



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
Maestría en Ciencias (Física)

Máquinas de Boltzmann Restringidas para aproximar el estado base del modelo  $\mathbb{Z}_2$   
de Bose-Hubbard

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
MAESTRÍA EN CIENCIAS

PRESENTA:  
ROMAN JOSUE ARMENTA RICO

DR. ISAAC PEREZ CASTILLO  
DEPARTAMENTO DE FÍSICA, UAM-I

DRA. ROCIO JAUREGUI RENAUD  
IF-UNAM

DR. JOSE RUFINO DIAZ URIBE  
ICAT-UNAM

MÉXICO, CDMX. NOVIEMBRE 2022



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*Para mi mamá Ma. Trinidad, mi papá Josué y mi hermana Jennifer.  
Para Mitzuky.*

# Agradecimientos

Gracias a mi familia, a mi mamá Trinidad, mi papá Josué y mi hermana Jennifer, por siempre apoyarme y creer en mí. Todos mis logros son suyos.

Gracias a mis amigas y amigos, Mitzuky, Tatiana, Carlos, Hugo, Abel, América, Fernando, Humberto, Silvia, Daniel, Patricio, Rosario, Axel, Jaziel y Yari, por darme siempre ánimos y estar ahí cuando lo necesité. Les quiero mucho.

Gracias a mi compañero de investigación, Gustavo, por ayudarme siempre y motivarme a seguir adelante.

Gracias al Instituto de Física de la UNAM por permitirme terminar mi proyecto de tesis en la biblioteca, y gracias a Guille, Lucy y Roberto por siempre tratarme con cariño y atención.

Gracias a mi asesor de tesis, Isaac, por tenerme paciencia y enseñarme todos los días a ser curioso. "*ABC: Always Be Curious*".

Gracias a mis asesores Rocío Jáuregui y Rufino Díaz.

Gracias a Filippo Vicentini, Giuseppe Carleo, Damian Hofmann y todos los desarrolladores de NetKet 3 por ser tan atentos y resolver mis dudas siempre.

Gracias a CONACyT por financiar mi maestría.

# Índice general

<b>Agradecimientos</b>	<b>II</b>
<b>Introducción</b>	<b>VIII</b>
<b>1. Sistemas físicos de muchos cuerpos correlacionados</b>	<b>1</b>
1.1. Sistemas cuánticos de red . . . . .	2
1.1.1. Sistemas ópticos de red . . . . .	3
1.1.2. Dos ejemplos paradigmáticos: Modelo de Ising y Modelo de Bose– Hubbard . . . . .	4
1.2. Métodos de Aproximación . . . . .	9
1.2.1. Métodos Monte Carlo: algoritmo de Metrópolis–Hastings . . . . .	9
1.2.2. Método variacional . . . . .	11
1.2.3. Descenso por gradiente . . . . .	12
1.2.4. Método de Monte Carlo variacional . . . . .	15
1.2.5. Reconfiguración estocástica . . . . .	18
1.3. Ejemplos de aproximación . . . . .	19
1.3.1. Modelo de Ising con campo externo transversal . . . . .	19
1.3.2. Modelo de Bose–Hubbard con la función de onda de Jastrow . . . . .	21
<b>2. Redes neuronales y aprendizaje automático</b>	<b>25</b>
2.1. Estructura de una red neuronal . . . . .	25
2.2. Aprendizaje automático . . . . .	29
2.2.1. Aprendizaje con profesor . . . . .	30
2.2.2. Aprendizaje sin profesor . . . . .	31

2.3.	Aproximar el estado base de sistemas cuánticos de muchos cuerpos a través de redes neuronales . . . . .	32
2.3.1.	Funciones de onda variacionales con redes neuronales . . . . .	33
2.3.2.	Máquina de Boltzmann Restringida . . . . .	35
2.4.	Ejemplos de aproximación con estados cuánticos neuronales . . . . .	38
2.4.1.	Modelo de Ising transversal . . . . .	39
2.4.2.	Modelo de Bose–Hubbard . . . . .	41
<b>3.</b>	<b>Modelo <math>\mathbb{Z}_2</math> Bose–Hubbard</b>	<b>45</b>
3.1.	Dinámica del sistema . . . . .	46
3.1.1.	Límite $U \rightarrow \infty$ de bosones robustos . . . . .	48
3.2.	Implementación de máquinas de Boltzmann restringidas . . . . .	51
3.3.	Resultados . . . . .	52
3.3.1.	Optimización de la energía . . . . .	52
3.3.2.	Diagrama de fase en límite adiabático . . . . .	54
3.3.3.	Llenado fermiónico respecto al potencial químico . . . . .	55
	<b>Conclusiones</b>	<b>57</b>
	<b>Apéndices</b>	
<b>A.</b>	<b>Modelos de Red</b>	<b>59</b>
A.1.	Modelo de Bose–Hubbard . . . . .	59
A.1.1.	Teoría de campo medio . . . . .	59
<b>B.</b>	<b>Métodos de Aproximación Variacionales</b>	<b>63</b>
B.1.	Descenso al gradiente . . . . .	63
B.1.1.	Técnicas de proyección . . . . .	66
B.1.2.	Reconfiguración estocástica . . . . .	67

# Índice de figuras

1.1. Representación gráfica del modelo de Ising unidimensional. . . . .	5
1.2. Representación gráfica del modelo de Bose–Hubbard unidimensional. . . .	7
1.3. Espacio fase del modelo de Bose–Hubbard con aproximación de campo medio.	8
1.4. Optimización de la energía del estado variacional de Jastrow para el modelo de Ising con campo transversal con $N = 20$ sitios de espín para $J = 1$ y a) $g = 0.5$ , b) $g = 1.0$ y c) $g = 1.5$ . . . . .	21
1.5. Optimización de la energía (líneas color morado) y de la densidad de ocupación (líneas color naranja) para un algoritmo variacional de 4000 pasos de optimización con la función de onda de Jastrow. Se comparan con los resultados obtenidos por diagonalización de Lanczos del modelo de Bose–Hubbard con $N = 6$ sitios bosónicos en la cadena y $n_{max} = 5$ número máximo de bosones por sitio, con a) $U = 20.0$ y $\mu = 11.0$ y b) $U = 20.0$ y $\mu = 20.0$ . . . . .	22
1.6. Comparación de diagrama de fase del modelo de Bose–Hubbard esbozado por la función de onda variacional de Jastrow (puntos azules) respecto al obtenido mediante diagonalización por Lanczos (líneas negras). Para el método variacional, la cadena cuenta con $N = 8$ sitios bosónicos con a lo más $n_{max} = 5$ bosones por espín, comparando la densidad de ocupación bosónica calculada por el método Monte Carlo respecto a su entero más cercano con tolerancia de $\pm 0.05$ . . . . .	23
2.1. Representación gráfica de una neurona. . . . .	26
2.2. Estructura de red neuronal realimentada de capa simple. . . . .	27
2.3. Estructura de red neuronal multicapa de realimentación. . . . .	28

2.4. Esquema gráfico del aprendizaje supervisado por un profesor. . . . . 30

2.5. Esquema gráfico del algoritmo de aprendizaje reforzado. . . . . 31

2.6. Estructura de una máquina de Boltzmann restringida. . . . . 36

2.7. Representación gráfica de la codificación tipo One–Hot en la entrada de una  
RBM. . . . . 37

2.8. Optimización de la energía del estado variacional de Jastrow para el modelo  
de Ising con campo transversal con  $N = 20$  sitios de espín para  $J = 1$  y a)  
 $g = 0.5$ , b)  $g = 1.0$  y c)  $g = 1.5$ . . . . . 40

2.9. Comparación de la optimización de la energía (líneas color morado) y den-  
sidad de ocupación (líneas color naranja) por la función de onda variacional  
de Jastrow respecto a la optimización de la energía (líneas color celeste) y  
la densidad de ocupación (líneas color rojo) del estado cuántico neuronal  
de la RBM . Se comparan con los resultados obtenidos por diagonalización  
de Lanczos del modelo de Bose–Hubbard con  $N = 8$  sitios bosónicos en la  
cadena y  $n_{max} = 5$  número máximo de bosones por sitio, con a)  $U = 20.0$  y  
 $\mu = 11.0$  y b)  $U = 20.0$  y  $\mu = 20.0$ . . . . . 42

2.10. División de fases entre el aislante de Mott  $\rho \in \mathbb{N}$  y superfluido esbozada por  
la RBM (puntos azules), comparando la densidad bosónica calculada por la  
optimización y su parte entera, con una tolerancia de  $\pm 0.05$ . Se compara  
con los resultados obtenidos por diagonalización de Lanczos. El modelo de  
Bose–Hubbard tiene  $N = 8$  sitios bosónicos con a lo más  $n_{max} = 5$  bosones  
por sitio. . . . . 43

3.1. Representación gráfica del modelo  $\mathbb{Z}_2$  de Bose–Hubbard. . . . . 47

3.2. Representación gráfica del modelo de RBM para codificar los estados del  $\mathbb{Z}_2$   
Bose–Hubbard. . . . . 51

3.3. Optimización de energía del modelo  $\mathbb{Z}_2$  de Bose–Hubbard en el límite adiabático para a)  $\Delta/t = 0.3$  mostrando el ordenamiento ferromagnético en las ligas de espín en b) y el llenado medio fermiónico en c), y la optimización para d)  $\Delta/t = 0.5$ , con un ordenamiento de Néel en las ligas de espín en e) y el llenado medio fermiónico f). . . . . 53

3.4. Resultados de a) magnetización total y b) magnetización escalonada para una malla en el espacio  $(\alpha/t, \Delta/t)$  con  $\beta/t = 0$  en el límite de bosones robustos. c) y d) muestran una línea de nivel más definida para  $\alpha/t = 1.0$ . 55

3.5. Llenado medio de ocupación fermiónica respecto al potencial químico para  $\alpha/t = 0.5$ ,  $\Delta/t = 0.5$  en el límite adiabático con bosones robustos. . . . . 56

# Introducción

Uno de los objetivos de la física es describir el comportamiento del mundo real a través de modelos matemáticos simples. Para esto, se define un *sistema físico* como un conjunto de objetos físicos que se desean estudiar, y un *modelo físico* como la descripción matemática de dichos sistemas [1]. De esta forma, es posible modelar cómo se comporta una porción del mundo real, sólo considerando las interacciones y los objetos físicos de interés resolviendo dichos modelos. Al resto del universo se le conoce como *entorno*, y la influencia de éste sobre el sistema es elección del análisis que se desee realizar. A través de estos modelos físicos, es posible aproximar el comportamiento del mundo real de manera tan precisa como se desee.

Una vez se ha definido un modelo del sistema físico, el siguiente paso es resolver dicho modelo para poder obtener las soluciones y comprender su comportamiento. Para ello, es necesario una *teoría física*, que está compuesta por tres componentes: el *formalismo*, las *leyes dinámicas* y las *reglas de correspondencia*.

Se puede identificar al formalismo de una teoría física como al conjunto de símbolos, objetos o reglas de deducción que describen al sistema que se estudia. Existen dos objetos básicos en el formalismo clásico o cuántico; el *estado* que describe exactamente la configuración del sistema en ese instante, y los *observables*, es decir, las cantidades físicas que se pueden medir experimentalmente.

Las leyes dinámicas son las relaciones que deben cumplir los objetos básicos de la teoría física contenidos por el formalismo. En estas leyes deben estar descritas las leyes de interacción entre los cuerpos y cómo afecta el entorno al sistema físico completo, así como otras consideraciones que puedan afectar al estado del sistema.

Por último, las reglas de correspondencia son las que asignan el significado empírico a los

objetos que describe el formalismo, es decir, es el que da la interpretación física a la teoría.

En la mecánica clásica, por ejemplo, el estado en el formalismo canónico está descrito por el espacio fase del sistema [2]. En particular, el estado de una sola partícula en tres dimensiones está descrito por un vector de seis entradas,  $(q_1(t), q_2(t), q_3(t), p_1(t), p_2(t), p_3(t))$ , donde  $q_i(t)$  y  $p_i(t)$  son los observables que describen la posición y el momento de la partícula en el tiempo a lo largo de la dirección  $i$ , respectivamente. Las leyes dinámicas en este formalismo son las llamadas *ecuaciones de Hamilton*,

$$\frac{dq_i(t)}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i(t)}{dt} = -\frac{\partial H}{\partial q_i}, \quad i = 1, 2, 3. \quad (1)$$

donde  $H(q_i, p_i; t)$  es el Hamiltoniano del sistema. En la mecánica clásica, las reglas de correspondencia son bastante directas: la interpretación física de los observables puede ser medida por las propiedades de los objetos macroscópicos, como la posición y el momento.

Por otro lado, en la mecánica cuántica, el estado en el llamado formalismo de Schrödinger está descrito por una función  $\Psi(\mathbf{x}; t)$  llamada función de onda, que toma valores de un espacio Euclidiano complejo que contiene todos los posibles estados del sistema, llamado *espacio de Hilbert*  $\mathcal{H}$  [3]. Los observables en este formalismo son operadores auto-adjuntos definidos en el espacio de Hilbert. La ley dinámica que rige este formalismo es la ecuación de Schrödinger,

$$i\hbar \frac{\partial \Psi}{\partial t}(\mathbf{x}; t) = \hat{H}\Psi(\mathbf{x}; t), \quad (2)$$

donde  $\hat{H}$  es el operador Hamiltoniano, el observable que mide la energía total y describe las interacciones entre los cuerpos del sistema.

A través de estos formalismos es posible resolver de manera analítica una gran cantidad de modelos físicos conocidos, tales como el oscilador armónico  $n$ -dimensional, el modelo de Ising en una y dos dimensiones o el movimiento gravitatorio de dos cuerpos celestes para el caso de sistemas clásicos a través del formalismo canónico [2]. Para la mecánica cuántica, algunos modelos exactamente resolubles a través del formalismo de Schrödinger son el oscilador armónico cuántico, el modelo del átomo hidrogenoide o el péndulo cuántico [3]. Sin embargo, existen modelos más interesantes para sistemas con interacciones más

complejas cuyas soluciones analíticas aún no han sido encontradas, como el sistema de tres cuerpos [4] o el sistema de un átomo en presencia de un campo magnético externo, lo que genera el llamado efecto Zeeman [5].

Aún cuando no se ha encontrado una solución exacta para un modelo físico, es posible obtener una solución a través de *métodos de aproximación*. Ejemplos de estos para resolver sistemas cuánticos de muchos cuerpos son los métodos perturbativos [6] y los métodos variacionales [7]. De estos salen métodos más específicos, como el método de Born–Oppenheimer [8] para aproximar sistemas moleculares, o el método de Hartree–Fock [9], (también llamado método autoconsistente) para obtener la energía y la función de onda de sistemas cuánticos de muchos cuerpos en un estado estacionario.

Últimamente, la comunidad se ha volcado en el estudio de la implementación de modelos de funciones de onda variacionales basadas en redes neuronales [10] y utilizar el principio variacional para aproximar el estado base de sistemas cuánticos de muchos cuerpos [11].

La manera en la que opera una red neuronal se puede resumir como un operador funcional  $F_{\mathcal{W}}$  dependiente de un conjunto de parámetros  $\mathcal{W}$  que transforma un *vector de entrada*  $\vec{x} \in \mathbb{R}^n$  en un *vector de salida*  $\vec{y} \in \mathbb{R}^m$ . De esta forma, codificando información en la entrada y salida de una red neuronal, y después de ser entrenada a través de un esquema de aprendizaje automático adecuado, la red neuronal es capaz de obtener la respuesta deseada para cualquier vector de entrada. Los modelos neuronales han demostrado gran adaptabilidad de aprendizaje en muchas áreas: desde reconocimiento facial [12], clasificación de imágenes [13], reconocer y trasladar automáticamente cualquier texto de un lenguaje a otro [14], crear imágenes totalmente nuevas a partir de frases [15], e incluso se ha enseñado a redes neuronales a jugar juegos como ajedrez a nivel superhumano [16]. Sin embargo, aún hay mucho provecho desde el punto de vista de la investigación científica que se puede obtener a través de estos modelos, tales como poder modelar sistemas físicos complejos de muchos cuerpos de manera eficiente y que, sobre todo, implique una reducción de costo computacional considerable respecto a los demás modelos artificiales.

Otra de las razones por las cuales los modelos de redes neuronales son llamativos para aproximar modelos físicos es el hecho de que son *aproximadores universales* [17], esto es, que brindada la cantidad de neuronas interconectadas suficientes, se puede aproximar tanto

como se desee cualquier función continua diferenciable a través de una estructura de red neuronal. En el contexto de la física cuántica, el interés por estos modelos neuronales es por aproximar las funciones de onda de los modelos cuánticos para los cuales no es posible obtener una forma analítica. Por lo tanto, las redes neuronales son una herramienta para aproximar dichos sistemas y obtener información de ellos.

Para entender la implicación de la universalidad de los modelos de red neuronal, considere el caso donde se desea aproximar una *función vectorial objetivo*  $f(\cdot)$ , que cumple con la relación  $\vec{y} = f(\vec{x})$ , donde  $\vec{y}$  es el vector de salida cuando se aplica al vector de entrada  $\vec{x}$ . Ahora considere el conjunto de  $N$  puntos de relación  $\mathcal{F} = \{(\vec{x}_i, \vec{y}_i)\}_{i=1}^N$ , donde  $\vec{y}_i = f(\vec{x}_i)$  para todas las  $i$ . El hecho de que las redes neuronales sean aproximadores universales implica que sea posible diseñar una red neuronal que aproxime a la función  $f(\cdot)$  tanto como se desee; esto es, podemos encontrar una *función neuronal*  $F_{\mathcal{W}'}(\cdot)$  descrita por el conjunto de parámetros de pesos  $\mathcal{W}'$  tal que [18],

$$|F_{\mathcal{W}'}(\vec{x}) - f(\vec{x})| < \epsilon, \quad \forall \vec{x}, \quad (3)$$

para un parámetro de error  $\epsilon$  arbitrario positivo y todos los vectores de entrada  $\vec{x}$ . Si la red tiene las suficientes neuronas, entonces  $\epsilon$  puede ser tan pequeño como uno desee.

En este proyecto de tesis, se estudia la implementación de un modelo de red neuronal llamado *Máquinas de Boltzmann Restringidas* (RBMs por sus siglas en inglés) para modelar funciones de onda variacionales de sistemas cuánticos de muchos cuerpos. En específico, se estudia el modelo  $\mathbb{Z}_2$  de Bose–Hubbard para capturar su diagrama de fase a través de funciones de onda variacionales modeladas con redes neuronales. En el Capítulo 1 se desarrollan las técnicas de aproximación a utilizar, y en específico se pone a prueba el método variacional en dos sistemas cuánticos de muchos cuerpos paradigmáticos: el modelo de Ising con campo magnético transversal y el modelo de Bose–Hubbard a través de una función de onda de Jastrow. En el Capítulo 2 se introducen los modelos de redes neuronales. En específico, se presentan las Máquinas de Boltzmann Restringidas, y se compara su efectividad en aproximar el estado base de los modelos paradigmáticos resueltos por la función de onda de Jastrow. En el Capítulo 3 se introduce el modelo  $\mathbb{Z}_2$  de

Bose–Hubbard, se prueba la efectividad de las RBMs para aproximar su estado base y se comparan los resultados de su diagrama de fase respecto a la solución analítica en el límite adiabático [19]. Por último, en el Capítulo 4 se dan las conclusiones finales del desarrollo del proyecto de investigación.

# Capítulo 1

## Sistemas físicos de muchos cuerpos correlacionados

Al estudiar un sistema físico de uno o dos cuerpos que interactúan entre sí y con algún agente externo, puede que sea sencillo obtener una solución analítica exacta para los estados y los observables de dicho sistema si estas interacciones no son muy complejas. Sin embargo, cuando más de dos cuerpos se toman en cuenta, se vuelve mucho más complejo obtener una solución exacta, no solo porque el espacio de Hilbert crece exponencialmente con el número de grados de libertad total, sino también por el hecho de que la interacción conjunta y simultánea de varios objetos físicos provoca *correlaciones* [20]. A consecuencia de este efecto existen comportamientos complejos en sistemas físicos, tales como la superconductividad [21] y el magnetismo [22].

Cuando las interacciones de un sistema físico cambian, las correlaciones también lo hacen, y se pueden presentar transiciones de fase donde el sistema se comporta de manera distinta sólo modificando los parámetros de interacción. Predecir estas transiciones de fase puede ser muy difícil, dado que es necesario calcular observables que cambian de manera significativa con estos parámetros de interacción, es necesario utilizar métodos de aproximación para poder calcular estos cambios de fase.

En este trabajo de investigación se estudian diagramas de fase cuyas transiciones se dan por efectos puramente cuánticos, por lo que se toman en cuenta modelos físicos que están sujetos a dos simplificaciones: 1) la temperatura del sistema es  $T = 0$ , de tal forma

que las correlaciones sean puramente cuánticas, y 2) se considera una discretización del espacio, tal que las interacciones entre los sistemas se puedan especificar en unidades de distancia que hay entre las partículas. A este tipo de sistemas simplificados se les conoce como sistemas cuánticos de red.

En este capítulo se estudian los sistemas cuánticos de muchos cuerpos y se introducen métodos de aproximación conocidos para calcular sus propiedades. En la Sección 1.1 se presentan los sistemas cuánticos de red, y se estudian dos ejemplos de modelos cuánticos paradigmáticos: el modelo de Ising con campo magnético transversal y el modelo de Bose–Hubbard. Después, en la Sección 1.2 se presentan modelos de aproximación conocidos para calcular el estado base de sistemas cuánticos de muchos cuerpos, mismos que se utilizan en la Sección 1.1.2 para aproximar las soluciones del estado base de los sistemas antes mencionados.

## 1.1. Sistemas cuánticos de red

Un *sistema cuántico de red* es un modelo de objetos cuánticos definido en una malla, ya sea periódica (cristalina) o aleatoria, y las interacciones se definen respecto a los sitios de la red. Suponiendo un sistema cuántico de red con  $N$  sitios, se define la *base* como el conjunto de posibles configuraciones accesibles del sistema,

$$\mathcal{C} = \{|\mathcal{S}\rangle = |s_0\rangle \otimes \cdots \otimes |s_{N-1}\rangle \mid s_i \in \mathcal{L}_i, i \in \{0, 1, \dots, N-1\}\}, \quad (1.1)$$

donde  $|\mathcal{S}\rangle$  son las posibles configuraciones del sistema de red y  $\mathcal{L}_i$  es el conjunto de números cuánticos locales discretos que definen el estado del objeto cuántico del sitio  $i$ . La elección de la base depende entonces del tipo de objetos cuánticos que se estudian en el sistema.

