http://dx.doi.org/10.1088/0031-8949/2011/T143/014010>.
- [17] R. L. B. George Casella, *Statistical Inference* (Duxbury Press, 2001).
- [18] A. C. Ian Goodfellow, Yoshua Bengio, *Deep Learning* (The MIT Press, 2016).
- [19] G. C. E.L. Lehmann, *Theory of Point Estimation* (Springer, 2003), segunda edición.
- [20] H. Zhu, *Quantum State Estimation and Symmetric Informationally Complete POMs.*, Tesis Doctoral, National University of Singapore (2012). <http://scholarbank.nus.edu.sg/handle/10635/35247>.

- [21] S. Gill, Richard D.; Massar, Physical Review A **61** (2000). <http://dx.doi.org/10.1103/PhysRevA.61.042312>.
- [22] J. C. Spall, IEEE Trans. Automat. Contr. **37**, 3 (1992). <http://dx.doi.org/10.1109/9.119632>.



Capítulo 9

Apéndice

```
1 function [SFID M] = FID(psi , psi1 , Nexp) %infidelidad simulada
2 dim = length(psi1);
3 M = rand(dim)*exp(i*2*pi*randn(dim));
4 M(:,1) = psi1;
5 M = GS(M);
6 prob = M'*kron(psi , psi')*M;
7 prob = diag(prob);
8 prob = prob/sum(prob);
9 SFID = 1- mnrnd(Nexp,prob)/Nexp; % distribucion multinomial
10 end
```

```
1 function [ LF,M3] = LF_Reall( t,n,vec) %funcion de verosimilitud
2 % vec – estados sobre los que se calculo infidelidad.
3 % n – valores de infidelidad asociados a vec.
4 % t – semilla
5 rho=itrans(t(:,1), t(:,2)); % transformacion de coordenadas.
6 rho = kron(rho,rho'); % variable en MLE
7 M= 1- vec'*rho*vec; % infidelidad teorica
8 M2=diag(M)';
9 M3=sqrt(0.5)*(M2-n) ./ sqrt(M2);
10 LF=norm(M3)^2; % funcion de verosimilitud
11 end
```



```

1 function [valor_optimo , rho] = CSPSASLD1(K,ESTADO, rho , dim , Nexp)
2 %CPSA con MLE para estados puros.
3 % K- numero de iteraciones.
4 % ESTADO- estado sobre el que realizamos tomografia.
5 % rho- semilla
6 % dim- dimension del estado
7 % Nexp- numero de particulas por medicion de infidelidad
8 s=0.606; t=1/6; a=3; A=0; b=0.02;
9 f= @(dens) (1 - abs(dot(dens, ESTADO))^2); %Infidelidad teorica
10 valor_optimo(1) = f(rho(:,1));
11 for k=0:K-2
12 alfa = a/(10*k+1+A)^s; %ganancias
13 beta = b/(k+1)^t; %ganancias
14 delta = (1i).^(randi(4,dim,1)); %vector delta
15 smas(:,k+1)= rho(:,k+1) + beta*delta;
16 smenos(:,k+1) = rho(:,k+1) - beta*delta;
17 smas(:,k+1) = smas(:,k+1)/norm(smas(:,k+1));
18 smenos(:,k+1) = smenos(:,k+1)/norm(smenos(:,k+1)); %vectores sobre
19 los que evaluamos infidelidad
20 [fmas(:, dim*k+1:dim*k+dim) sma(:, dim*k+1:dim*k+dim) ]= FID(ESTADO,
21 smas(:,k+1), Nexp ); % Infidelidad simulada
22 [fmenos(:, dim*k+1:dim*k+dim) sme(:, dim*k+1:dim*k+dim)]= FID(ESTADO,
23 smenos(:,k+1), Nexp); % Infidelidad simulada
24 % smas smenos - bases en las que se mide.
25 % fmas fmenos- infidelidad en las bases medidas.
26 gra = ( fmas(:, dim*k+1)- fmenos(:, dim*k+1))./(2*beta*conj(delta)); %
27 evaluacion gradiente.
28 rho(:,k+2) = (rho(:,k+1) - alfa*gra)/norm(rho(:,k+1) - alfa*gra); %
29 guess para MLE
30 % MLE
31 ss = [sma sme]; % array con estados.
32 ff = [fmas fmenos]; % array con mediciones de infidelidad.
33 [theta phi] = trans(rho(:,k+2)); % cambio a coordenadas esfericas.
34 t_re = [theta phi]; % guess MLE
35 options = optimset('TolFun',10e-8);
36 fun = @(x)LF_Reall(x, ff , ss);
37 t_matlab = fminsearch(fun,t_re , options); % minimizacion funcion de
38 verosimilitud
39 rho(:,k+2)=itrans(t_matlab(:,1) , t_matlab(:,2)); % vuelta a coordenadas
40 normales. Guess para la proxima iteracion
41 valor_optimo(k+2) = f(rho(:,k+2));
42 end

```