Un estado genérico  $|\Psi\rangle$  del sistema cuántico de red es definido como una combinación lineal de la base elegida [1],

$$|\Psi\rangle = \sum_{\mathcal{S}} \psi(\mathcal{S}) |\mathcal{S}\rangle, \quad (1.2)$$

por lo que el espacio de Hilbert del sistema es generado por esta base.

### 1.1.1. Sistemas ópticos de red

Los modelos físicos que suponen sitios donde las partículas están confinadas en una red cristalina espacial pueden ser llevados a cabo en la vida real a través de los llamados *sistemas ópticos de red* [23]. Cuando un átomo es perturbado por un campo de luz láser dependiente de la posición  $\mathbf{x}$  y del tiempo  $t$  descrito por el campo eléctrico monocromático  $\mathbf{E}(\mathbf{x}, t)$ , este induce un momento dipolar  $\mathbf{p}$  por el efecto AC Stark [24] dado por,

$$\mathbf{p}(\mathbf{x}, t) = \hat{e}\tilde{p}(\mathbf{x})e^{-i\omega t} + \text{c.c.}, \quad (1.3)$$

donde  $\omega$  es la frecuencia angular de la onda eléctrica,  $\hat{e}$  es el vector de polarización,  $\tilde{p} = \alpha_p \tilde{E}$  es la amplitud del momento dipolar con  $\alpha_p$  la polarizabilidad compleja y  $\tilde{E}$  la amplitud del campo eléctrico. El potencial de interacción entre el momento inducido y el campo eléctrico es [25],

$$V(\mathbf{x}) = -\frac{1}{2} \langle \mathbf{p} \cdot \mathbf{E} \rangle = -\frac{1}{2\epsilon_0 c} \text{Re}(\alpha_p) I(\mathbf{x}), \quad (1.4)$$

donde  $I(\mathbf{x}) = 2\epsilon_0 c |\tilde{E}|^2$  es la intensidad del campo eléctrico. Para encontrar la polarizabilidad, se considera el átomo en el modelo de Lorentz, y considerando el límite semiclásico donde el átomo se puede aproximar a un sistema cuántico de dos niveles en un campo de luz clásico, la polarizabilidad resulta [25],

$$\alpha_p = \frac{e^2}{m_e} \frac{1}{\omega_0^2 - \omega^2 - i\omega\Gamma}, \quad (1.5)$$

donde  $\Gamma = (\omega_0^3/3\pi\epsilon_0\hbar c^3) |\langle e | \hat{\mu} | g \rangle|$  es la razón de decaimiento del estado excitado, con  $\hat{\mu}$  el operador de momento dipolar eléctrico [25],  $\omega_0$  la frecuencia asociada a la diferencia de energía  $\Delta = \hbar\omega_0$  entre el estado base  $|g\rangle$  y el estado excitado  $|e\rangle$  y  $\omega$  la frecuencia de la luz láser. De esta forma, el potencial de interacción es,

$$V(\mathbf{x}) = \frac{3\pi c^2}{2\omega_0^3} \frac{\Gamma}{\Delta} I(\mathbf{x}), \quad (1.6)$$

que depende linealmente de la intensidad del campo eléctrico externo. De esta forma, es posible crear un potencial de red óptico superponiendo dos haces de luz láser  $\vec{E}_{\pm}(\mathbf{x}, t) = E_0 \exp(\pm i k x)$  que se propagan en la dirección  $x$  con amplitud  $E_0$ , con número de onda  $k$  y longitud de onda  $2\pi/k$ , creando así un potencial  $V(x) \propto \cos^2(kx)$  con periodicidad  $a = \lambda/2$ . Usando otro par de haces contra-propagándose en la dirección  $y$  y  $z$ , se puede crear un potencial periódico de red tridimensional de la forma [25],

$$V(\vec{r}) = V_{0x} \cos^2(kx) + V_{0y} \cos^2(ky) + V_{0z} \cos^2(kz), \quad (1.7)$$

donde la profundidad de cada pozo de potencial es determinada por el par de haces de luz correspondientes.

### 1.1.2. Dos ejemplos paradigmáticos: Modelo de Ising y Modelo de Bose–Hubbard

A manera de ejemplo, se estudian dos modelos de red paradigmáticos para la mecánica cuántica: el modelo de Ising con campo transversal [26] y el modelo de Bose–Hubbard [27]. En ambos modelos se considera que la temperatura del sistema es cero,  $T = 0$ , y que los objetos cuánticos están confinados espacialmente en una red unidimensional cristalina.

#### Modelo de Ising con campo transversal

La dinámica del sistema cuántico de Ising con campo transversal está dada por el Hamiltoniano [28],

$$\hat{H}_I = -J \sum_{\langle i, j \rangle} \hat{\sigma}_i^z \hat{\sigma}_j^z - g \sum_{i=1}^N \hat{\sigma}_i^x, \quad (1.8)$$

donde  $J$  es la constante de intercambio entre los sitios de espín vecinos  $\langle i, j \rangle$  y  $g$  denota el valor del campo magnético externo en la dirección transversal  $x$ . En este ejemplo, se suponen condiciones periódicas de frontera en el sistema cuántico.

Cuando el campo externo es transversal a la dirección de los espines de la cadena en el

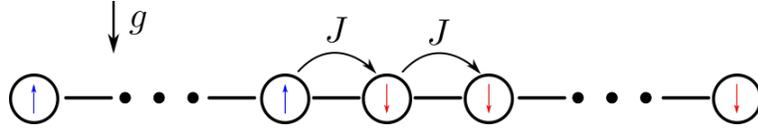


Figura 1.1: Representación gráfica del modelo de Ising unidimensional.

modelo de Ising, el problema deja de ser clásico y requiere un tratamiento cuántico dado que los operadores  $\hat{\sigma}_z$  y  $\hat{\sigma}_x$  no conmutan, por lo que no pueden medirse simultáneamente según el principio de incertidumbre de Heisenberg. Este modelo se puede resolver analíticamente aplicando una transformación de Jordan–Wigner [29].

### Modelo de Bose–Hubbard

El *modelo de Bose–Hubbard* describe la interacción entre bosones sin espín en una red  $d$ -dimensional [27], y su interacción ayuda a describir sistemas físicos tales como átomos bosónicos en una red óptica [30] y algunos aislantes magnéticos [31].

Se considera un gas de partículas que interactúan entre ellas en una red óptica, como las descritas en la Sección 1.1.1. La dinámica de dicho sistema está dada por el Hamiltoniano [32],

$$\hat{H}_{int} = \int d^3x \hat{\Psi}^\dagger(\mathbf{x}) \left( \frac{\hat{p}^2}{2m} + V_0(\mathbf{x}) \right) \hat{\Psi}(\mathbf{x}) + \frac{g}{2} \int d^3x \hat{\Psi}^\dagger(\mathbf{x}) \hat{\Psi}^\dagger(\mathbf{x}) \hat{\Psi}(\mathbf{x}) \hat{\Psi}(\mathbf{x}), \quad (1.9)$$

donde  $\hat{\Psi}(\mathbf{x})$  es el operador de campo que aniquila un bosón en la posición  $\mathbf{x}$ ,  $V_0(\mathbf{x})$  es el potencial de red óptico como en la Ec. (1.7) y  $g = 4\pi\hbar^2 a/m$  es la constante de acoplamiento de las colisiones de onda- $s$  descritas por una longitud de dispersión  $a$ . Para átomos individuales, el teorema de Bloch establece que las funciones propias del Hamiltoniano pueden tener la forma de funciones de onda de Bloch [33]. Entonces, usando la base de funciones de Wannier  $W_\nu(\mathbf{x} - \mathbf{x}_i)$ , donde  $\mathbf{x}_i$  denota la posición del sitio  $i$  de la red y  $\nu$  es el índice de banda [34], el operador de campo bosónico en segunda cuantización se escribe como [30],

$$\hat{\Psi}(\mathbf{x}) = \sum_{\nu,i} W_\nu(\mathbf{x} - \mathbf{x}_i) \hat{b}_{\nu,i}, \quad (1.10)$$

donde el espacio de Fock está en la base  $|n_{\nu,i}\rangle$ , con  $n_{\nu,i}$  el número de ocupación de bosones en la banda  $\nu$  del sitio  $i$  y  $\hat{b}_{\nu,i}$  es el operador que aniquila un bosón en el estado  $W_{\nu}(\mathbf{x} - \mathbf{x}_i)$ . Ahora, haciendo una aproximación de banda sencilla, es decir, considerando que todos los bosones del gas están ocupando el estado de banda más bajo, el Hamiltoniano resulta [30],

$$\hat{H} = - \sum_{i,j} J_{ij} \hat{b}_i^\dagger \hat{b}_j + \frac{1}{2} \sum_{i,j,k,l} U_{ijkl} \hat{b}_i^\dagger \hat{b}_j^\dagger \hat{b}_k \hat{b}_l, \quad (1.11)$$

donde se han definido los parámetros,

$$\begin{aligned} J_{ij} &= - \int d^3x W(\mathbf{x} - \mathbf{x}_i) \left( \frac{\hat{p}^2}{2m} + V_0(\mathbf{x}) \right) W(\mathbf{x} - \mathbf{x}_j), \\ U_{ijkl} &= g \int d^3x W(\mathbf{x} - \mathbf{x}_i) W(\mathbf{x} - \mathbf{x}_j) W(\mathbf{x} - \mathbf{x}_k) W(\mathbf{x} - \mathbf{x}_l), \end{aligned} \quad (1.12)$$

con  $J_{ij}$  los elementos de la *matriz de tunelaje* que fija la probabilidad de que un bosón en el sitio  $i$  pase al sitio  $j$  a través del efecto de tunelaje cuántico, y  $U_{ijkl}$  son los elementos de la *matriz de interacción*, y fijan la energía de interacción entre los bosones de los sitios  $i, j, k$  y  $l$ .

Ahora, considerando el límite de malla profunda, donde  $V_0 \gg E_R$ , donde  $E_R = k^2/2m$  es la llamada *energía de retroceso* [30], utilizando la aproximación de atadura fuerte (*tight binding approximation*), todos los elementos de la matriz de tunelaje más allá de primeros vecinos no son considerados, al igual que las interacciones de bosones en sitios distintos. Por lo tanto, definiendo  $U_{iiii} \equiv U$ ,  $J_{\langle i,j \rangle} = t$  con  $\langle i, j \rangle$  denota que  $i$  es primer vecino con  $j$ , y  $J_{ii} = \mu$ , el Hamiltoniano se transforma en [30],

$$\hat{H}_{BH} = -t \sum_{\langle i,j \rangle} \left( \hat{b}_i^\dagger \hat{b}_j + H.c. \right) + \frac{U}{2} \sum_i \hat{n}_i (\hat{n}_i - 1) - \mu \sum_i \hat{n}_i, \quad (1.13)$$

que es el Hamiltoniano del llamado *modelo de Bose-Hubbard*. El parámetro  $U$  fija la fuerza de interacción en el sitio cuando hay más de un bosón, que puede ser atractiva ( $U < 0$ ) o repulsiva ( $U > 0$ ) y  $\mu$  es el potencial químico que fija la energía necesaria para agregar o retirar bosones del modelo. Este sistema de red conserva una simetría global  $U(1)$ , dado que es invariante ante transformaciones del tipo  $\hat{b}_i \rightarrow e^{i\theta} \hat{b}_i$ , donde  $\theta$  es una fase global.

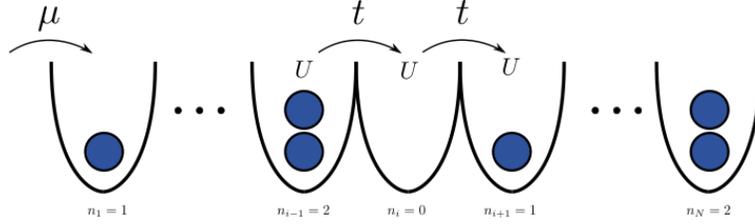


Figura 1.2: Representación gráfica del modelo de Bose–Hubbard unidimensional.

Aunque el modelo de Bose–Hubbard puede aparentar simplicidad, aún no se ha obtenido una solución exacta. Sin embargo, es posible aproximar la solución del estado base y su energía utilizando teoría de campo medio (ver Apéndice A.1), de donde se obtiene la información de las fases. Las fases del modelo de Bose–Hubbard son [35],

- Fase de superfluido (SF) cuando  $J \gg U$ . Esta fase es caracterizada por un número promedio de bosones por sitio no entero y compresibilidad distinta de cero.
- Fase de aislante de Mott (MI) cuando  $U \gg J$ . Se distingue por tener un promedio de bosones por sitio entero, además de ser incompresible.

Una visualización burda del diagrama de fase de este modelo se puede obtener aplicando teoría de campo medio al Hamiltoniano del sistema (ver Apéndice A.1.1). La idea principal de esta teoría es aproximar todas las interacciones de muchos cuerpos y reducirlas a interacciones de un solo cuerpo o *interacciones efectivas* [36].

Para el modelo de Bose–Hubbard, se supone que los observables  $\hat{b}_i$  y  $\hat{b}_i^\dagger$  oscilan alrededor de su valor promedio,  $\langle \hat{b}_i \rangle \equiv \phi \in \mathbb{C}$ , y  $\langle \hat{b}_i^\dagger \rangle \equiv \phi^* \in \mathbb{C}$ .

Utilizando la teoría de Landau para calcular las transiciones de fase a segundo orden (Ver Apéndice A.1), la transición de fase se da cuando,

$$a_2 = tz \left[ 1 + tz \left( \frac{n_0 + 1}{\mu - Un_0} + \frac{n_0}{U(n_0 - 1) - \mu} \right) \right] = 0, \quad (1.14)$$

donde  $\phi = 0$  corresponde a la fase de aislante de Mott, mientras que  $\phi \neq 0$  corresponde a la fase de superfluido.

En la Fig. (1.3) se muestra la transición de fase del modelo de Bose–Hubbard en el espacio de parámetros  $t/U, \mu/U$ , y se puede observar que para distintos números de ocupación promedio, se forman los llamados *módulos de Mott*, donde dentro de estas regiones

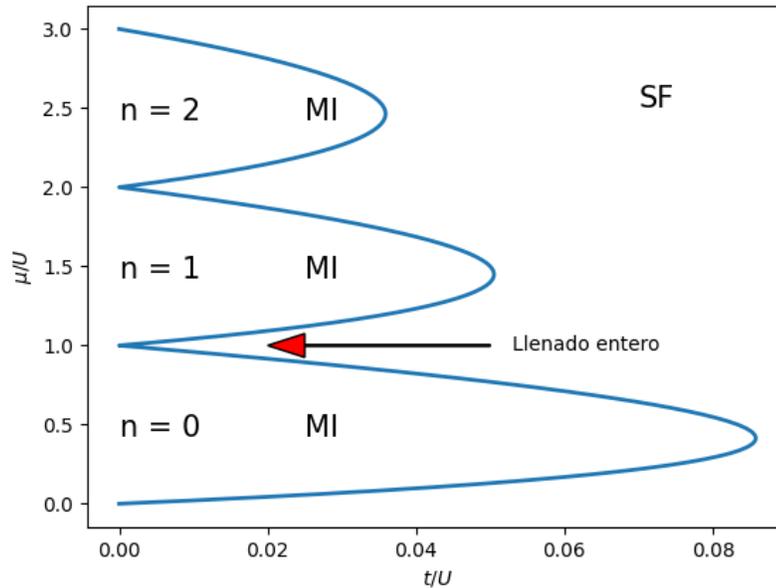


Figura 1.3: Espacio fase del modelo de Bose–Hubbard con aproximación de campo medio.

el modelo se comporta como un aislante de Mott. Además, esta aproximación permite ver que cuando  $\mu/U \in \mathbb{N}$  el modelo de Bose–Hubbard sólo se encuentra en fase de superfluido.

De esta aproximación teórica se puede obtener mucha información valiosa acerca del modelo de Bose–Hubbard:

- El modelo presenta dos fases: superfluido y aislante de Mott,
- las regiones del espacio de fase donde el modelo se comporta como un aislante de Mott está divididas por módulos que disminuyen en tamaño conforme aumenta la densidad de bosones promedio por sitio, que tiene que ser entera, y
- cuando  $\mu/U \in \mathbb{N}$ , el sistema se comporta como un superfluido.

Con esta aproximación fue posible darse una idea de la forma que tiene el diagrama de fase del modelo de Bose–Hubbard. Sin embargo, dado que la teoría de campo medio simplifica las correlaciones cuánticas del sistema, es necesario utilizar otros métodos de aproximación para obtener un diagrama de fase más preciso del modelo.

## 1.2. Métodos de Aproximación

Como se vio en la sección anterior, suponer en el modelo de Bose–Hubbard que el parámetro de orden  $\langle \hat{b} \rangle$  varíe alrededor de un valor promedio para obtener el diagrama de fase resta mucha validez a los cálculos. Sin embargo, tratar de resolver estos sistemas diagonalizando el Hamiltoniano a través de métodos numéricos exactos es un problema; como se ha reiterado, la dimensión del espacio de Hilbert escala exponencialmente con el número de partículas, por lo que llega a ser computacionalmente imposible utilizar métodos de diagonalización exacta en los sistemas más interesantes que constan de muchos cuerpos interactuantes [37]. La existencia de esta característica en los sistemas de muchos cuerpos nos obliga a buscar otros métodos para aproximar una solución de dichos modelos.

En esta sección se especificarán algunos de estos métodos para aproximar el estado base y los observables de sistemas físicos de muchos cuerpos, tales como los métodos estocásticos de Monte Carlo, el algoritmo de Metropolis–Hastings, y los basados en el principio variacional y las técnicas de proyección.

### 1.2.1. Métodos Monte Carlo: algoritmo de Metrópolis–Hastings

Suponga un sistema físico cuya distribución de probabilidad sobre los estados accesibles es  $p(x)$ , donde  $\{x\}$  es la base del sistema, con  $x$  las posibles configuraciones que puede tomar. Para calcular observables de interés, es necesario tomar muestras de la distribución de probabilidad de configuraciones.

Una solución ampliamente usada para tomar muestras de una distribución de probabilidad son los métodos Monte Carlo; procesos estocásticos que toman muestras aleatoriamente, pero siguiendo la distribución de probabilidad deseada. En particular, en este proyecto se abordará el método proporcionado por el *algoritmo de Metropolis–Hastings*, que tiene por propósito generar una secuencia de posibles configuraciones acorde con la distribución de probabilidad del estado del sistema  $p(x)$  [37]. A esta secuencia se le conoce como *cadena de Markov*, y a cada estado muestreado se le conoce como un *eslabón*. Se define el *kernel* de la cadena de Markov  $Q(x' | x)$  como la probabilidad de que una nueva configuración  $x'$  sea propuesta dada una configuración  $x$  aceptada anteriormente.

El algoritmo de Metrópolis–Hastings consiste en,

- **Inicialización:** se inicia el algoritmo con una configuración inicial aleatoria  $x_0$ , y se coloca en el primer eslabón de la cadena de Markov.
- **Iteración:** para un tiempo algorítmico  $t$ ,
  - Se propone una configuración  $y$  del *kernel* de transición  $Q(y | x_{t-1})$ , donde  $x_{t-1}$  es la configuración aceptada en el eslabón  $t - 1$  de la cadena de Markov.
  - Calcular la *probabilidad de aceptación*,

$$A(x_{t-1} \rightarrow y) = \min \left( 1, \frac{p(y)Q(x_{t-1} | y)}{p(x_{t-1})Q(y | x_{t-1})} \right). \quad (1.15)$$

- Con probabilidad  $A$ , aceptar la configuración propuesta  $y$  como el eslabón  $x_t$ . De no ser el caso, mantener  $x_{t-1}$  como la configuración del eslabón  $x_t$ .

Dado que el algoritmo de Metropolis–Hastings empieza con una configuración totalmente aleatoria de la base del sistema, las primeras configuraciones de la cadena de Markov no son representativas de la distribución de probabilidad de la que se están tomando las muestras. Se define la *cola* de la cadena de Markov como el conjunto de configuraciones que sí son representativas de la distribución de probabilidad.

Aún cuando el algoritmo de Metropolis–Hastings es muy útil para sistemas clásicos en ensambles con distribuciones de probabilidad conocidas, para sistemas cuánticos donde la distribución de probabilidad depende de la función de onda del sistema  $|\Psi\rangle$  es necesario primero calcular esta distribución de probabilidad y después utilizar el método de Metrópolis–Hastings para calcular los observables. Sin embargo, conocer completamente la función de onda de un sistema cuántico de muchos cuerpos puede llegar a ser imposible dada la cantidad de correlaciones cuánticas e interacciones que se deben de tomar en cuenta, por lo que se requiere acudir a métodos de aproximación para calcularla. En este proyecto se propone un método variacional para calcular dicha distribución de probabilidad.

### 1.2.2. Método variacional

El principio variacional se sostiene en el hecho de que *el estado base de cualquier sistema cuántico es aquel que tiene la energía mínima global* [7]. Con eso en cuenta, la idea principal es proponer una función de onda "de prueba"  $|\Psi_\alpha\rangle$  que dependa de un conjunto de *parámetros variacionales*  $\alpha = \{\alpha_1, \dots, \alpha_p\}$ , y encontrar el conjunto de parámetros óptimo  $\alpha'$  tal que la función de onda de prueba, llamada *función de onda variacional*  $|\Psi_{\alpha'}\rangle$  tenga la energía mínima en el espacio de parámetros.

La energía asociada a un estado variacional  $|\Psi_\alpha\rangle$  es llamada *energía variacional* y está definida por,

$$E_\alpha = \frac{\langle \Psi_\alpha | \hat{H} | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle}, \quad (1.16)$$

donde  $\hat{H}$  es el Hamiltoniano del sistema cuántico de muchos cuerpos.

La forma en la que se construye esta función de onda variacional es directa: de la teoría de espacios de Hilbert, el conjunto de vectores propios de cualquier operador Hermitiano en un espacio de Hilbert es una base para ese espacio de Hilbert [1]. Por lo tanto, si  $\{|x_i\rangle\}$  es el conjunto de vectores propios del operador Hermitiano  $\hat{O}$  definido en el espacio de Hilbert, es posible representar cualquier estado accesible del sistema cuántico como una combinación lineal de la forma,

$$|\Psi_\alpha\rangle = \sum_i \psi_\alpha(x_i) |x_i\rangle. \quad (1.17)$$

donde  $\psi_\alpha(x_i) = \langle x_i | \Psi_\alpha \rangle$  son los coeficientes que dependen de los parámetros variacionales, y,  $p_\alpha(x_i) = |\psi_\alpha(x_i)|^2$  denota la amplitud de probabilidad de que el sistema cuántico se encuentre en el estado  $|x_i\rangle$ . Ahora, dado que el operador Hamiltoniano  $\hat{H}$  del sistema cuántico también es Hermitiano, el conjunto de estados propios  $\{|\varphi_j\rangle\}$  del Hamiltoniano conforma una base para el espacio de Hilbert [1]. Entonces, existen coeficientes  $\varphi_j(x_i)$  tales que se puedan expresar los vectores propios de  $\hat{O}$  como una combinación lineal de la forma  $|x_i\rangle = \sum_j \varphi_j(x_i) |\varphi_j\rangle$ , y la función de onda variacional en la base del operador

Hamiltoniano resulta,

$$|\Psi_\alpha\rangle = \sum_{i,j} \psi_\alpha(x_i) \varphi_j(x_i) |\varphi_j\rangle. \quad (1.18)$$

Por lo tanto, si  $E_j$  es el valor propio del operador Hamiltoniano  $\hat{H}$  respecto a su vector propio  $|\varphi_j\rangle$  (es decir,  $E_j$  es la energía del estado  $|\varphi_j\rangle$ ), la diferencia entre la energía variacional definida en la Ec. (1.16) y la energía del estado base está dada por,

$$\epsilon \equiv E_\alpha - E_0 = \frac{\sum_{i,j} |\psi_\alpha(x_i)|^2 |\varphi_j(x_i)|^2 (E_i - E_0)}{\sum_{i,j} |\psi_\alpha(x_i)|^2 |\varphi_j(x_i)|^2}, \quad (1.19)$$

y dado que  $E_i - E_0 \geq 0$ , esto asegura que  $E_0$  es una cota inferior para la energía variacional  $E_\alpha$ .

Existen muchos métodos y técnicas para minimizar la energía variacional del sistema, pero en este proyecto se abordará el método del *descenso por gradiente* y la *reconfiguración estocástica*.

### 1.2.3. Descenso por gradiente

El descenso por gradiente es un algoritmo de optimización de primer orden diferenciable, y es utilizado para encontrar mínimos locales de funciones diferenciables multi-variables [38].

La idea principal detrás de este algoritmo está basado en el hecho de que cualquier función diferencial multi-variable  $F$  evaluada en algún punto de su espacio de dominio  $\vec{x}$  en el cual esta función es diferenciable, desciende más rápido a lo largo del camino a donde apunta la dirección contraria al gradiente en ese punto  $-\nabla F(\vec{x})$ .

Ahora, en el contexto de una función de onda variacional, de la Ec. (1.17), si elegimos la base como el conjunto  $\{|x_i\rangle\}$  de configuraciones posibles del sistema, entonces la función de onda variacional está dada por la Ec. (1.17), y la probabilidad  $p_\alpha(x_i)$  de encontrar al sistema en el estado  $|x_i\rangle$  está dada por,

$$p_\alpha(x_i) = \frac{|\langle x_i | \Psi_\alpha \rangle|^2}{\sum_j |\langle x_j | \Psi_\alpha \rangle|^2} = \frac{|\psi_\alpha(x_i)|^2}{\langle \Psi_\alpha | \Psi_\alpha \rangle}, \quad (1.20)$$

donde  $\psi_\alpha(x_i) = \langle x_i | \Psi_\alpha \rangle$  son los coeficientes de la expansión sobre la base  $\{|x_i\rangle\}$  y  $\sum_j |\langle x_j | \Psi_\alpha \rangle|^2 = \langle \Psi_\alpha | \Psi_\alpha \rangle$  es la norma de la función de onda variacional.

Sea  $\hat{O}$  un observable de interés del sistema. Entonces, el valor esperado está dado por la ecuación,

$$\langle \hat{O} \rangle = \frac{\langle \Psi_\alpha | \hat{O} | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle} = \sum_i p_\alpha(x_i) \tilde{O}(x_i) = \mathbb{E}[\tilde{O}], \quad (1.21)$$

donde  $\tilde{O}(x_i)$  es el *estimador local* del observable  $\hat{O}$  (ver Apéndice B.1), y está dado por,

$$\tilde{O}(x_i) = \frac{\langle x_i | \hat{O} | \Psi_\alpha \rangle}{\langle x_i | \Psi_\alpha \rangle} = \sum_j \frac{\langle x_j | \Psi_\alpha \rangle}{\langle x_i | \Psi_\alpha \rangle} \langle x_i | \hat{O} | x_j \rangle. \quad (1.22)$$

Cuando el operador  $\hat{O} = \hat{H}$  es el Hamiltoniano,  $\tilde{H}(x_i) = E_{\text{loc}}(x_i)$  es la *energía local* asociada a la configuración específica  $|x_i\rangle$ .

Ahora, se desea minimizar el valor esperado de la energía  $\langle \hat{H} \rangle$  respecto al espacio de parámetros variacionales. Una forma de hacer esto es a través del descenso por gradiente en el espacio de parámetros, aplicando la *regla de optimización*,

$$\alpha \rightarrow \alpha - \eta \nabla_\alpha \langle \hat{H} \rangle, \quad (1.23)$$

donde  $\nabla_\alpha$  denota el gradiente en el espacio de parámetros, y  $\eta$  es la *razón de aprendizaje*; entre más grande sea  $\eta$ , más rápido será la convergencia al mínimo local; sin embargo, habrá más error estadístico, y viceversa.

La derivada parcial del observable  $\langle \hat{H} \rangle$  se puede calcular usando la Ec. (1.17) y la Ec. (1.20), obteniendo así, (ver Apéndice B.1),

$$\frac{\partial}{\partial \alpha_k} \langle \hat{H} \rangle = \mathbb{E}[(\tilde{H} - \mathbb{E}[\tilde{H}])^* \mathcal{O}_k], \quad (1.24)$$

donde se ha definido a la *derivada logarítmica* sobre el parámetro variacional  $\alpha_k$  de la

función de onda variacional como,

$$\mathcal{O}_k(x) \equiv \frac{\partial}{\partial \alpha_k} \ln \psi_\alpha(x), \quad (1.25)$$

respecto a la configuración específica del sistema  $x$ .

Entonces, resumiendo, aproximar la función de onda variacional al estado base de un sistema de red cuántico de muchos cuerpos, así como su energía mínima a través de un esquema de principio variacional utilizando el descenso por gradiente puede hacerse siguiendo los siguientes pasos:

- **Inicialización:** Proponer una distribución de probabilidad variacional de la base del sistema  $\psi_\alpha(x) = \langle x | \Psi_\alpha \rangle$  que sea diferenciable respecto al espacio de parámetros  $\alpha$ , tal que la función de onda variacional sea como en la Ec. (1.17),

$$|\Psi_\alpha\rangle = \sum_i \psi_\alpha(x_i) |x_i\rangle,$$

donde  $\{x\}$  es la base del sistema. Se inicia el algoritmo con un conjunto de parámetros iniciales  $\alpha_1$ .

- **Iteración:** Para un tiempo algorítmico  $t$ , siendo  $\alpha_t$  el conjunto de parámetros al tiempo  $t$ ,
  1. Calcular la energía variacional  $E_{\alpha_t}$  del sistema dada por,

$$E_{\alpha_t} = \frac{\langle \Psi_{\alpha_t} | \hat{H} | \Psi_{\alpha_t} \rangle}{\langle \Psi_{\alpha_t} | \Psi_{\alpha_t} \rangle}, \quad (1.26)$$

a través de la Ec. (1.21).

2. Calcular las derivadas parciales de la energía variacional con respecto a todos los parámetros variacionales como en la Ec. (1.24),

$$\frac{\partial}{\partial \alpha_{t,k}} E_{\alpha_t} = \mathbb{E}[(E_{\text{loc}} - \mathbb{E}[E_{\text{loc}}])^* \mathcal{O}_k],$$

donde  $\alpha_{t,k}$  es el  $k$ -ésimo parámetro variacional en el paso  $t$  de optimización,

y  $E_{\text{loc}}(x) = \tilde{H}(x)$  es la energía local del sistema asociada a la configuración específica  $|x\rangle$ .

3. Actualizar todos los parámetros variacionales a través de un esquema del descenso por gradiente como en la Ec. (1.23),

$$\alpha_{t+1,k} = \alpha_{t,k} - \eta \frac{\partial}{\partial \alpha_{t,k}} E_{\alpha_t}$$

4. Repetir el proceso de optimización hasta que se alcance un objetivo de precisión.

Las Ecs. (1.20), (1.21) y (1.22) muestran que, en cada paso de optimización, es necesario sumar sobre todos los estados del espacio de configuraciones del sistema. Como se ha reiterado antes, el espacio de Hilbert crece exponencialmente con el número de partículas, por lo que para sistemas cuánticos de interés con una gran cantidad de partículas interactuando, llega a ser muy costoso guardar todas las configuraciones accesibles del sistema cuántico. Para solucionar este problema, se propone utilizar un esquema de Monte Carlo para aproximar las Ecs. (1.20), (1.21) y (1.22) en cada paso de optimización sin necesidad de considerar todas las configuraciones posibles.

#### 1.2.4. Método de Monte Carlo variacional

Suponga que se desea tomar un número  $N_s$  de muestras de configuraciones posibles de un sistema cuántico de muchos cuerpos a través de su distribución de probabilidad variacional. El conjunto de configuración  $\{|x_i\rangle\}_{i=1}^{N_s}$  es entonces tomado acorde a una distribución de probabilidad  $p_\alpha(x) \propto |\psi_\alpha(x)|^2$  dada por la Ec. (1.20). Entonces, el valor promedio de la Ec. (1.21) está dado por,

$$\langle \hat{O} \rangle = \mathbb{E}[\tilde{O}] \approx \frac{1}{N_s} \sum_{i=1}^{N_s} \tilde{O}(x_i), \quad (1.27)$$

donde  $\hat{O}$  es un observable de interés del sistema cuántico.

En el algoritmo de muestreo, se genera una cadena de Markov de configuraciones  $|x_1\rangle \rightarrow |x_2\rangle \rightarrow \dots \rightarrow |x_{N_s}\rangle$ , en la cual, en cada paso  $i$  de muestreo se guarda el módulo al

cuadrado de la función de onda variacional  $|\psi_\alpha(x_i)|^2$  para un conjunto de parámetros  $\alpha$ . Este muestreo puede llevarse a cabo a través de un *algoritmo de Metropolis–Hastings* simple, mostrado en la Sección 1.2.1, en el cual, en cada paso de la cadena de Markov, el número cuántico de un sitio aleatorio de la red es modificado (un espín se voltea, un número de ocupación bosónico cambia, etc.), y la nueva configuración es aceptada acorde a la *probabilidad de aceptación*,

$$A(|x_k\rangle \rightarrow |x_{k+1}\rangle) = \min\left(1, \left|\frac{\psi_\alpha(x_{k+1})}{\psi_\alpha(x_k)}\right|^2\right). \quad (1.28)$$

Una vez obtenida todas las muestras de configuraciones, se puede empezar el proceso de optimización a través de la Ec. (1.27). En el proceso de muestreo, la primera configuración tiene que ser tomada de manera completamente aleatoria del conjunto de todas las configuraciones posibles del sistema; sin embargo, esta configuración no es representativa de la distribución de probabilidad  $p_\alpha(x)$ . De hecho, dado que el siguiente eslabón de la cadena de Markov depende de la elección de la primera configuración, habrá un número  $N_c$  de muestras, al que se le llama *cola de la cadena de Markov*, que no son representativas de la distribución de probabilidad. En este sentido, es necesario descartar ese número de muestras de la cadena de Markov. Una vez que las muestras empiezan a ser representativas de la distribución de probabilidad, se dice que la cadena ha llegado al punto de *termalización*, y a partir de ahí, se pueden tomar las muestras para hacer los cálculos variacionales.

Como se vio, es posible obtener una función de onda variacional sin la necesidad de considerar todas las configuraciones posibles del sistema, y sólo calcular los observables con las muestras que se encuentran en la cola de la cadena de Markov y con la Ec. (1.27), para así minimizar la energía a través de un algoritmo extendido del principio variacional:

- **Inicialización:** Proponer una distribución de probabilidad variacional de la base del sistema  $\psi_\alpha(x) = \langle x | \Psi_\alpha \rangle$  que sea diferenciable respecto al espacio de parámetros  $\alpha$ , tal que la función de onda variacional sea como en la Ec. (1.17),

$$|\Psi_\alpha\rangle = \sum_i \psi_\alpha(x_i) |x_i\rangle,$$

donde  $\{x\}$  es el conjunto variacional del sistema. Inicializar el algoritmo de optimización con un conjunto de parámetros iniciales  $\alpha_1$ .

- **Iteración:** Para un tiempo algorítmico  $t$ , siendo  $\alpha_t$  el conjunto de parámetros al tiempo  $t$ ,

1. Utilizar el algoritmo de Metropolis–Hastings para tomar  $N_s$  muestras de configuraciones posibles de la distribución de probabilidad  $p_{\alpha_t}(x_i) = |\psi_{\alpha_t}(x_i)|^2$ , con  $\{|x_i\rangle\}_{i=1}^{N_s}$  la base del sistema al paso  $t$ .
2. Definir la función de onda variacional de Monte Carlo para el paso  $t$  como,

$$|\Psi_{\alpha_t}\rangle = \sum_{i=1}^{N_s} \psi_{\alpha_t}(x_i) |x_i\rangle, \quad (1.29)$$

donde se suma sobre los elementos de la base del sistema en el paso  $t$ .

3. Calcular la energía variacional  $E_{\alpha_t}$  del sistema dada por,

$$E_{\alpha_t} = \langle \hat{H} \rangle \approx \frac{1}{N_s} \sum_{i=1}^{N_s} \tilde{H}(x_i), \quad (1.30)$$

a través de la Ec. (1.27).

4. Calcular las derivadas parciales de la energía variacional con respecto a todos los parámetros variacionales como en la Ec. (1.24),

$$\frac{\partial}{\partial \alpha_{t,k}} E_{\alpha_t} = \mathbb{E}[(E_{\text{loc}} - \mathbb{E}[E_{\text{loc}}])^* \mathcal{O}_k],$$

donde  $\alpha_{t,k}$  es el  $k$ -ésimo parámetro variacional en el paso  $t$  de optimización, y  $E_{\text{loc}}(x_i) = \tilde{H}(x_i)$  es la energía local del sistema asociada a la configuración específica  $|x_i\rangle$ .

5. Actualizar todos los parámetros variacionales a través de un esquema del descenso por gradiente como en la Ec. (1.23),

$$\alpha_{t+1,k} = \alpha_{t,k} - \eta \frac{\partial}{\partial \alpha_{t,k}} E_{\alpha_t}$$

6. Repetir el proceso de optimización hasta que se alcance un objetivo de precisión.

La rapidez con la que la energía variacional converge a la energía del estado base depende totalmente del número de muestras que se toma, así como de la eficiencia con la que el algoritmo de muestreo Monte Carlo empieza a ser representativo de la distribución de probabilidad que quiere modelar. La mayoría de las veces, es cuestión de prueba y error modificar los parámetros como el número de muestras, la razón de aprendizaje, el *kernel* de transición del algoritmo de muestreo, entre otros, para que se obtenga la precisión deseada en los cálculos en un tiempo de cómputo razonable. Sin embargo, existen más técnicas que se pueden utilizar a la par con el descenso por gradiente para acelerar esta convergencia, como las técnicas de proyección y en especial, la reconfiguración estocástica.

### 1.2.5. Reconfiguración estocástica

La reconfiguración estocástica [39] es un algoritmo variacional que fue desarrollado inicialmente para resolver de manera parcial el famoso "problema del signo" que aparece al tratar de solucionar ciertos problemas mediante el método de Monte Carlo de la función de Green en sistemas cuánticos de red [40]. Una de las ventajas de este método de aproximación es el hecho de que requiere más información acerca de la función de prueba de la que se necesita en el método de descenso por gradiente simple, permitiendo así una optimización más rápida de la función de onda de muchos cuerpos.

Este algoritmo es un método de proyección, en el sentido de que la idea principal es modificar los parámetros de una función de onda de prueba de tal forma que se aproxime al estado base a lo largo de un camino en el espacio de los parámetros variacionales dictado por la proyección  $(\Lambda - \hat{H})$ , donde  $\Lambda$  se elige de tal manera que  $\Lambda - \hat{H} \geq 0$ .

El algoritmo de la reconfiguración estocástica consiste en optimizar los parámetros variacionales de la forma (ver Apéndice B.1.2),

$$\alpha_i \rightarrow \alpha_i - \Delta t \sum_j S_{i,j}^{-1} f_j, \quad (1.31)$$

donde  $\Delta t$  es la razón de aprendizaje,  $f_j$  son los elementos del vector de *fuerzas generalizadas*

zadas  $\vec{f}$  y  $S_{i,j}^{-1}$  son los elementos de la pseudo-inversa de Moore–Penrose de la *matriz de covarianza*  $\mathbf{S}$  definidos por,

$$S_{i,j} = \langle \hat{\mathcal{O}}_i^\dagger \hat{\mathcal{O}}_j \rangle - \langle \hat{\mathcal{O}}_j \hat{\mathcal{O}}_i^\dagger \rangle, \quad (1.32)$$

$$f_i = \langle \hat{\mathcal{O}}_i^\dagger \hat{H} \rangle - \langle \hat{H} \rangle \langle \hat{\mathcal{O}}_i^\dagger \rangle, \quad (1.33)$$

con  $\mathcal{O}_i$  el operador de la derivada logarítmica de la función de onda variacional respecto al parámetro  $\alpha_i$ , definida en la Ec. (1.24).

### 1.3. Ejemplos de aproximación

Una vez presentados los métodos de aproximación a utilizar, en esta sección se hará uso de ellos a manera de ejemplo para acercar el estado base de los sistemas paradigmáticos expuestos en la sección 1.1.2. Primero se hará uso de un método de Monte Carlo muestreando configuraciones a través de una distribución de Maxwell–Boltzmann de una red bidimensional de sistemas de espín, el llamado modelo de Ising clásico en dos dimensiones. Luego, para ejemplificar la implementación del método variacional en sistemas cuánticos, se utilizará una función de onda variacional para aproximar el estado base del modelo de Ising cuántico con campo magnético transversal, y el modelo de Bose–Hubbard en dos dimensiones. En ambos casos se comparan los resultados obtenidos a través del método de Monte Carlo variacional aunado con la reconfiguración estocástica con los obtenidos a través de diagonalización exacta de la matriz Hamiltoniana mediante el método de Lanczos [41].

#### 1.3.1. Modelo de Ising con campo externo transversal

La dinámica del sistema cuántico de Ising con campo transversal está dada por el Hamiltoniano de la Ec. (1.8). En esta sección, se pondrá a prueba el método variacional con este modelo para aproximar la función de onda del estado base.

Como primera aproximación variacional, se considera la *función de onda de Jastrow*,

definida por [42],

$$|\Psi_\nu\rangle = \sum_{\boldsymbol{\sigma}} \mathcal{J}_\nu(\boldsymbol{\sigma}) |\boldsymbol{\sigma}\rangle, \quad (1.34)$$

$$\mathcal{J}_\nu(\boldsymbol{\sigma}) = \exp \left[ -\frac{1}{2} \sum_{i,j} \nu_{ij} \sigma_i^z \sigma_j^z \right], \quad (1.35)$$

donde  $\boldsymbol{\sigma} = (\sigma_1^z, \sigma_2^z, \dots, \sigma_N^z)$  es la base del sistema, y  $\mathcal{J}_\nu(\boldsymbol{\sigma})$  es el llamado *factor de Jastrow* que depende del conjunto de parámetros  $\nu$ . Este factor puede interpretarse como un término de *pseudo-potencial* que introduce las correlaciones cuánticas entre los espines de la cadena, y  $\nu = \{\nu_{ij}\}$  es el conjunto de parámetros variacionales a optimizar en el algoritmo de aproximación. Estos parámetros miden las correlaciones entre los sitios de espín.

En la Fig. (1.4) se muestran los resultados obtenidos de aproximar la función de onda de Jastrow a través del algoritmo presentado en la sección 1.2.4 para el modelo de Ising con campo transversal con  $N = 20$  sitios de espín en las tres fases, en la fase ordenada con  $|g| < 1$  (1.4 a), en la fase sin brecha con  $|g| = 1$  (1.4 b), y en la fase desordenada  $|g| > 1$  (1.4 c). Para este modelo, es necesario optimizar  $N(N - 1)/2 = 190$  parámetros variacionales. En los tres casos, se toman 1008 muestras a través de un algoritmo de Metropolis–Hastings, con las primeras 500 muestras como la cola de las cadenas de Markov. Los parámetros variacionales se actualizan a través de 1000 pasos de optimización a una razón de aprendizaje de 0.1. El promedio  $E_J$  de los últimos 300 pasos de la energía variacional optimizada se compara con la energía  $E_{ext}$  calculada a través de la diagonalización de la matriz Hamiltoniana mediante un algoritmo de Lanczos.

Los resultados para la energía son a) una varianza de  $\sigma^2 = 0.1525$  de la energía y un error de  $\Delta E = E_J - E_{ext} = 0.0248$ , b)  $\sigma^2 = 0.1850$  y  $\Delta E = 0.0386$ , y c)  $\sigma^2 = 0.0091$  y  $\Delta E = 0.0007$ .

Como se puede apreciar, la función de onda de Jastrow tiene mejores resultados para la fase desordenada, donde el error entre las energías disminuye en un orden de magnitud. Por lo tanto, podemos acertar que, al menos para sistemas alrededor de  $N = 20$  espines, la función de onda variacional de Jastrow es capaz de aproximar con éxito al estado base a través del principio variacional.

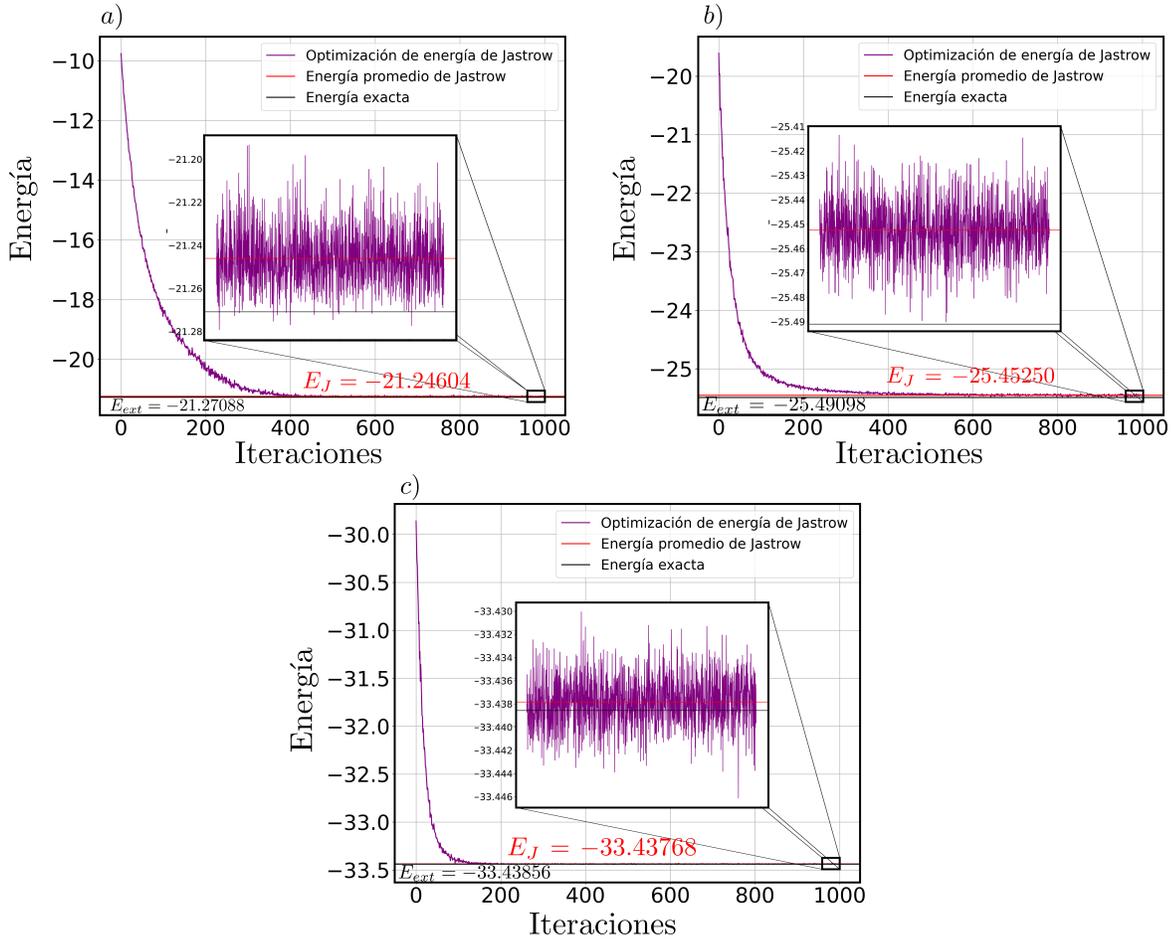


Figura 1.4: Optimización de la energía del estado variacional de Jastrow para el modelo de Ising con campo transversal con  $N = 20$  sitios de espín para  $J = 1$  y a)  $g = 0.5$ , b)  $g = 1.0$  y c)  $g = 1.5$ .

### 1.3.2. Modelo de Bose–Hubbard con la función de onda de Jastrow

Considere el modelo de Bose–Hubbard que consiste en una cadena de  $N$  sitios bosónicos, con a lo más  $n_{\text{max}}$  bosones por sitio, presentado en la sección 1.1.2. Al igual que en ejemplo anterior, como primera aproximación variacional se considera la función de onda

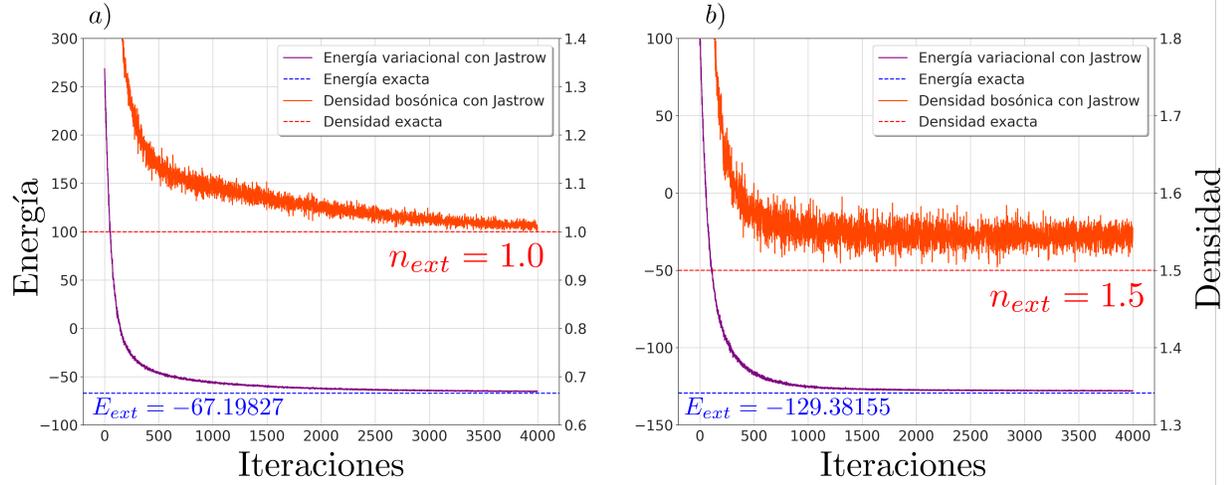


Figura 1.5: Optimización de la energía (líneas color morado) y de la densidad de ocupación (líneas color naranja) para un algoritmo variacional de 4000 pasos de optimización con la función de onda de Jastrow. Se comparan con los resultados obtenidos por diagonalización de Lanczos del modelo de Bose–Hubbard con  $N = 6$  sitios bosónicos en la cadena y  $n_{max} = 5$  número máximo de bosones por sitio, con a)  $U = 20.0$  y  $\mu = 11.0$  y b)  $U = 20.0$  y  $\mu = 20.0$ .

de Jastrow, donde un estado general del sistema está dado por [42],

$$|\Psi_\alpha\rangle = \sum_{\mathbf{n}} \mathcal{J}_\alpha(\mathbf{n}) |\mathbf{n}\rangle, \quad (1.36)$$

$$\mathcal{J}_\alpha(\mathbf{n}) = \exp \left[ -\frac{1}{2} \sum_i a_i n_i + \sum_{i,j} J_{ij} n_i n_j \right], \quad (1.37)$$

en donde  $\mathbf{n} = (n_1, n_2, \dots, n_N)$  es la base de posibles configuraciones del sistema en la base del operador número, y  $\mathcal{J}_\alpha(\mathbf{n}) = \langle \mathbf{n} | \Psi_\alpha \rangle$  es la distribución de probabilidad de Jastrow respecto a la base, y  $\alpha = (\vec{a}, \mathbf{J})$  es el conjunto de parámetros variacionales a optimizar para encontrar el conjunto de parámetros que minimice la energía variacional  $E_\alpha$  en el espacio de parámetros.

En la Fig.(1.5) se muestran los resultados obtenidos al optimizar la función de onda de Jastrow a través de un algoritmo de Monte Carlo variacional para encontrar el estado base del modelo de Bose–Hubbard en dos zonas de los parámetros de interacción: cuando el sistema se encuentra en la fase de aislante de Mott ( $U = 20.0$ ,  $\mu = 11.0$  y  $t = 1.0$ ), y cuando está en la fase de superfluido ( $U = 20.0$ ,  $\mu = 20.0$  y  $t = 1.0$ ). Para este ejemplo,

se utilizó una cadena de  $N = 6$  sitios bosónicos, con  $n_{max} = 5$  bosones máximos por sitio, por lo que el espacio de Hilbert tiene  $6^6 = 46656$  estados en la base. Esta función de onda variacional cuenta con  $N(N - 1)/2 + N = 21$  parámetros variacionales a optimizar. En el ejemplo se utiliza un algoritmo de Metrópolis–Hastings para tomar 1308 muestras de configuraciones posibles de la distribución de probabilidad dada por la función de onda de Jastrow para cada paso de optimización, tomando las últimas 1080 como la cola de la cadena de Markov.

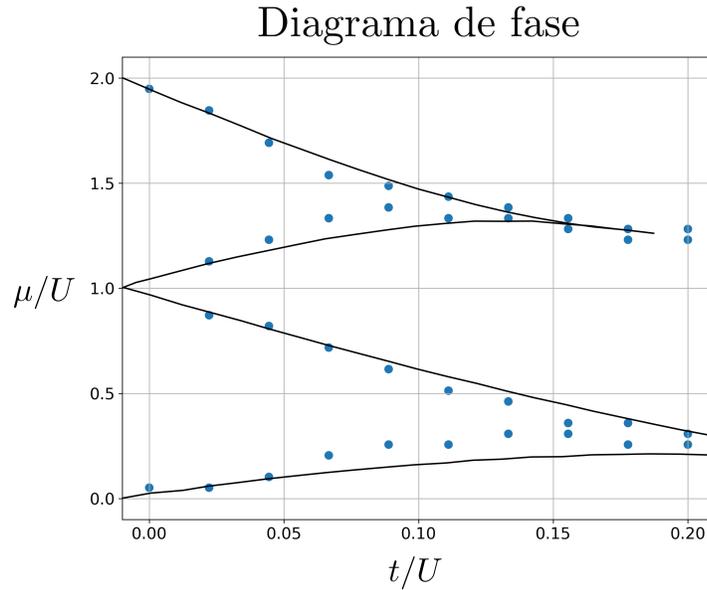


Figura 1.6: Comparación de diagrama de fase del modelo de Bose–Hubbard esbozado por la función de onda variacional de Jastrow (puntos azules) respecto al obtenido mediante diagonalización por Lanczos (líneas negras). Para el método variacional, la cadena cuenta con  $N = 8$  sitios bosónicos con a lo más  $n_{max} = 5$  bosones por espín, comparando la densidad de ocupación bosónica calculada por el método Monte Carlo respecto a su entero más cercano con tolerancia de  $\pm 0.05$ .

En la Fig. (1.6) se muestra el diagrama de fase donde se hizo una cuadrícula en el espacio de fase  $(t/U, \mu/U)$  de  $40 \times 10$ , y en cada punto se hace un algoritmo de optimización para el Hamiltoniano. Para cada punto de optimización, se promedia la densidad de ocupación bosónica  $\rho$  obtenida por el algoritmo de optimización y se considera que el modelo se encuentra en la fase de aislante de Mott si  $\rho$  es entera, con una tolerancia de  $\pm 0.05$ .

En la Fig. (1.4) se muestra como la varianza en la energía variacional es del orden del error de energía. Esto ocasiona que, aunque la energía sea satisfactoriamente aproximada

a la del estado base, los demás observables no sean aproximados, como se puede apreciar en la Fig. (1.5), donde el operador de densidad bosónica falla para el caso de superfluido. Esto ocasiona también que las fronteras entre las fases del modelo de Bose–Hubbard de la Fig. (1.6) no estén bien definidas respecto a las obtenidas de manera exacta.

Para solucionar esto, se propone utilizar una función de onda variacional utilizando *redes neuronales* que sea capaz de aprender de las correlaciones del sistema de manera más eficiente, y así poder capturar la física del estado base sin tener muchas fluctuaciones numéricas por el método Monte Carlo.

# Capítulo 2

## Redes neuronales y aprendizaje automático

### 2.1. Estructura de una red neuronal

Una red neuronal consiste en un conjunto de unidades de información, llamadas *neuronas* que se conectan entre ellas a través de *enlaces sinápticos* (o pesos de conexión). La red neuronal está caracterizada por una *matriz de conexión*  $\mathbf{W}$  que contiene la información de los enlaces sinápticos entre neuronas, donde  $W_{ij}$  es el peso de conexión entre la neurona  $i$  y la neurona  $j$ . También, a cada neurona  $i$  se le asigna un *término de sesgo* (o bias)  $b_i$ , que fija un valor constante a la señal de la neurona  $i$ , además de una *función de activación*  $f_i(\cdot)$  que depende de todos los valores de las demás neuronas, y especifica si la neurona  $i$  da una señal de activación o no. Por lo tanto, el estado de una red neuronal puede ser caracterizado por el conjunto  $\mathcal{W} = (\mathbf{W}, \mathbf{b})$  de pesos sinápticos y términos de sesgo de las neuronas, así como del conjunto  $\mathbf{f} = \{f_1, f_2, \dots\}$  de funciones de activación de cada neurona.

Generalmente, las neuronas están agrupadas por capas, permitiendo conexiones sinápticas sólo entre neuronas de capas contiguas. En este tipo de modelos, la red neuronal se *alimenta* dando valores a las neuronas de la primera capa, llamada *capa de entrada*, y las señales de las neuronas se van propagando a través de las capas hasta que la información llega a las neuronas de la última capa, llamada *capa de salida*, que se interpreta como el

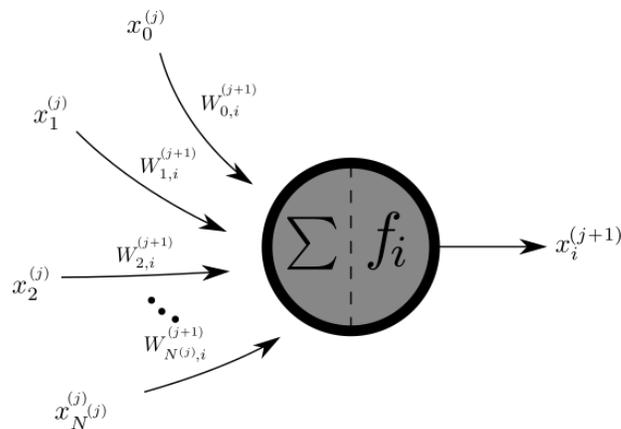


Figura 2.1: Representación gráfica de una neurona.

resultado de las operaciones neuronales.

Para ejemplificar la estructura de una red neuronal artificial, considere el ejemplo donde el modelo sólo tiene dos capas: la capa de entrada y la capa de salida, como la que se muestra en la Fig. (2.2). A este modelo de red neuronal se le conoce como *red de capa simple de realimentación*, donde se dice que una red neuronal es *realimentada* si la información la recibe desde las capas inferiores, empezando por la capa de entrada, hacia las capas superiores, terminando en la capa de salida, pero no viceversa. La capa de entrada está compuesta por  $N^{(0)}$  neuronas, donde el superíndice (0) denota el número que se le asigna a la capa de entrada. Las neuronas de esta capa están conectadas a través de una matriz de pesos  $\mathbf{W}^{(1)}$ , donde su elemento  $W_{ij}^{(1)}$  son los pesos de conexión de la  $i$ -ésima neurona de la capa de entrada que la conecta con la  $j$ -ésima neurona de la capa de salida. Suponga que los valores de las neuronas de la capa de entrada son denotados con  $x_i^{(0)}$ , entonces se define el *campo local de inducción* asociado a la  $j$ -ésima neurona de la capa de salida definido como la combinación lineal de estos valores,

$$z_j^{(1)} = \sum_{i=1}^{N^{(0)}} W_{ij}^{(1)} x_i^{(0)} + b_j^{(1)}, \quad (2.1)$$

donde  $b_j^{(1)}$  es el término de sesgo asociado a la  $j$ -ésima neurona de la capa de salida. Después de calcular el campo inducido, se aplica la llamada función de activación  $f$  para

obtener el valor de la  $j$ -ésima neurona de la capa de salida como,

$$\begin{aligned} x_j^{(1)} &= f\left(z_j^{(1)}\right) \\ &= f\left(\sum_{i=1}^{N^{(0)}} W_{ij}^{(1)} x_i^{(0)} + b_j^{(1)}\right), \end{aligned} \quad (2.2)$$

donde se supuso que la función de activación es la misma para todas las neuronas de la capa de salida. En la Fig. (2.1) se muestra una representación gráfica de una neurona que está conectada con  $N^{(j-1)}$  neuronas de la capa anterior, y donde el término de sesgo se interpreta como el término  $b_i^{(j+1)} = x_0^{(j)} W_{0,i}^{(j+1)}$ , donde se fija el valor de la neurona  $x_0^{(j)} = 1$  y  $W_{0,i}^{(j+1)} = b_i^{(j+1)}$ , por lo que el campo inducido se puede reescribir como,

$$z_j^{(k+1)} = \sum_{i=0}^{N^{(k)}} W_{ij}^{(k+1)} x_i^{(k)}. \quad (2.3)$$

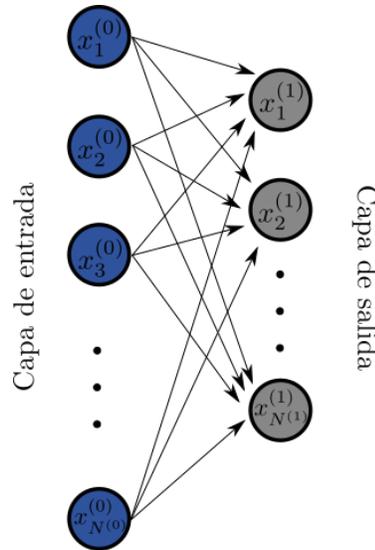


Figura 2.2: Estructura de red neuronal realimentada de capa simple.

Aún cuando esta estructura de red neuronal es simple, se ha demostrado que es suficiente para poder aproximar cualquier tipo de función diferenciable y lo suficientemente suave, proporcionando el número adecuado de neuronas y ajustando los parámetros  $\mathcal{W} = (\mathbf{W}^{(1)}, \mathbf{b}^{(1)})$  para que disminuya el error hasta obtener la precisión deseada. Sin

embargo, es posible mejorar esta precisión y facilitar el aprendizaje del modelo agregando más capas de neuronas entre la capa de entrada y la capa de salida.

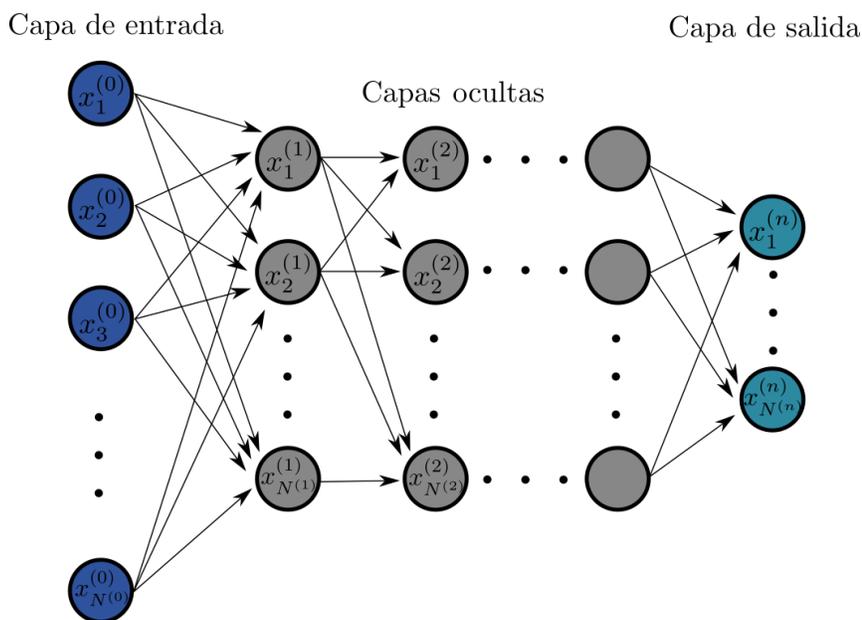


Figura 2.3: Estructura de red neuronal multicapa de realimentación.

Extrapolando para el caso donde se tienen  $n$  capas de neuronas, donde a las capas que están entre la capa de entrada y la capa de salida son llamadas *capas ocultas*, dado que los valores de sus neuronas –llamadas *neuronas ocultas*– están completamente definidas por sus conexiones sinápticas entre ellas. A este tipo de modelos neuronales se les conoce como *redes multicapa de realimentación*. Para una red neuronal multicapa de realimentación con  $n - 1$  capas ocultas, y denotando  $x_i^{(k)}$  como el valor de la  $i$ -ésima neurona en la capa  $k$  donde  $0 < k < n$  que tiene  $N^{(k)}$  neuronas, entonces el campo inducido de la neurona  $j$  de la capa  $k + 1$  está dado por,

$$z_j^{(k+1)} = \sum_{i=1}^{N^{(k)}} W_{ij}^{(k+1)} x_i^{(k)} + b_j^{(k+1)}, \quad (2.4)$$

donde  $\mathbf{W}^{(k+1)}$  es la matriz de conexión entre la capa  $k$  y la capa  $k + 1$ . Por lo tanto, el

valor de dicha neurona está dado por,

$$\begin{aligned} x_j^{(k+1)} &= f^{(k+1)} \left( z_j^{(k+1)} \right) \\ &= f^{(k+1)} \left( \sum_{i=1}^{N^{(k)}} W_{ij}^{(k+1)} x_i^{(k)} + b_j^{(k+1)} \right), \end{aligned} \quad (2.5)$$

donde se denota  $f^{(k+1)}$  a la función de activación de las neuronas de la capa  $k + 1$ .

En ciertos problemas a resolver con redes neuronales, es conveniente que algunas de las neuronas de la capa de salida estén conectadas con las de capas inferiores, e incluso con la propia capa de entrada. De esta forma, las salidas en un tiempo  $t$  dado son introducidas a la red de nuevo en un tiempo posterior  $t + 1$ . A este tipo de redes neuronales se les conoce como *redes neuronales recurrentes*, dado que se permite que los valores de capas superiores influyan en la activación de neuronas que se encuentran en las capas inferiores.

Estos modelos de redes neuronales son muy útiles, dado que lo único que se debe hacer para que la red neuronal aprenda es codificar la entrada y la salida, y elegir un esquema de aprendizaje adecuado para que el modelo se adapte a través de iteraciones de entrenamiento. A esto se le conoce como *aprendizaje automático*, y es la parte más esencial en el proceso del modelado a través de redes neuronales.

## 2.2. Aprendizaje automático

Como idea general, para que las redes neuronales aprendan a realizar tareas, es necesario que previamente hayan sido *entrenadas*. El *entrenar* a la red neuronal implica exponerla al entorno para que pueda capturar su funcionamiento, y así poder replicarlo. Para ello, se requiere de dos elementos a tener en cuenta en el proceso de aprendizaje de la red neuronal: el *conjunto de entrenamiento*, que son un conjunto de vectores de entrada que codifican la información contenida en el entorno del que se quiere aprender, y la *señal de error*, una función que permite visualizar el error entre la respuesta dada por la red neuronal y la respuesta deseada. Se puede categorizar el tipo de aprendizaje automático clasificando el conjunto de entrenamiento que se tenga en el problema; en ciertos casos se



El aprendizaje supervisado es muy útil, dado que es posible entrenar una red neuronal a través de este esquema para que pueda reconocer patrones en datos (imágenes, texto, lenguaje, etc), así como en bioinformática para reconocer huellas dactilares, textura del iris de ojos, identificación de individuos, entre otras cosas.

### 2.2.2. Aprendizaje sin profesor

En este esquema de aprendizaje, por otra parte, la red neuronal es expuesta a los datos de entrenamiento, y debe aprender a identificar patrones en ellos sin necesidad de que haya un profesor dando las respuestas correctas. Este esquema de aprendizaje se puede clasificar en dos tipos de subesquemas: el *aprendizaje no supervisado*, donde la red neuronal aprende a identificar patrones de datos no-etiquetados, clasificando grupos de datos en llamados *clusters* que comparten características similares. Esta habilidad de aprender sólo de los datos de entrada vuelve ideal a este esquema de aprendizaje para análisis de datos, reconocimiento de imágenes, o para la reducción de dimensionalidad en un conjunto de datos.

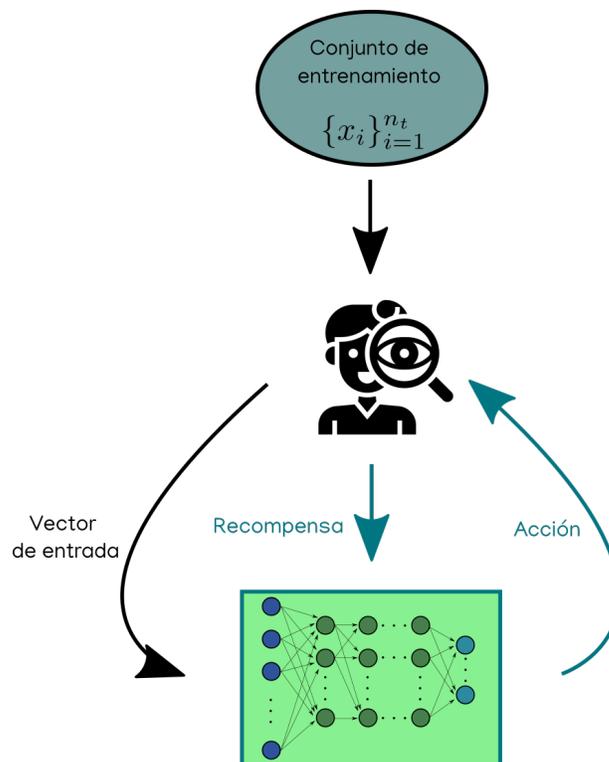


Figura 2.5: Esquema gráfico del algoritmo de aprendizaje reforzado.

El otro caso especial de aprendizaje sin profesor es el *aprendizaje reforzado*, en el cual, el proceso de aprendizaje consiste en alimentar a un *agente* con los vectores de entrada del conjunto de entrenamiento. Este agente después propone una *acción* sobre el estado de entrada, la cual es analizada por un *observador* que, si bien no conoce la respuesta correcta para cada uno de los ejemplos del conjunto de entrenamiento, es capaz de evaluar cuando una respuesta dada por el agente se acerca a la deseada. Después de evaluar el resultado obtenido por el agente, el observador da una *recompensa* si la respuesta se acerca a la deseada después de la acción propuesta por el agente, o le da una *penalización* si es que se aleja. La forma en la que el observador evalúa la salida es a través de una *función coste* que depende del problema específico que se quiera resolver. El agente puede ser una red neuronal, tal como se muestra en la Fig. (2.5). Entonces, en cada iteración del algoritmo de aprendizaje, los pesos sinápticos de la red neuronal se modifican de tal forma que esta función coste disminuya hasta que haya aprendido a obtener la respuesta deseada.

En el contexto de la física, y en especial, de la mecánica cuántica de muchos cuerpos, se sabe que el estado base de cualquier sistema físico es el que tiene la energía mínima. Por lo tanto, el aprendizaje reforzado es muy llamativo de utilizar para aproximar el estado base de un sistema cuántico. Para lograr esto, es necesario encontrar una estructura neuronal capaz de modelar la función de onda del sistema cuántico a estudiar, y a partir de ahí utilizar un esquema de aprendizaje reforzado para que la red aprenda a modificar sus parámetros sinápticos, tal que se minimice su función coste -que en este caso, es la energía total del sistema- hasta llegar al mínimo, y así aproximar la función de onda del estado base del sistema.

### 2.3. Aproximar el estado base de sistemas cuánticos de muchos cuerpos a través de redes neuronales

La función de onda de un sistema cuántico es el objeto que contiene toda la información del estado completo del sistema. Operando con ella, es posible calcular observables del sistema en ese estado, lo que permite comparar con alguna realización experimental de ese

sistema, e incluso nos permite inferir propiedades como cambios de fase de la materia. Dado que el estudio experimental de sistemas cuánticos complejos de muchos cuerpos está siendo tendencia entre la comunidad científica, es importante también tener una contraparte teórica que sostenga los resultados obtenidos en el laboratorio. Sin embargo, no siempre se tiene la información requerida de cómo actúan las correlaciones cuánticas en el sistema para construir una función de onda variacional como la de Jastraw. Para solucionar esto, se puede hacer uso de la capacidad inherente de las redes neuronales de aprender estas correlaciones, sin necesidad de inferirlas en un principio en el modelo teórico.

Como se dijo antes, las redes neuronales son aproximadores universales de funciones diferenciables, y dado que se puede expresar la función de onda de un sistema cuántico como una distribución de probabilidad sobre posibles estados en una base de un operador Hermitiano definido en el espacio de Hilbert, entonces las redes neuronales son una propuesta ideal para modelar dichas funciones de onda. Además, dado que computacionalmente las redes neuronales no son muy demandantes en términos operacionales, también son llamativas por el posible escalamiento que se pueda obtener en el tiempo de cómputo respecto a otros métodos de aproximación.

### 2.3.1. Funciones de onda variacionales con redes neuronales

Para mostrar cómo es el modelo de la función de onda a través de una red neuronal, suponga un sistema cuántico cuya dinámica está definida por el operador Hamiltoniano  $\hat{H}$ , y cuenta con  $N$  partículas interactuantes. La meta es utilizar una estructura de red neuronal para poder modelar la distribución de probabilidad respecto a una base computacional del sistema. Dado que se desea que la red neuronal aprenda del sistema físico, la capa de entrada deberá tener codificado el estado, teniendo entonces  $N$  unidades en la capa de entrada, y por salida, se desea que el modelo dé la amplitud de probabilidad de dicho estado, por lo que la capa de salida deberá tener sólo una neurona. Entonces, se puede construir una función de onda variacional a través de dicha red neuronal como,

$$|\Psi_{\mathcal{W}}\rangle = \sum_x \psi_{\mathcal{W}}(x) |x\rangle, \quad (2.6)$$

donde  $\psi_{\mathcal{W}}(x)$  es la amplitud de probabilidad modelada por la red neuronal con conjunto de pesos sinápticos  $\mathcal{W} = (\mathbf{W}, \mathbf{b})$ , y  $\{x\}$  es la base del sistema. Dado que el objetivo es aproximar la función de onda del estado base, la forma de hacerlo es minimizando la energía del modelo neuronal.

Al igual que en la Sección (1.2.2), la energía de una distribución  $\psi_{\mathcal{W}}(x)$  respecto a los posibles estados del sistema está dado por,

$$E_{\mathcal{W}} = \langle \hat{H} \rangle_{\mathcal{W}} = \mathbb{E}[\tilde{H}]_{\mathcal{W}}, \quad (2.7)$$

donde  $\hat{H}$  es el operador Hamiltoniano del sistema cuántico, y se define la *energía local* del estado  $x$  respecto a la distribución de probabilidad  $\psi_{\mathcal{W}}(x)$  como,

$$\tilde{H}(x) = \sum_{x'} \frac{\psi_{\mathcal{W}}(x')}{\psi_{\mathcal{W}}(x)} \langle x | \hat{H} | x' \rangle \quad (2.8)$$

y se define al operador  $\mathbb{E}[\cdot]_{\mathcal{W}}$  como el *promedio* respecto a la distribución de probabilidad  $\psi_{\mathcal{W}}(x)$  como,

$$\mathbb{E}[A] = \sum_{x'} \frac{|\psi_{\mathcal{W}}(x')|^2}{\langle \Psi_{\mathcal{W}} | \Psi_{\mathcal{W}} \rangle} A(x), \quad (2.9)$$

con  $\langle \Psi_{\mathcal{W}} | \Psi_{\mathcal{W}} \rangle = \sum_x |\psi_{\mathcal{W}}(x)|^2$  es la norma del estado cuántico neuronal.

Entonces, dado que no se conoce la distribución de probabilidad de configuraciones del estado base, es necesario hacer uso de aprendizaje reforzado. El conjunto de entrenamiento es el espacio de todas las posibles configuraciones  $\{x\}$ , y la red neuronal da por respuesta la probabilidad para cada una de ellas, a través de la distribución  $\psi_{\mathcal{W}}(x)$ . Después, el agente propone un cambio en los pesos sinápticos de la red a  $\mathcal{W}'$ , tal que cambie dicha distribución de probabilidad a una distinta  $\psi_{\mathcal{W}'}(x)$ . Las dos probabilidades modeladas por la red neuronal para cada una de las configuraciones es evaluada por el observador, calculando la energía para ambas distribuciones  $E_{\mathcal{W}}$  y  $E_{\mathcal{W}'}$  con la Ec. (2.7). Si  $E_{\mathcal{W}'} < E_{\mathcal{W}}$ , entonces, se recompensa positivamente al agente, y se penaliza negativamente si pasa lo contrario. De esta forma, la red neuronal tenderá a cambiar sus pesos sinápticos de manera que  $E_{\mathcal{W}}$  disminuya.

Aún cuando el procedimiento antes descrito es suficiente para que la energía disminuya, es posible acelerar el proceso utilizando *optimizadores* que guíen al agente a tomar acciones más acertadas. Estos optimizadores pueden ser procedimientos de minimización de funciones, como el descenso al gradiente o la reconfiguración estocástica respecto a los pesos sinápticos.

### 2.3.2. Máquina de Boltzmann Restringida

Las máquinas de Boltzmann Restringidas [43] (RBM por sus siglas en inglés) son un tipo de red neuronal estocásticas utilizadas en el aprendizaje automático para modelar la distribución de probabilidad de un conjunto de vectores binarios que fueron muestreados desde una distribución desconocida. La razón de usar este tipo de redes neuronales recurrentes para aproximar la función de onda del estado base de sistemas cuánticos de muchos cuerpos es el hecho de que se ha probado de manera satisfactoria la universalidad del modelo en aproximar distribuciones de probabilidad con la precisión que se desee [44], además de que la estructura neuronal recurrente permite que se consideren muchos tipos de correlaciones cuánticas en el sistema, mejorando la precisión de los cálculos [45]. Ya se han reportado trabajos donde se utiliza este modelo de red neuronal para aproximar distintos sistemas cuánticos de manera satisfactoria, como el modelo de Heisenberg [11], el modelo  $J_1 - J_2$  [46] [47], el modelo de Bose–Hubbard [48], entre otros.

Las RBM se componen de dos capas de neuronas que toman valores binarios, la *capa visible*, donde las neuronas codifican la entrada de los vectores binarios, y la *capa oculta*, que se compone de unidades binarias estocásticas. Las conexiones se limitan sólo entre neuronas de capas distintas. La capa oculta se encarga de crear correlaciones entre las unidades visibles sin la necesidad de que haya interacción directa entre las neuronas de entrada. Dependiendo de los valores de los pesos sinápticos entre las neuronas, así como los términos de sesgo, la red neuronal asigna una probabilidad a cada vector del conjunto de vectores binarios de entrenamiento.

La distribución de probabilidad que modela una RBM con  $n_v$  neuronas binarias en la capa visible y  $n_h$  en la capa oculta, y con parámetros sinápticos dados por  $\mathcal{W} = (\mathbf{W}, \vec{a}, \vec{b})$ , donde  $W_{ij}$  es el peso sináptico que conecta a la neurona  $i$  de la capa visible con la neurona

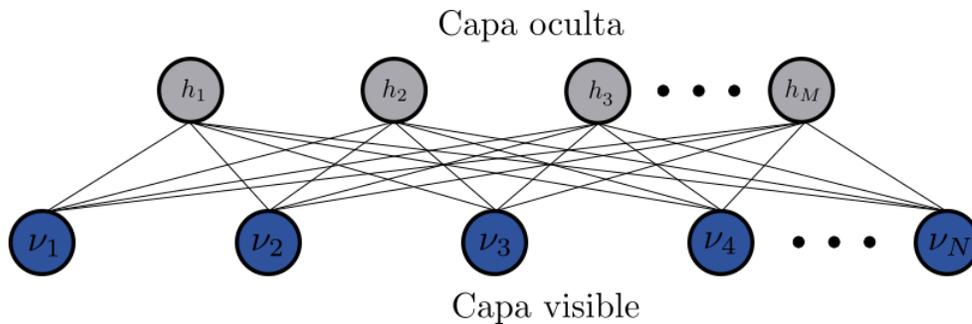


Figura 2.6: Estructura de una máquina de Boltzmann restringida.

$j$  de la capa oculta,  $a_i$  es el término de sesgo de la neurona  $i$  de la capa visible y  $b_j$  el término de sesgo de la neurona  $j$  de la capa oculta, con vector de entrada  $\vec{v} = (v_1, \dots, v_{n_v})$  y vector oculto  $\vec{h} = (h_1, \dots, h_{n_h})$  está dada por,

$$P_{RBM}(\vec{v}, \vec{h}) = \frac{1}{Z(\mathcal{W})} \exp(-E_{RBM}(\vec{v}, \vec{h}; \mathcal{W})), \quad (2.10)$$

donde  $Z(\mathcal{W})$  es la *función de partición* que normaliza la distribución de probabilidad de la RBM, y  $E_{RBM}(\vec{v}, \vec{h}; \mathcal{W})$  es la *energía libre de la máquina de Boltzmann restringida* dada por,

$$E_{RBM}(\vec{v}, \vec{h}; \mathcal{W}) = - \sum_{i=1}^{n_v} a_i v_i - \sum_{j=1}^{n_h} b_j h_j - \sum_{i=1}^{n_v} \sum_{j=1}^{n_h} W_{ij} v_i h_j. \quad (2.11)$$

Dado que las unidades de la capa oculta  $\vec{h}$  son estocásticas, se puede calcular la probabilidad condicional  $P_{RBM}(\vec{v} | \vec{h}; \mathcal{W})$  de que la capa de entrada contenga los valores  $\vec{v}$  cuando la capa oculta es  $\vec{h}$  con parámetros sinápticos  $\mathcal{W}$ ,

$$\begin{aligned} P_{RBM}(\vec{v} | \vec{h}; \mathcal{W}) &\propto \exp(a_i v_i + \sum_{j=1}^{n_h} W_{ij} v_i h_j) \\ &\propto \prod_{i=1}^{n_v} P_{RBM}(v_i | \vec{h}; \mathcal{W}), \end{aligned} \quad (2.12)$$

y una solución similar para la probabilidad condicional de que se tenga el vector oculto  $\vec{h}$  cuando la entrada es  $\vec{v}$ . Por lo tanto, la distribución de probabilidad respecto a los vectores posibles de entrada  $\vec{v}$  se calcula *marginalizando* sobre todos los posibles estados de la capa

oculta,

$$P_{RBM}(\vec{v}; \mathcal{W}) = \sum_{\vec{h}} P_{RBM}(\vec{v}, \vec{h}; \mathcal{W}), \quad (2.13)$$

por lo que se puede tomar muestras de un conjunto de posibles entradas respecto al modelo de distribución del RBM.

Cuando se intenta calcular el estado base de un sistema cuántico de muchos cuerpos, en principio no se conoce la distribución de probabilidad de los posibles configuraciones del sistema en dicho estado, ya que eso mismo es la amplitud de la función de onda  $P(\vec{v}) = |\langle \vec{v} | \Psi \rangle|^2$ , donde  $\vec{v}$  son los vectores de la base computacional en la que se está estudiando el sistema cuántico. Entonces, la meta es aproximar dicha distribución de probabilidad a través de una máquina de Boltzmann restringida, modificando los parámetros de conexión y los términos de sesgo a través de un algoritmo de aprendizaje automático. Por lo tanto, utilizando el método variacional, en principio es posible aproximar el estado base de un sistema cuántico de muchos cuerpos a través de una RBM, optimizando los valores sinápticos con un algoritmo de aprendizaje reforzado, donde la recompensa se da por positiva cuando la energía asociada a la RBM disminuye al actualizar los parámetros sinápticos, y se castiga con recompensa negativa cuando la energía aumenta.

Dado que las neuronas de las RBMs son binarias, codificar modelos con sistemas de espines es directo: si la red cuenta con  $N$  sitios de espín, entonces un estado accesible  $|\sigma_1^z, \dots, \sigma_N^z\rangle$ , con  $\sigma_i^z \in \pm 1$  el valor de la proyección en  $z$  del espín en el sitio  $i$  de la cadena, se puede codificar en una RBM con  $N$  neuronas en la capa de entrada, y donde se asigna el valor  $v_i = 1$  ( $v_j = 0$ ) sólo si  $\sigma_i^z = +1$  ( $\sigma_j^z = -1$ ).

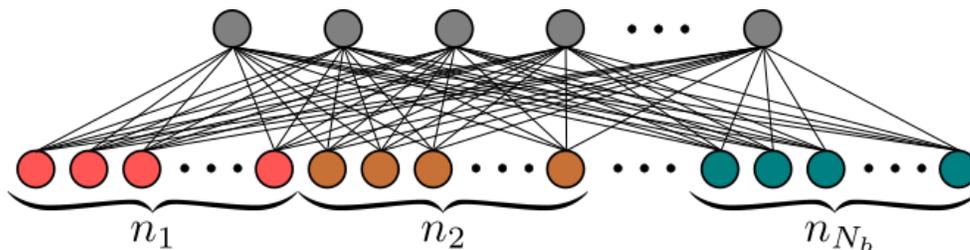


Figura 2.7: Representación gráfica de la codificación tipo One-Hot en la entrada de una RBM.

Sin embargo, esto cambia cuando se tiene otro tipo de objetos cuánticos cuyos números cuánticos no son binarios, tales como los sitios de ocupación bosónica como en el modelo de Bose–Hubbard. Para este tipo de casos, es necesario hacer otro tipo de codificación en la entrada, de tal forma que la red neuronal pueda interpretar los diferentes estados accesibles de ocupación bosónica del sistema de manera distintiva. Para ello, se hace uso de una *codificación tipo one-hot* en la entrada.

A modo de ejemplo, suponga un sistema cuántico que cuenta con  $N_b$  sitios bosónicos, en los cuales se permiten, a lo más,  $n_{max}$  bosones por sitio, y cuya distribución de probabilidad se quiere modelar a través de una RBM. La codificación en la capa de entrada se hace como sigue: a cada sitio de ocupación bosónica se le asigna un conjunto de  $n_{max} + 1$  neuronas de la capa de entrada. Por ejemplo, al sitio  $j$  se le asigna un vector  $\vec{v}_j = (v_0, v_1, \dots, v_{n_{max}})$ , donde el valor de cada neurona está dado por  $v_i = \delta_{n_j, i}$ , con  $n_j$  el número de ocupación bosónica del sitio  $j$ , como se muestra en la Fig. (2.7). De esta forma, es posible capturar las distintas configuraciones del sistema a través de una RBM con  $n_v = N_b * (n_{max} + 1)$  neuronas en la capa visible, y  $n_h$  neuronas en la capa oculta.

Ahora que se ha presentado el modelo de red neuronal, se pondrá a prueba su efectividad en los modelos de muchos cuerpos ya estudiados con la función de onda de Jastrow en el capítulo anterior.

## 2.4. Ejemplos de aproximación con estados cuánticos neuronales

Como se vio en el capítulo anterior, a pesar de que la función de onda de Jastrow es capaz de optimizar la energía del estado base del modelo de Ising con campo magnético transversal y el modelo paradigmático de Bose–Hubbard, aún sigue teniendo problemas para capturar la física del sistema, fallando al tratar de calcular el diagrama de fase para el último modelo mencionado. En esta sección, se pondrá a prueba el estado cuántico neuronal de las máquinas de Boltzmann restringidas para comparar los resultados con la función de onda de Jastrow. En estos ejemplos, se hizo uso de la biblioteca Netket 3 [49]

para definir el modelo de red neuronal que modela la función de onda variacional, así como el proceso de optimización.

### 2.4.1. Modelo de Ising transversal

Suponga un modelo de Ising con campo externo transversal con  $N$  sitios de espín, donde su dinámica está definida por el Hamiltoniano de la Ec. (1.8). Ahora, se desea aproximar un modelo de estado cuántico neuronal a través de una máquina de Boltzmann restringida. Para construir dicha arquitectura, es necesario que haya  $n_v = N$  neuronas binarias en la capa visible, y  $n_h$  unidades ocultas. Entonces, la probabilidad modelada por la red neuronal recurrente para una configuración  $\boldsymbol{\sigma} = (\sigma_1^z, \sigma_2^z, \dots, \sigma_N^z)$  de la cadena de espines está dada por,

$$\psi_{\mathcal{W}}(\boldsymbol{\sigma}) = P_{RBM}(\boldsymbol{\sigma}; \mathcal{W}) = \frac{\exp\left(\sum_{i=1}^N a_i \sigma_i^z\right)}{Z(\mathcal{W})} \prod_{j=1}^{n_h} 2 \cosh\left(b_j + \sum_{i=1}^N W_{ij} \sigma_i^z\right), \quad (2.14)$$

donde  $\mathcal{W} = (\mathbf{a}, \mathbf{b}, \mathbf{W})$  es el conjunto de parámetros sinápticos de la red neuronal, con  $a_i$  el término de sesgo de la neurona  $i$  en la capa de entrada,  $b_j$  el término de sesgo de la neurona  $j$  en la capa oculta y  $W_{ij}$  el peso sináptico que mide la interacción entre la neurona  $i$  de la capa de entrada con la neurona  $j$  de la capa de salida. Entonces, el algoritmo de optimización busca el conjunto  $\mathcal{W}$  en el espacio de parámetros tal que la energía variacional  $E_{\mathcal{W}}$  se minimice. Para hacer esto, en cada paso de optimización se calcula la energía asociada a la función de onda a través de las Ecs. (1.21) y 1.22, donde,

$$E_{\mathcal{W}} = \frac{\langle \Psi_{\mathcal{W}} | \hat{H}_I | \Psi_{\mathcal{W}} \rangle}{\langle \Psi_{\mathcal{W}} | \Psi_{\mathcal{W}} \rangle}, \quad (2.15)$$

con  $|\Psi_{\mathcal{W}}\rangle = \sum_{\boldsymbol{\sigma}} \psi_{\mathcal{W}}(\boldsymbol{\sigma}) |\boldsymbol{\sigma}\rangle$  el estado cuántico neuronal de la Ec. (2.6) para el caso de la RBM en el modelo de Ising con campo magnético externo transversal.

En la Fig. (2.8) se muestran los resultados obtenidos al optimizar el estado cuántico variacional de la Ec. (2.14) a través de un algoritmo de Monte Carlo variacional, presentado en la sección 1.2.4, para a) la fase ordenada, con  $|g| < 1$ , b) la fase sin brecha con  $|g| = 1$  y c) la fase desordenada, con  $|g| > 1$ . La red neuronal utilizada cuenta con  $n_h = N = 20$

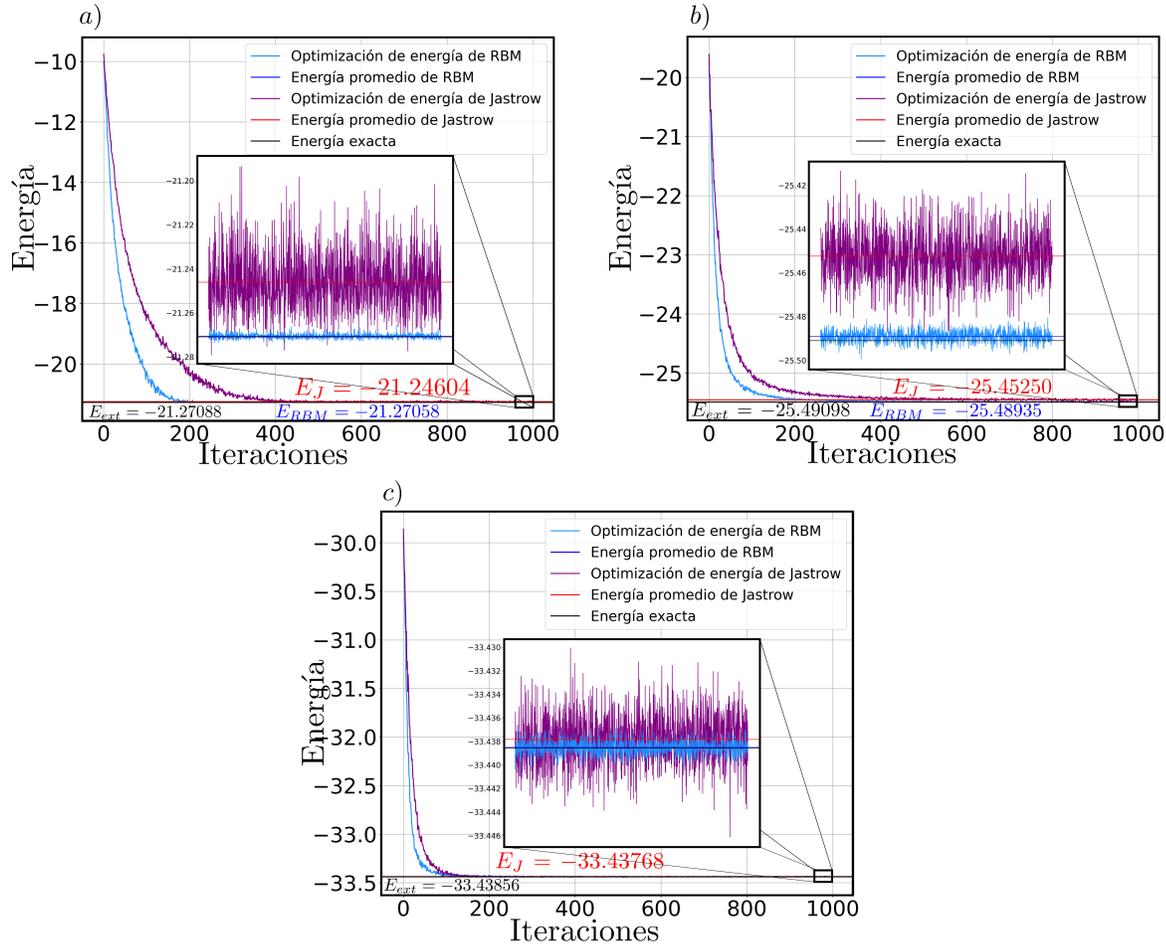


Figura 2.8: Optimización de la energía del estado variacional de Jastrow para el modelo de Ising con campo transversal con  $N = 20$  sitios de espín para  $J = 1$  y a)  $g = 0.5$ , b)  $g = 1.0$  y c)  $g = 1.5$ .

neuronas en la capa oculta, por lo que es necesario optimizar  $2N + N^2 = 440$  parámetros variacionales. En los tres procesos se tomaba la función de onda con 1380 muestras de configuraciones, descartando las primeras 300. Se actualizaron los parámetros a través de 1000 pasos de optimización a una razón de aprendizaje de 0.1, y se toma por promedio de la energía variacional los últimos 300 pasos de optimización. Los resultados obtenidos son: a)  $\sigma^2 = 0.0021$  y  $\Delta E = 0.0003$ , b)  $\sigma^2 = 0.0144$  y  $\Delta E = 0.0019$  y c)  $\sigma^2 = 0.0006$  y  $\Delta E = 0.00005$ .

Como se puede apreciar por los resultados, existe una mejora sustancial en el error de la energía del estado base de la función de onda variacional modelada por una RBM respecto

a la de Jastrow, al igual que una reducción considerable de la fluctuación de la energía. Además, aún cuando el número de parámetros a optimizar aumentó 20 veces, dado que se puede controlar el número de muestras de configuraciones por paso de optimización, es posible controlar el tiempo de cómputo sin perder la mejora de precisión en los cálculos de la función de onda variacional de la RBM.

### 2.4.2. Modelo de Bose–Hubbard

Suponga un modelo de Bose–Hubbard, donde la red consta de  $N$  sitios bosónicos, y en cada uno se permiten a lo más  $n_{max}$  bosones. Para poder interpretar las diferentes configuraciones del sistema en la base del operador número en una RBM, es necesario hacer una codificación del tipo One–Hot en la capa de entrada, como en la Fig. (2.7). Haciendo esto, la distribución de probabilidad sobre el conjunto de configuraciones accesibles del sistema  $\mathbf{n} = (n_1, \dots, n_N)$ , con  $0 \leq n_i \leq n_{max}$ , para  $i = 1, \dots, N$  modelada por una RBM con  $n_v = N * (n_{max} + 1)$  neuronas en la capa visible y  $n_h$  neuronas en la capa oculta está dada por,

$$\psi_{\mathcal{W}}(\mathbf{n}) = P_{RBM}(\vec{v}(\mathbf{n}); \mathcal{W}) = \frac{\exp\left(\sum_{i=1}^N \sum_{j=1}^{n_{max}} a_i^j v_i^j\right)}{Z(\mathcal{W})} \prod_{k=1}^{n_h} 2 \cosh\left(b_k + \sum_{i=1}^{N_b} \sum_{j=0}^{n_{max}} W_{ij}^k v_i^j\right) \quad (2.16)$$

donde  $\vec{v}(\mathbf{n})$  es el vector que representa a la configuración  $\mathbf{n}$  de la cadena en la capa de entrada de la RBM a través de una codificación One–Hot donde su elemento  $v_i^j$  es la neurona que representa al sitio bosónico  $i$  en el sentido de que  $v_i^j = 1$  significa que hay  $j$  bosones en el sitio bosónico  $i$  de la cadena. El conjunto de parámetros  $\mathcal{W} = (\mathbf{a}, \mathbf{b}, \mathbf{W})$  son las conexiones sinápticas de la red neuronal, donde  $a_i^j$  y  $b_k$  son los parámetros de sesgo de la neurona  $v_i^j$  en la capa de entrada y de la neurona  $h_k$  de la capa oculta, respectivamente, y  $W_{ij}^k$  es el parámetro de conexión entre la neurona  $v_i^j$  de la capa de entrada con la neurona  $h_k$  de la capa oculta. En el algoritmo de minimización, estos parámetros se optimizan en pos de minimizar la energía variacional  $E_{\mathcal{W}}$  obtenida calculando el valor esperado del

operador Hamiltoniano  $\hat{H}_{BH}$  del modelo de Bose–Hubbard,

$$E_{\mathcal{W}} = \frac{\langle \Psi_{\mathcal{W}} | \hat{H}_{BH} | \Psi_{\mathcal{W}} \rangle}{\langle \Psi_{\mathcal{W}} | \Psi_{\mathcal{W}} \rangle}, \quad (2.17)$$

donde  $|\Psi_{\mathcal{W}}\rangle = \sum_{\mathbf{n}} \psi_{\mathcal{W}}(\mathbf{n}) |\mathbf{n}\rangle$  es el estado cuántico neuronal del sistema respecto al estado sináptico  $\mathcal{W}$  de la RBM.

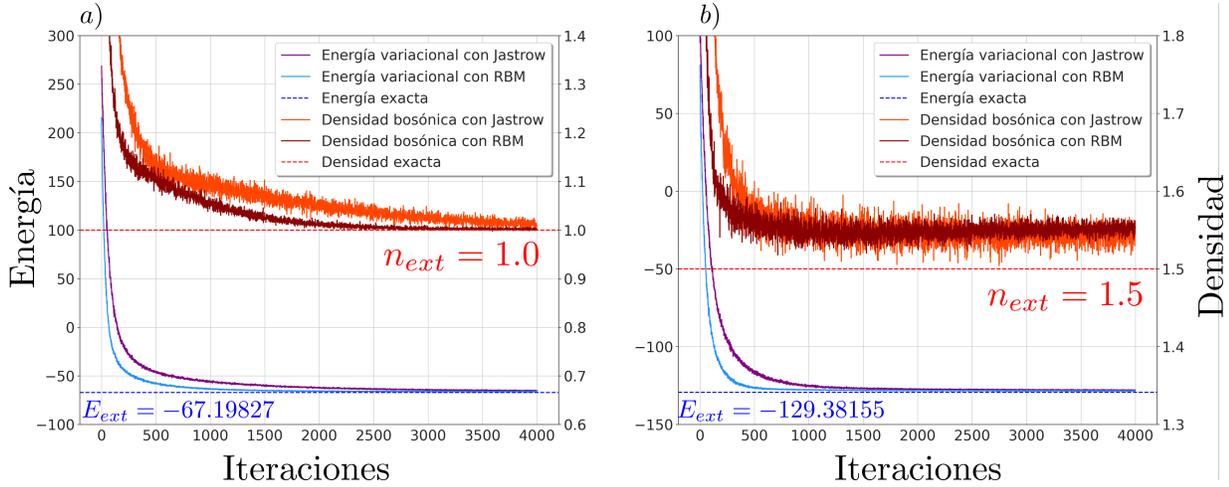


Figura 2.9: Comparación de la optimización de la energía (líneas color morado) y densidad de ocupación (líneas color naranja) por la función de onda variacional de Jastrow respecto a la optimización de la energía (líneas color celeste) y la densidad de ocupación (líneas color rojo) del estado cuántico neuronal de la RBM. Se comparan con los resultados obtenidos por diagonalización de Lanczos del modelo de Bose–Hubbard con  $N = 8$  sitios bosónicos en la cadena y  $n_{max} = 5$  número máximo de bosones por sitio, con a)  $U = 20.0$  y  $\mu = 11.0$  y b)  $U = 20.0$  y  $\mu = 20.0$ .

En la Fig. (2.9) se comparan los resultados obtenidos por la optimización de la función de onda de Jastrow con la del estado cuántico neuronal de la RBM. Como se puede apreciar, aún cuando la cadena de sitios bosónicos es pequeña con  $N = 8$  sitios bosónicos y un número máximo de  $n_{max} = 5$  bosones por sitio, se nota la aceleración en la optimización en el sentido de que energía variacional converge más pronto para el estado cuántico neuronal respecto a la función de onda variacional de Jastrow. Aún así, el modelo neuronal sigue teniendo problemas capturando la física del operador de densidad bosónica para cuando el sistema se encuentra en la fase de superfluido, pero en la fase de aislante de Mott, la RBM presenta una notable mejora capturando el valor exacto de la densidad bosónica, haciéndolo más rápido y con menos varianza.

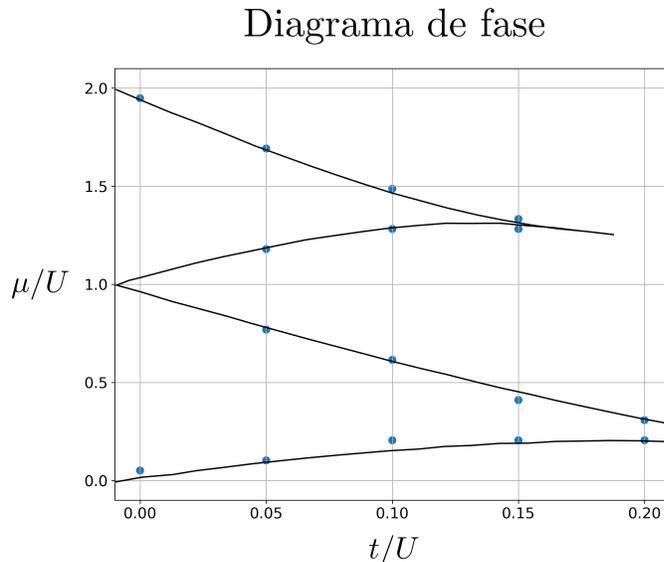


Figura 2.10: División de fases entre el aislante de Mott  $\rho \in \mathbb{N}$  y superfluido esbozada por la RBM (puntos azules), comparando la densidad bosónica calculada por la optimización y su parte entera, con una tolerancia de  $\pm 0.05$ . Se compara con los resultados obtenidos por diagonalización de Lanczos. El modelo de Bose–Hubbard tiene  $N = 8$  sitios bosónicos con a lo más  $n_{max} = 5$  bosones por sitio.

Aún cuando el modelo de red neuronal sigue presentando problemas para capturar la física del sistema cuando se encuentra en la fase de superfluido, es interesante observar qué tanto mejora este estado cuántico neuronal en capturar el diagrama de fase, y cómo se comparan las divisiones entre las dos fases. Calculando el observable de densidad bosónica, promediando las últimas 300 cadenas de Markov del sistema y considerando que si ésta es entera con una tolerancia de  $\pm 0.05$  el sistema se encuentra en la fase de aislante de Mott, en la Fig. (2.10) se muestra como la RBM es capaz de delimitar las zonas de las distintas fases en el espacio de parámetros  $(t/U, \mu/U)$  de forma más nítida y precisa de lo que lo hace la función de onda variacional de Jastrow, como se aprecia en la Fig. (1.6). Esto se debe a que es más precisa en capturar la física en la fase estable de la fase de aislante de Mott. En la figura también se comparan los resultados con los obtenidos al diagonalizar la matriz con el método de Lanczos.

Una vez que se ha probado la efectividad de los estados cuánticos neuronales a través de máquinas de Boltzmann restringidas para aproximar el estado base de un sistema cuántico de muchos cuerpos con los dos ejemplos paradigmáticos antes mencionados, el

paso natural a seguir es probar en sistemas mucho más complejos, como en sistemas mixtos que combinen varios tipos de números cuánticos. En el siguiente capítulo se tomará por ejemplo un modelo de red de muchos cuerpos cuánticos que combina una interacción del tipo Bose–Hubbard acoplado a un sistema  $\mathbb{Z}_2$  de dos niveles de espines en las interacciones, y se usará un modelo de red neuronal para tratar de aproximar su estado base: el llamado *modelo  $\mathbb{Z}_2$  de Bose–Hubbard*.

# Capítulo 3

## Modelo $\mathbb{Z}_2$ Bose–Hubbard

Una vez que se ha mostrado la capacidad de las máquinas de Boltzmann restringidas para modelar las funciones de onda de sistemas cuánticos de muchos cuerpos constituidos por partículas descritas en términos de la misma base de operadores, como en la base de espines para el modelo de Ising con campo transversal o el modelo de Bose–Hubbard para la base del operador número bosónico, el siguiente paso es poner a prueba este modelo en sistemas con partículas descritas por bases compuestas y observar si es capaz de capturar las correlaciones que estas interacciones conllevan.

En este capítulo se estudia un sistema de muchos cuerpos cuánticos que consta de una cadena de sitios bosónicos que interactúan entre ellos como en el modelo de Bose–Hubbard tradicional, con la diferencia de que la interacción entre los sitios está dominada por *ligas* de espines que condiciona el tipo de integral de salto de los bosones en los sitios bosónicos, así como la amplitud de probabilidad de salto. Este ejemplo ha sido estudiado analíticamente por medio de una aproximación en el límite de bosones rígidos, donde  $U \rightarrow \infty$ , obteniendo así un diagrama de fase donde el sistema bosónico se encuentra a medio llenado y las ligas pasan de una fase de total ferromagnetismo a total anti-ferromagnetismo con una configuración de Néel [19]. En este capítulo se mostrará cómo el estado modelado por una máquina de Boltzmann restringida puede replicar los resultados en el límite de bosones robustos ya obtenidos en este artículo sin necesidad de hacer inferencias sobre la función de onda [19], lo que nos permite estudiar otros límites con este modelo de red neuronal.

En la primera sección se presenta el modelo  $\mathbb{Z}_2$  Bose–Hubbard y se demuestran las

soluciones obtenidas en [19]. Luego, en la segunda sección, se introduce el modelo de máquina de Boltzmann restringida para hacer los cálculos en el sistema cuántico, presentando los resultados obtenidos en el límite de bosones robustos. Por último, en la tercera sección se muestran los resultados obtenidos para el modelo fuera del límite estudiado en el artículo [19].

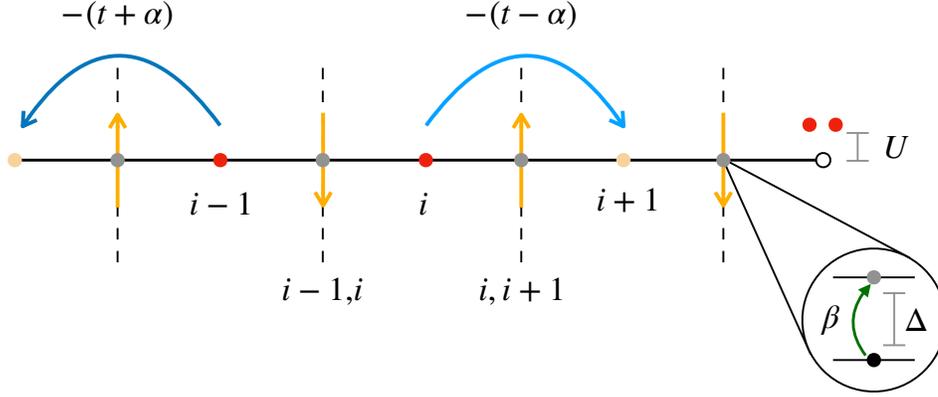
### 3.1. Dinámica del sistema

El Hamiltoniano del modelo de  $\mathbb{Z}_2$  Bose–Hubbard describe una red unidimensional de bosones acoplados a un sistema de campos  $\mathbb{Z}_2$ , donde la interacción está descrita por [19],

$$H_{\mathbb{Z}_2} = -\alpha \sum_i (\hat{b}_i^\dagger \hat{\sigma}_{i,i+1}^z \hat{b}_{i+1} + \text{H.c.}) + \beta \sum_i \hat{\sigma}_{i,i+1}^x, \quad (3.1)$$

donde  $\hat{b}_i^\dagger$  ( $\hat{b}_j$ ) es el operador de creación (aniquilación) bosónica que actúa en el sitio  $i$  ( $j$ ) de la cadena de bosones, y el campo  $\mathbb{Z}_2$  de Ising es descrito por las matrices de Pauli en la dirección  $z$  ( $x$ ) denotadas por  $\hat{\sigma}_{i,i+1}^z$  ( $\hat{\sigma}_{i,i+1}^x$ ) que residen en las ligas de la cadena entre los sitios bosónicos  $i$  e  $i+1$ . El parámetro  $\alpha > 0$  representa la intensidad del salto de los bosones entre sitios vecinos, y  $\beta > 0$  fija la intensidad del campo magnético transversal que induce una dinámica en el campo  $\mathbb{Z}_2$  de espines. Este Hamiltoniano describe que la interacción de salto de bosones entre sitios está totalmente controlada por el sistema de campos  $\mathbb{Z}_2$  de espines, pudiendo cambiar de sitio a sitio, haciendo que la dinámica bosónica del sistema cambie completamente respecto al modelo de Bose–Hubbard. Además, este modelo presenta una simetría local  $\mathbb{Z}_2$ , dado que es invariante ante transformaciones en el campo bosónico dadas por  $\hat{b}'_j \rightarrow e^{i\varphi_j} \hat{b}_j$  y  $\vec{b}'_j \rightarrow e^{-i\varphi_j} \vec{b}_j$  mientras el campo bosónico en dirección  $x$  se mantiene  $\hat{\sigma}'_{j,j+1}^x \rightarrow \hat{\sigma}_{j,j+1}^x$ , pero en la dirección  $z$ , el campo es  $\hat{\sigma}'_{j,j+1}^z \rightarrow e^{i\varphi_j} \hat{\sigma}_{j,j+1}^z e^{-i\varphi_{j+1}}$ , donde  $\varphi_j \in \{0, \pi\}$  es una fase asociada al campo  $\mathbb{Z}_2$  de Ising para el sitio  $j$ .

Adicional a estas interacciones, se consideran términos que rompan la simetría del


 Figura 3.1: Representación gráfica del modelo  $\mathbb{Z}_2$  de Bose–Hubbard.

sistema, tales como el modelo de Bose–Hubbard tradicional,

$$H_{BH} = -t \sum_i \left( \hat{b}_i^\dagger \hat{b}_{i+1} + \text{H.c.} \right) + \frac{U}{2} \sum_i \hat{n}_i (\hat{n}_i - 1) - \mu \sum_i \hat{n}_i, \quad (3.2)$$

donde a la integral de salto controlada por el campo  $\mathbb{Z}_2$  de Ising de la Ec. (3.1) se le agrega un *término de sesgo* fijo  $t$ , además de una interacción de contacto en el sitio bosónico con fuerza  $U$  que mide la interacción entre bosones de un mismo sitio, y  $\mu$  es el potencial químico que controla el número de bosones en la cadena. Finalmente, se introduce el término más simple que rompe la simetría en el campo  $\mathbb{Z}_2$  que modifica la dinámica de espines,

$$H_\Delta = \frac{\Delta}{2} \sum_i \hat{\sigma}_{i,i+1}^z, \quad (3.3)$$

donde  $\Delta$  fija la diferencia de energía entre las configuraciones locales  $|\uparrow_{i,i+1}\rangle$  y  $|\downarrow_{i,i+1}\rangle$  del campo  $\mathbb{Z}_2$  de espines. Con las tres contribuciones de dinámicas Hamiltonianas combinadas, el modelo de  $\mathbb{Z}_2$  Bose–Hubbard resulta descrito por el operador Hamiltoniano,

$$\begin{aligned} H_{\mathbb{Z}_2 BH} &= H_{\mathbb{Z}_2} + H_{HB} + H_\Delta \\ &= -\alpha \sum_i \left( \hat{b}_i^\dagger \hat{\sigma}_{i,i+1}^z \hat{b}_{i+1} + \text{H.c.} \right) + \beta \sum_i \hat{\sigma}_{i,i+1}^x - t \sum_i \left( \hat{b}_i^\dagger \hat{b}_{i+1} + \text{H.c.} \right) \\ &\quad + \frac{U}{2} \sum_i \hat{n}_i (\hat{n}_i - 1) - \mu \sum_i \hat{n}_i + \frac{\Delta}{2} \sum_i \hat{\sigma}_{i,i+1}^z. \end{aligned} \quad (3.4)$$

Este sistema aún no se ha podido resolver analíticamente de manera general. Sin embargo, se ha podido obtener una solución para el estado base en el límite de bosones fuertemente interactuantes [19], introduciendo una transformación de Jordan–Wigner en el *límite de bosones robustos*, como se explora en la siguiente sección.

### 3.1.1. Límite $U \rightarrow \infty$ de bosones robustos

Cuando la interacción en los sitios bosónicos tiende a infinito, es decir,  $U \rightarrow \infty$ , la energía necesaria para tener dos o más bosones en un mismo sitio tiende a infinito, por lo que el sistema sólo permitirá un bosón a lo más por sitio bosónico. Por lo tanto, si además consideramos un potencial químico nulo  $\mu = 0$  en dicho límite, llamado el *límite de bosones robustos*, los bosones se comportan como fermiones.

Para pasar de una dinámica bosónica a una dinámica fermiónica en el límite de bosones robustos, es necesario aplicar una transformación de Jordan–Wigner [50], donde los operadores de creación y aniquilación bosónicos se transforman como,

$$\begin{aligned}\hat{b}_j &\rightarrow e^{i\pi \sum_{k<j} \hat{c}_k^\dagger \hat{c}_k} \hat{c}_j, \\ \hat{b}_j^\dagger &\rightarrow e^{-i\pi \sum_{k<j} \hat{c}_k^\dagger \hat{c}_k} \hat{c}_j^\dagger,\end{aligned}\tag{3.5}$$

por lo que el modelo  $\mathbb{Z}_2$  de Bose–Hubbard en el límite de bosones robustos tiene una dinámica definida por el Hamiltoniano transformado (con  $\mu = 0$ ),

$$\hat{H}_{\mathbb{Z}_2 BH}^{U \rightarrow \infty} = - \sum_i \left[ (t + \alpha \hat{\sigma}_{i,i+1}^z) \hat{c}_i^\dagger \hat{c}_{i+1} + \text{H.c.} \right] + \beta \sum_i \hat{\sigma}_{i,i+1}^x + \frac{\Delta}{2} \sum_i \hat{\sigma}_{i,i+1}^z,\tag{3.6}$$

donde  $\hat{c}_i^\dagger$  ( $\hat{c}_i$ ) es el operador de creación (aniquilación) fermiónico en el sitio fermiónico  $i$  de la cadena.

El modelo descrito por la dinámica del Hamiltoniano de la Ec. (3.6) sigue sin ser resoluble analíticamente, dado que las interacciones se ven fuertemente afectadas por la dinámica de los espines dado el campo magnético transversal externo del segundo término del Hamiltoniano, además de la ruptura de simetría.

**Régimen cuasiadiabático**  $\beta \ll t$ 

Cuando el campo magnético transversal externo es despreciable respecto a la amplitud de salto entre fermiones, es decir, en el límite  $\beta \ll t$ , el cambio en los campos de espín es mucho más lento que los posibles saltos que hay entre los sitios fermiónicos. Esto se puede interpretar como que los campos fermiónicos se adaptan instantáneamente a los espines, y no existe correlación entre los espines y los fermiones, por lo que los espines fungen como un tipo de *potencial estático* visto desde el punto de vista de los campos fermiónicos.

Dadas estas consideraciones, es posible definir una función de onda variacional del estado base para el modelo en el límite cuasiadiabático en el cual las correlaciones cuánticas entre los fermiones y los espines sea nula; es decir, de la forma,

$$|\Psi_{\mathcal{K}}\rangle = \left( \sum_{\mathbf{n}} a_{\mathbf{n}} |\mathbf{n}\rangle_f \right) \otimes \left[ \exp \left( -i \sum_j \frac{\theta_{j,j+1}}{2} \hat{\sigma}_{j,j+1}^y \right) |-\rangle_s \right], \quad (3.7)$$

donde  $\mathcal{K} = \{\mathbf{a}, \boldsymbol{\theta}\}$  es el conjunto de parámetros variacionales, la suma de la parte fermiónica de la función de onda es sobre todo el conjunto computacional del sistema  $\mathbf{n} = (n_1, \dots, n_{n_N})$  en la base del operador número, con  $n_i \in \{0, 1\}$ ,  $i = 1, \dots, n_N$ , y  $|-\rangle_s = \otimes_j (|\uparrow_{j,j+1}\rangle - |\downarrow_{j,j+1}\rangle) / \sqrt{2}$  es un estado auxiliar en el campo de espines. Por lo tanto, la energía variacional en este límite está dada por,

$$\begin{aligned} E_{\mathcal{K}} &= \langle \Psi_{\mathcal{K}} | \hat{H}_{\mathbb{Z}_2 BH}^{U \rightarrow \infty} | \Psi_{\mathcal{K}} \rangle \\ &= - \langle \Psi_{\mathbf{a}}^f | \sum_j (t + \alpha \sin(\theta_{j,j+1})) \hat{c}_j^\dagger \hat{c}_{j+1} | \Psi_{\mathbf{a}}^f \rangle + \frac{\beta}{2} \sum_j \left[ \sin^2 \left( \frac{\theta_{j,j+1}}{2} \right) - \cos^2 \left( \frac{\theta_{j,j+1}}{2} \right) \right] \\ &\quad + \frac{\Delta}{2} \sum_j \cos \left( \frac{\theta_{j,j+1}}{2} \right) \sin \left( \frac{\theta_{j,j+1}}{2} \right), \end{aligned} \quad (3.8)$$

donde  $|\Psi_{\mathbf{a}}^f\rangle = \sum_{\mathbf{n}} a_{\mathbf{n}} |\mathbf{n}\rangle$  es la parte fermiónica de la función de onda variacional en el límite cuasiadiabático.

Un caso especial resulta cuando se dobla la celda de los campos de espín, es decir, cuando se consideran sólo dos parámetros variacionales para los espines:  $\boldsymbol{\theta} = (\theta_A, \theta_B)$ , donde  $\theta_A$  es el parámetro variacional para las ligas pares, y  $\theta_B$  para las impares. De esta

forma, la energía variacional resulta,

$$\begin{aligned}
 E_{\mathcal{K}} = & - \langle \Psi_{\mathbf{a}}^f | \left\{ \sum_j t(\boldsymbol{\theta}) \left[ 1 + (-1)^j \delta(\boldsymbol{\theta}) \right] \hat{c}_j^\dagger \hat{c}_{j+1} + \text{H.c.} \right\} | \Psi_{\mathbf{a}}^f \rangle \\
 & - \frac{L\beta}{2} [\cos(\theta_A) + \cos(\theta_B)] + \frac{L\Delta}{4} [\sin(\theta_A) + \sin(\theta_B)], \tag{3.9}
 \end{aligned}$$

donde los parámetros  $t(\boldsymbol{\theta})$  y  $\delta(\boldsymbol{\theta})$  están dados por,

$$t(\boldsymbol{\theta}) = t + \frac{\alpha}{2} (\sin(\theta_A) + \sin(\theta_B)), \tag{3.10}$$

$$\delta(\boldsymbol{\theta}) = \frac{\alpha(\sin(\theta_A) - \sin(\theta_B))}{2(t + \frac{\alpha}{2}(\sin(\theta_A) + \sin(\theta_B)))}, \tag{3.11}$$

y se identifica al operador  $\hat{H}_H(t, \delta) = \sum_i \left[ -t(1 + (-1)^i \delta) \hat{c}_i^\dagger \hat{c}_{i+1} + \text{H.c.} \right]$  como el Hamiltoniano del modelo de Hückel [51] [52], que se puede interpretar como un tipo de modelo de Fermi–Hubbard con integrales de salto dimerizadas: es decir, que siguen un patrón del tipo  $\{t(1 - \delta), t(1 + \delta), \dots, t(1 + \delta), t(1 - \delta)\}$ . Dicho modelo tiene solución analítica para el estado base, y está dada por,

$$\begin{aligned}
 \varepsilon_{\boldsymbol{\theta}} = & -\frac{2}{\pi} t(\boldsymbol{\theta}) E(1 - \delta^2(\boldsymbol{\theta})) + \frac{\Delta}{2} (\sin(\theta_A) + \sin(\theta_B)) \\
 & - \frac{\beta}{2} (\cos(\theta_A) + \cos(\theta_B)), \tag{3.12}
 \end{aligned}$$

donde la función  $E(x)$  está dada por,

$$E(x) = \int_0^{\pi/2} dk \sqrt{1 - x \sin^2 k}, \tag{3.13}$$

que se identifica como la integral elíptica completa de segundo tipo.

### 3.2. Implementación de máquinas de Boltzmann restringidas

El modelo de RBM que se implementa para la resolución de este sistema cuántico es la modificación de la versión de codificación One–Hot para el modelo de Bose–Hubbard mostrada en la Fig. (3.2), donde la capa de entrada recibe los vectores que codifican el valor esperado del operador número  $\hat{n}_i$  para el sitio  $i$  a través del vector  $\mathbf{n}_1 = (n_1^0, n_1^1, n_1^2, \dots, n_1^L)$ , para  $1 \leq i \leq N_b$ , donde  $n_i^j = \delta_{n_i^j, n_i}$ , y se agregan las neuronas que codifican el estado del campo de espines del modelo. De esta forma, el vector de entrada está codificado de la forma  $\mathcal{S} = (\mathbf{n}_1, \sigma_{1,2}^z, \mathbf{n}_2, \sigma_{2,3}^z, \dots, \sigma_{N_s-1, N_s}^z, \mathbf{n}_{N_b})$ .

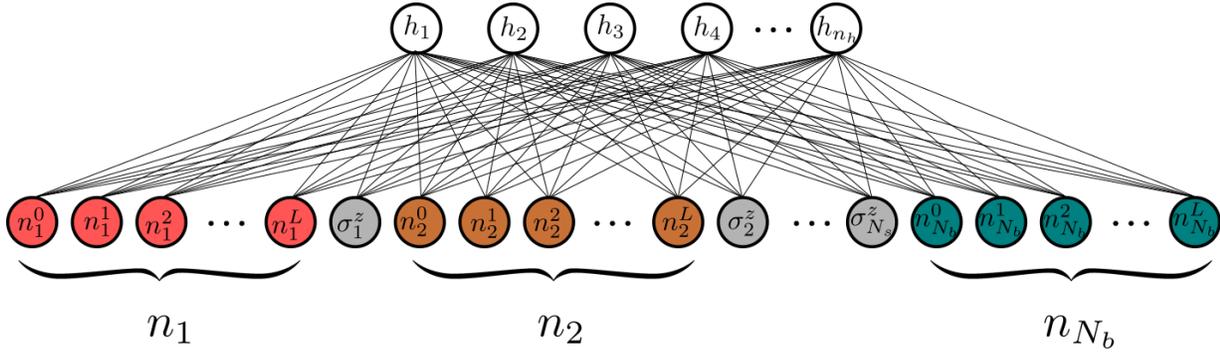


Figura 3.2: Representación gráfica del modelo de RBM para codificar los estados del  $\mathbb{Z}_2$  Bose–Hubbard.

La función de onda variacional modelada por la RBM está dada entonces por,

$$|\Psi_{\mathcal{W}}\rangle = \sum_{\mathcal{S}} \psi_{\mathcal{W}}(\mathcal{S}) |\mathcal{S}\rangle, \quad (3.14)$$

$$\psi_{\mathcal{W}}(\mathcal{S}) = \mathcal{A} \times \prod_{i=1}^{n_h} 2 \cosh \left\{ \sum_{j=1}^{N_s} A_{ij} \sigma_j^z + \sum_{k=0}^L \sum_{l=1}^{N_b} B_{il}^k n_l^k + c_i \right\}, \quad (3.15)$$

$$\mathcal{A} = \exp \left( \sum_{j=1}^{N_s} a_j \sigma_j^z + \sum_{k=0}^L \sum_{l=1}^{N_b} b_l^k n_l^k \right), \quad (3.16)$$

donde  $\mathcal{W} = (\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{A}, \mathbf{B})$  son los parámetros sinápticos de la red neuronal, con  $a_j$  es el término de sesgo correspondiente a la neurona en la capa de entrada que codifica el valor

del sitio de espín  $j$ ,  $\mathbf{b}_l$  es el vector de sesgos de la codificación  $\mathbf{n}_l$  en la entrada,  $c_i$  es el término de sesgo de la neurona  $i$  en la capa de salida,  $\mathbf{A}_{ij}$  es la matriz de conexiones sinápticas que conecta la neurona oculta  $i$  con la neurona visible que codifica el valor del espín en el sitio  $j$ , y  $\mathbf{B}_{il}$  es el vector de conexiones sinápticas entre la neurona oculta  $i$  y el vector de neuronas visibles que codifica  $\mathbf{n}_l$ . De esta forma, se propone utilizar la distribución de probabilidad de la Ec. (3.16) en un algoritmo variacional para encontrar el estado base del modelo  $\mathbb{Z}_2$  Bose–Hubbard.

### 3.3. Resultados

En esta sección se presentan los resultados obtenidos al optimizar el modelo de Bose–Hubbard en distintos escenarios. Primero, se verifica su capacidad de aproximar la energía del estado base, alcanzando precisiones satisfactorias para distintos puntos del espacio de parámetros del sistema. Luego, se pone a prueba respecto a los resultados obtenidos en [19] para el sistema en el límite adiabático y cuasiadiabático, mostrando que el modelo de red neuronal presentado es capaz de capturar la física de los observables lo suficientemente bien como para observar las distintas fases del sistema en dicho límite. Por último, se presentan resultados obtenidos para el número de ocupación bosónica respecto al potencial químico.

#### 3.3.1. Optimización de la energía

A través de un algoritmo de minimización, se optimizan los parámetros sinápticos de la función de onda modelada por una RBM de la forma de la Ec. (3.14) tal que se minimice la energía variacional dada por,

$$E_{\mathcal{W}} = \frac{\langle \Psi_{\mathcal{W}} | \hat{H}_{\mathbb{Z}_2 BH} | \Psi_{\mathcal{W}} \rangle}{\langle \Psi_{\mathcal{W}} | \Psi_{\mathcal{W}} \rangle}. \quad (3.17)$$

Utilizando las técnicas en conjunto del descenso al gradiente y la reconfiguración estocástica, es posible encontrar la función de onda del estado base para el modelo  $\mathbb{Z}_2$  Bose–Hubbard.

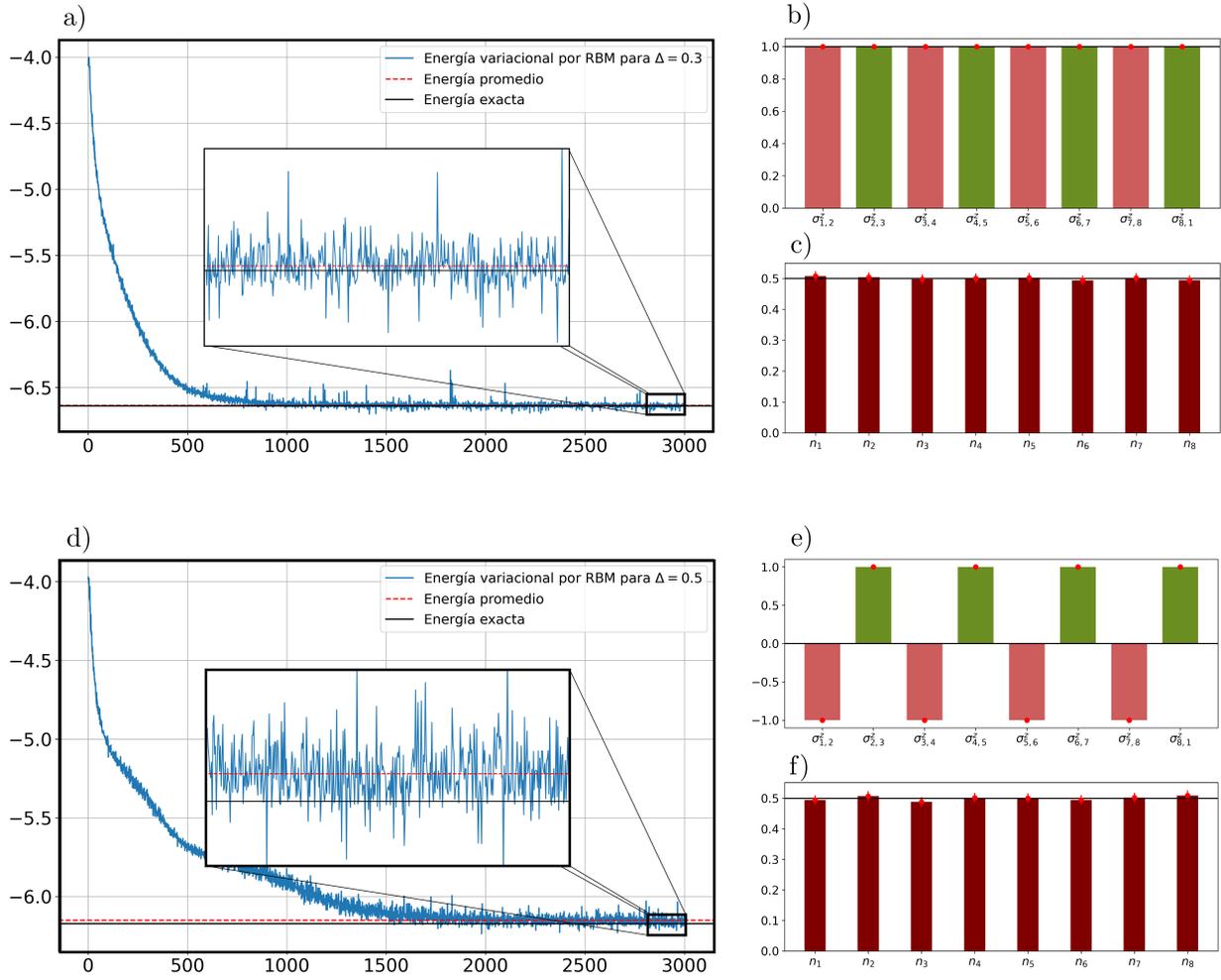


Figura 3.3: Optimización de energía del modelo  $\mathbb{Z}_2$  de Bose–Hubbard en el límite adiabático para a)  $\Delta/t = 0.3$  mostrando el ordenamiento ferromagnético en las ligas de espín en b) y el llenado medio fermiónico en c), y la optimización para d)  $\Delta/t = 0.5$ , con un ordenamiento de Néel en las ligas de espín en e) y el llenado medio fermiónico f).

En la Fig. (3.3) se muestran los resultados obtenidos para dos casos del Hamiltoniano en el límite adiabático. En la Fig. (3.3) se muestra la optimización de la energía para  $\alpha/t = 0.5$  y  $\Delta/t = 0.3$ . Estos parámetros corresponden a la fase ferromagnética, como se muestra en la Fig. (3.3 b)), y hay llenado medio de los campos fermiónicos, como se muestra en la Fig. (3.3 c)). La Fig. (3.3 d) e) f)) muestran la optimización de la energía, el ordenamiento de las ligas de espín y el número promedio de fermiones por sitio del sistema, respectivamente, para  $\alpha/t = 0.5$  y  $\Delta/t = 0.5$ . En Este caso, el sistema se encuentra en la fase antiferromagnética, y presenta un ordenamiento de Néel en las ligas de espines,

mientras que la ocupación fermiónica promedio por sitio sigue siendo de medio llenado.

Como se puede apreciar, la optimización de la energía presenta más varianza cuando el sistema se encuentra en la fase antiferromagnética. Sin embargo, los observables de la magnetización por liga y el número de ocupación fermiónico por sitio presentan barras de error bajas.

### 3.3.2. Diagrama de fase en límite adiabático

En la Fig. (3.4) se muestran los resultados obtenidos para una malla de  $10 \times 10$  en el espacio  $(\alpha/t, \Delta/t)$  en el límite adiabático con  $\beta = 0$  con bosones robustos en  $U \rightarrow \infty$ . Todas las optimizaciones se hicieron con una RBM con  $n_h/n_v = 1$ . En cada paso de optimización se toman 3500 muestras de configuraciones posibles de la base computacional, las cuales se toman las últimas 3000 como cola de las cadenas de Markov.

En la Fig. (3.4 a)) se muestra la magnetización total en el espacio de espines para la malla completa, en la cual se puede apreciar cómo la parte de espines del estado base del sistema tiene una transición de un estado ferromagnético con todos los espines hacia arriba ( $|\uparrow, \uparrow\uparrow \dots \uparrow\rangle_s$ ) a un estado ferromagnético con todos los espines hacia abajo ( $|\downarrow\downarrow\downarrow \dots \downarrow\rangle$ ), pasando por un estado antiferromagnético: es decir, el estado de Néel ( $|\uparrow\downarrow\uparrow\downarrow \dots \downarrow\rangle$  o  $|\downarrow\uparrow\downarrow\uparrow \dots \uparrow\rangle$ ) corroborado por la Fig. (3.4 b)) que muestra los resultados para la magnetización escalonada.

En las Figs. (3.4 c)) y (3.4 d)) se muestra la línea de nivel punteada con mejor definición para  $\alpha/t = 1.0$  para la magnetización total y la magnetización escalonada, respectivamente, dividiendo  $\Delta/t$  en 100 puntos en el intervalo  $(0.0, 1.0)$ . En estas figuras se muestra cómo el modelo de RBM es capaz de observar la ruptura de simetría espontánea en el límite adiabático, mostrando así cómo el cambio de fases es discontinuo en dicho límite.

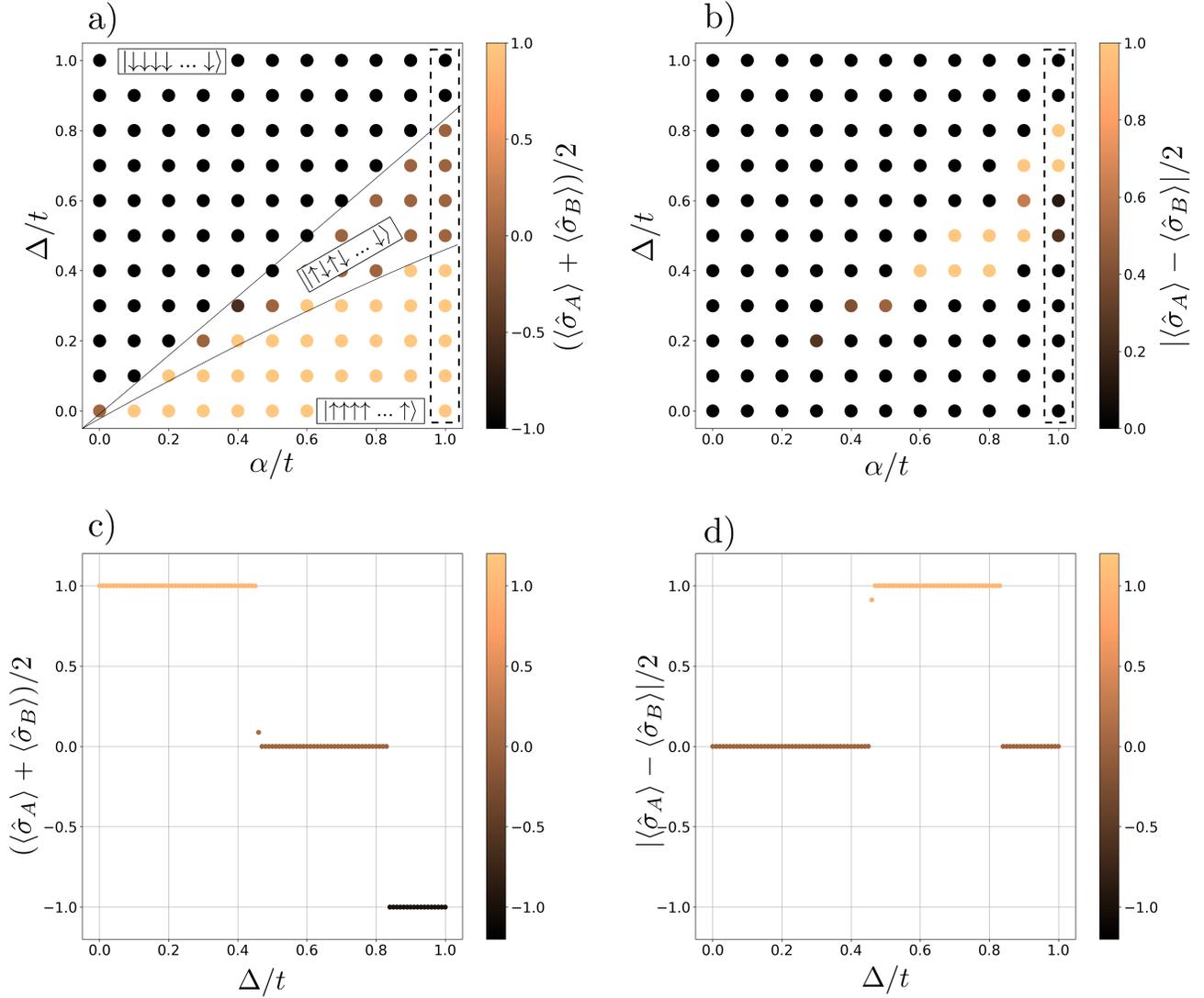


Figura 3.4: Resultados de a) magnetización total y b) magnetización escalonada para una malla en el espacio  $(\alpha/t, \Delta/t)$  con  $\beta/t = 0$  en el límite de bosones robustos. c) y d) muestran una línea de nivel más definida para  $\alpha/t = 1.0$ .

### 3.3.3. Llenado fermiónico respecto al potencial químico

En los resultados presentados anteriormente se dejó al potencial químico nulo  $\mu = 0$ . Sin embargo, haciendo un análisis del llenado fermiónico promedio de los sitios del sistema variando  $\mu$  optimizando la energía variacional de la RBM es posible encontrar otros llenados accesibles.

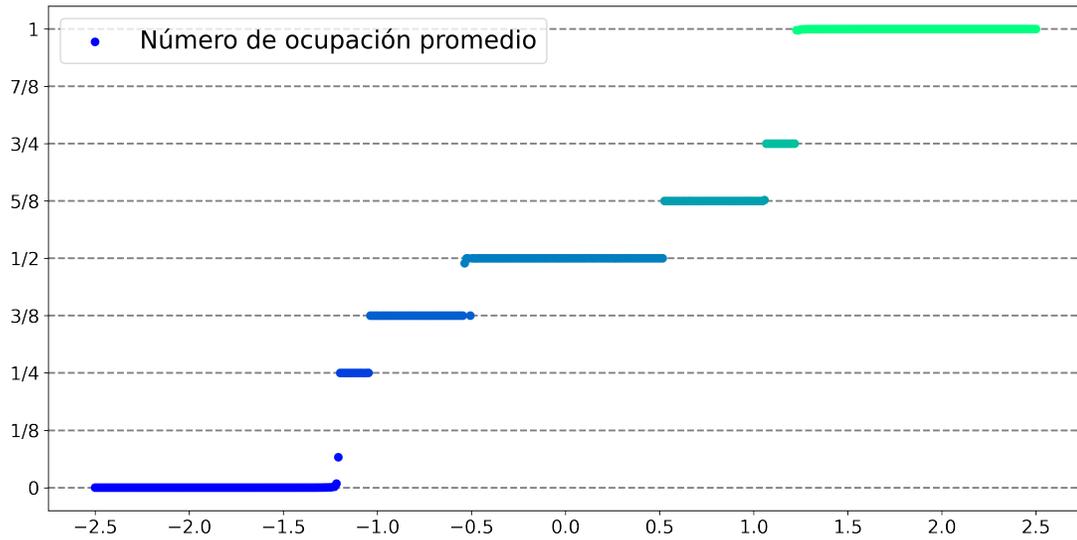


Figura 3.5: Llenado medio de ocupación fermiónica respecto al potencial químico para  $\alpha/t = 0.5$ ,  $\Delta/t = 0.5$  en el límite adiabático con bosones robustos.

En la Fig. (3.5) se muestran los resultados obtenidos para  $\alpha/t = 0.5$  y  $\Delta/t = 0.5$  en el límite adiabático y de bosones robustos de un modelo de  $\mathbb{Z}_2$  Bose-Hubbard con  $N_b = 8$  sitios bosónicos y con condiciones periódicas de frontera. En este caso, modificando  $\mu$  es posible acceder a 9 distintos llenados, dependiendo de los ordenamientos en las ligas de espines. Esto se debe a que el número de sitios bosónicos es 8, pudiendo acceder a cualquier llenado  $k/8$ , con  $0 \leq k \leq 8$ .

# Conclusiones

En el Capítulo 2 se propuso un modelo de red neuronal estocástica como lo son las Máquinas de Boltzmann Restringidas para modelar la función de onda del estado base de sistemas cuánticos de muchos cuerpos, y se mostró su efectividad con dos ejemplos paradigmáticos -el modelo de Ising con campo transversal y el modelo de Bose-Hubbard- comparando los resultados con métodos de diagonalización exacta (diagonalización de Lanczos) y otras funciones de onda variacionales (función de onda de Jastrow).

En el Capítulo 3 se mostró que utilizar modelos de Máquinas de Boltzmann Restringidas para aproximar el estado base del modelo  $\mathbb{Z}_2$  de Bose-Hubbard resulta satisfactorio para el límite adiabático y cuasiadiabático, siendo capaz de obtener un esboce del diagrama de fase con una buena precisión en dichos límites aún para un sistema pequeño de 8 sitios bosónicos.

Aún cuando para este sistema el número de neuronas en la capa de entrada aumenta exponencialmente con el número de sitios bosónicos y con el número máximo de bosones que se permiten por sitio, la simplicidad de las operaciones necesarias para optimizar los parámetros variacionales del modelo neuronal de las Máquinas de Boltzmann Restringidas permite que sea posible la optimización de los sistemas con un gran número de sitios bosónicos. Además, la capacidad de las RBM para capturar las correlaciones físicas del modelo permite explorar otros espacios de parámetros que no se hayan estudiado, como considerar el potencial químico para acceder a distintos llenados fermiónicos del sistema dependiendo del ordenamiento de las ligas de espín.

El siguiente paso de este proyecto de investigación es poner a prueba la precisión de este modelo de red para sistemas más grandes, y realizar estudios de escalabilidad de tiempo y recursos computacionales respecto al tamaño del sistema. También, dado que es posible

acceder a otros llenados respecto al número de sitios bosónicos y el potencial químico, otro resultado interesante es obtener los diagramas de fase del sistema en otro tipo de llenados bosónicos.

Otra parte es buscar obtener resultados para el diagrama de fase del sistema para potenciales de interacción en sitio  $U$  finitas, y observar el tipo de ordenamientos en los sistemas de espín.

También es posible implementar reconfiguración estocástica dependiente del tiempo para capturar la evolución temporal del sistema a través de una función de onda dependiente del tiempo modelada con RBMs.

# Apéndice A

## Modelos de Red

### A.1. Modelo de Bose–Hubbard

El Hamiltoniano del modelo de Bose–Hubbard está dado por,

$$\hat{H}_{BH} = -t \sum_{\langle i,j \rangle} (\hat{b}_i^\dagger \hat{b}_j + H.c.) + \frac{U}{2} \sum_i \hat{n}_i (\hat{n}_i - 1) - \mu \sum_i \hat{n}_i, \quad (\text{A.1})$$

donde las fases a llenados enteros  $\bar{n} \in \mathbb{N}$  están definidos en los límites  $U/t \ll 1$  para la fase de superfluido, y  $t/U \ll 1$  para el aislante de Mott.

#### A.1.1. Teoría de campo medio

Para obtener un poco más de información sobre el modelo, se utilizan los métodos otorgados por teoría de campo medio. Dado que, como se dijo antes, el parámetro de orden de la fase de superfluido es el valor esperado  $\langle \hat{b} \rangle$ , suponemos una aproximación de campo medio en nuestro sistema, dejando que los observables oscilen alrededor de su valor promedio  $\langle \hat{b}_i \rangle \equiv \phi \in \mathbb{C}$  y  $\langle \hat{b}_i^\dagger \rangle \equiv \phi^* \in \mathbb{C}$ . Por lo tanto,

$$\hat{b}_i = \phi + \delta \hat{b}_i, \quad \hat{b}_i^\dagger = \phi^* + \delta \hat{b}_i^\dagger, \quad (\text{A.2})$$

donde  $\delta \hat{b}_i$  y  $\delta \hat{b}_i^\dagger$  representan una variación infinitesimal de los observables  $\hat{b}_i$  y  $\hat{b}_i^\dagger$  alrededor de  $\phi$  y  $\phi^*$ , respectivamente. Por lo tanto, el producto  $\hat{b}_i^\dagger \hat{b}_j$  presente en el Hamiltoniano de

Bose–Hubbard resulta,

$$\hat{b}_i^\dagger \hat{b}_j = \phi^* \hat{b}_i + \phi \hat{b}_j^\dagger - |\phi|^2, \quad (\text{A.3})$$

donde se ha supuesto que el producto  $\delta \hat{b}_i \delta \hat{b}_j^\dagger$  es demasiado pequeño, por lo que se puede eliminar ese término. Por lo tanto, introduciendo esta aproximación de campo medio en el Hamiltoniano de Bose–Hubbard, el Hamiltoniano *efectivo* resulta,

$$\hat{H}_{eff} = -t \sum_{\langle i,j \rangle} \left( \phi^* \hat{b}_i + \phi \hat{b}_i^\dagger - |\phi|^2 \right) - \mu \sum_i \hat{n}_i + \frac{U}{2} \sum_i \hat{n}_i (\hat{n}_i - 1), \quad (\text{A.4})$$

y, dado que la suma sobre  $\langle i, j \rangle$  es sobre primeros vecinos, definiendo  $z$  como el *número de coordinación*, que cuenta cuántos primeros vecinos hay por sitio en la red, el Hamiltoniano efectivo se reduce a,

$$\begin{aligned} \hat{H}_{eff} &= -tz \sum_i \left( \phi^* \hat{b}_i + \phi \hat{b}_i^\dagger \right) + \sum_i \left[ \frac{U}{2} \hat{n}_i (\hat{n}_i - 1) - \mu \hat{n}_i + tz |\phi|^2 \right] \\ &= \sum_i \left[ h_i^{(0)} + V_i^t \right], \end{aligned} \quad (\text{A.5})$$

donde se ha definido  $h_i^{(0)} = (U/2) \hat{n}_i (\hat{n}_i - 1) - \mu \hat{n}_i + tz |\phi|^2$  y  $V_i^t = -tz (\phi^* \hat{b}_i + \phi \hat{b}_i^\dagger)$ .

Como se puede observar, gracias a la aproximación de campo medio, el Hamiltoniano de Bose–Hubbard se redujo a una suma de *operadores locales efectivos*  $\hat{H}_{loc} = h^{(0)} + V^t$ ,  $V^t = -tz (\phi^* \hat{b} + \phi \hat{b}^\dagger)$ . Considerando sólo el término  $h^{(0)}$ , el costo energético necesario para agregar una partícula a sistema es,

$$\epsilon_{n_0+1} - \epsilon_{n_0} = U n_0 - \mu > 0, \quad (\text{A.6})$$

por lo tanto,  $n_0 > \mu/U$ . Ahora, el costo energético de remover una partícula del sistema es,

$$\epsilon_{n_0-1} - \epsilon_{n_0} = U(1 - n_0) + \mu > 0, \quad (\text{A.7})$$

por lo tanto,  $\mu/U > n_0 - 1$ . Entonces,  $n_0 - 1 < \mu/U < n_0$ , por lo que el número de ocupación bosónico en el estado base depende de  $\mu/U$  y están dados por,

$$n_0 = \begin{cases} 0, & \text{con } \mu/U < 0, \\ 1, & \text{con } 0 < \mu/U < 1, \\ \vdots & \\ n, & \text{con } n - 1 < \mu/U < n. \end{cases} \quad (\text{A.8})$$

Por lo tanto, la energía del sistema es mínima cuando el número de ocupación promedio por sitio es  $n_0$ , y está dada por,

$$E_{n_0}^{(0)} = \frac{U}{2}n_0(n_0 - 1) - \mu n_0 + tz|\phi|^2, \quad (\text{A.9})$$

entonces es posible aproximar el estado base del Hamiltoniano local obtenido con la aproximación de campo medio mediante teoría de perturbaciones, donde  $h^{(0)}$  es el sistema sin perturbar, y  $V^t$  es la perturbación.

El  $p$ -ésimo orden de corrección del estado base  $|n_0\rangle$  está dado por,

$$|n_0^{(p)}\rangle = \sum_{m \neq n} \frac{\langle m_0^{(0)} | V^t | n_0^{(p-1)} \rangle}{E_{n_0}^{(0)} - E_{m_0}^{(0)}} - \sum_{i=1}^p E_{n_0}^i \sum_{m \neq n} |m_0^{(0)}\rangle \frac{\langle m_0^{(0)} | n_0^{(p-i)} \rangle}{E_{n_0}^{(0)} - E_{m_0}^{(0)}}, \quad (\text{A.10})$$

mientras que las correcciones de energía están dadas por,

$$E_{n_0}^{(p)} = \langle n_0^{(0)} | V^t | n_0^{(p-1)} \rangle. \quad (\text{A.11})$$

Calculando hasta segundo orden de perturbación, obtenemos que la corrección de energía está dada por,

$$\begin{aligned} E(\phi, \phi^*) &= \frac{U}{2}n_0(n_0 - 1) - \mu n_0 + tz \|\phi\|^2 + (tz)^2 \|\phi\|^2 \left( \frac{n_0 + 1}{\mu - Un_0} + \frac{n_0}{U(n_0 - 1) - \mu} \right) + \mathcal{O}(\|\phi\|^4) \\ &= a_0 + a_2 \|\phi\|^2 + \mathcal{O}(\|\phi\|^4), \end{aligned} \quad (\text{A.12})$$

donde se han definido los parámetros,

$$a_0 = \frac{U}{2}n_0(n_0 - 1) - \mu n_0 \quad (\text{A.13})$$

$$a_2 = tz + (tz)^2 \left( \frac{n_0 + 1}{\mu - Un_0} + \frac{n_0}{U(n_0 - 1) - \mu} \right). \quad (\text{A.14})$$

Ahora, siguiendo el procedimiento usual de Landau para identificar transiciones de fase de segundo orden, dado que  $\phi$  es un parámetro de orden de la fase superfluida del sistema, una transición de fase se da entonces cuando se minimiza  $E(\phi, \phi^*)$ , es decir, cuando,

$$\frac{\partial}{\partial \phi} E(\phi, \phi^*) = a_2 \phi^* + \mathcal{O}(|\phi|^2, \phi^*). \quad (\text{A.15})$$

La solución que minimiza la energía es aquella para la cual la segunda derivada es positiva. Esto depende del parámetro  $a_2$ , dado que,

$$\frac{\partial^2}{\partial \phi \partial \phi^*} E(\phi, \phi^*) = a_2 + \mathcal{O}(|\phi|^2), \quad (\text{A.16})$$

entonces si  $a_2 > 0$ , el parámetro de orden  $\phi$  debe ser cero tal que  $\partial E / \partial \phi = 0$  y eso significa que el sistema está en la fase de superfluido. Ahora, si  $a_2 < 0$ , entonces  $\phi$  debe ser distinto de cero para que  $\partial^2 E / \partial \phi \partial \phi^* > 0$  y  $\partial E / \partial \phi = 0$ , lo que define una fase de aislante de Mott. Entonces, la transición de fase se da en la línea  $a_2 = 0$ , definida por,

$$tz \left[ 1 + tz \left( \frac{n_0 + 1}{\mu - Un_0} + \frac{n_0}{U(n_0 - 1) - \mu} \right) \right] = 0. \quad (\text{A.17})$$

# Apéndice B

## Métodos de Aproximación Variacionales

### B.1. Descenso al gradiente

La idea principal detrás del algoritmo del descenso al gradiente está basada en el hecho de que cualquier función diferencial multi-variable  $F$  valuada en algún punto de su espacio de dominio  $\vec{x}$  en el cual esta función es diferenciable desciende más rápido a lo largo del camino a donde apunta la dirección contraria al gradiente en ese punto  $-\nabla F(\vec{x})$ .

Suponga una función de onda variacional de un sistema cuántico de muchos cuerpos en la base computacional  $\{|x\rangle\}$  dada por  $|\Psi_\alpha\rangle = \sum_x \psi_\alpha(x) |x\rangle$ , donde  $p_\alpha(x) = |\psi_\alpha(x)|^2 / \langle \Psi_\alpha | \Psi_\alpha \rangle$  es la distribución de probabilidad de la función de onda respecto a la base computacional, con  $\psi_\alpha(x) = \langle x | \Psi_\alpha \rangle$  los coeficientes de la expansión sobre la base computacional  $\{|x\rangle\}$ . Sea  $\hat{O}$  un observable de interés del sistema. Entonces, el valor

esperado está dado por la ecuación,

$$\begin{aligned}
 \langle \hat{O} \rangle &= \frac{\langle \Psi_\alpha | \hat{O} | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle} = \frac{\langle \Psi_\alpha | (\sum_{x'} |x'\rangle \langle x'|) \hat{O} | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle} \\
 &= \frac{\sum_{x'} \langle \Psi_\alpha | x' \rangle \frac{\langle x' | \Psi_\alpha \rangle}{\langle x' | \Psi_\alpha \rangle} \langle x' | \hat{O} | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle} \\
 &= \sum_{x'} \frac{|\langle x' | \Psi_\alpha \rangle|^2}{\langle \Psi_\alpha | \Psi_\alpha \rangle} \frac{\langle x' | \hat{O} | \Psi_\alpha \rangle}{\langle x' | \Psi_\alpha \rangle} \\
 &= \sum_{x'} \frac{|\psi_\alpha(x')|^2}{\langle \Psi_\alpha | \Psi_\alpha \rangle} \frac{\langle x' | \hat{O} | \Psi_\alpha \rangle}{\langle x' | \Psi_\alpha \rangle} \\
 &= \sum_{x'} p_\alpha(x') \tilde{O}(x') \\
 &= \mathbb{E}[\tilde{O}], \tag{B.1}
 \end{aligned}$$

donde se ha usado que  $1 = \sum_x |x\rangle \langle x|$ , y se ha definido la función,

$$\tilde{O}(x) = \frac{\langle x | \hat{O} | \Psi_\alpha \rangle}{\langle \Psi_\alpha | \Psi_\alpha \rangle} \tag{B.2}$$

$$\mathbb{E}[\tilde{O}] = \sum_x p_\alpha(x) \tilde{O}(x), \tag{B.3}$$

donde  $\tilde{O}(x)$  es el *estimador local* del observable  $\hat{O}$  respecto al elemento de la base computacional  $x$ . Este se puede reescribir como,

$$\begin{aligned}
 \tilde{O}(x) &= \frac{\langle x | \hat{O} | \Psi_\alpha \rangle}{\langle x | \Psi_\alpha \rangle} = \frac{\langle x | \hat{O} (\sum_{x'} |x'\rangle \langle x'|) | \Psi_\alpha \rangle}{\langle x | \Psi_\alpha \rangle} \\
 &= \sum_{x'} \frac{\langle x' | \Psi_\alpha \rangle}{\langle x | \Psi_\alpha \rangle} \langle x | \hat{O} | x' \rangle \\
 &= \sum_{x'} \frac{\psi_\alpha(x')}{\psi_\alpha(x)} \langle x | \hat{O} | x' \rangle. \tag{B.4}
 \end{aligned}$$

Aún cuando la Ec. (B.4) es una suma sobre el espacio de Hilbert completo, el estimador local puede ser aproximado de manera eficiente para operadores que sean lo suficientemente dispersos en la base computacional dada, es decir, que todos los elementos de matriz  $\langle x | \hat{O} | x' \rangle$  son cero excepto por los que están cerca de la diagonal. Cuando el operador  $\hat{O} = \hat{H}$  es el Hamiltoniano,  $\tilde{H}(x) = E_{\text{loc}}(x)$  es la *energía local* asociada a la configuración

específica  $x$ .

Ahora, se desea minimizar el valor esperado  $\langle \hat{O} \rangle$  respecto al espacio de parámetros variacionales. Es decir, se busca encontrar la función de onda variacional  $|\Psi_\alpha\rangle$  que minimice la energía del sistema. Una forma de hacer esto es usar la forma variacional de la función de onda y obtener el gradiente del observable en el espacio de parámetros, para así actualizar los parámetros a lo largo de la dirección opuesta del gradiente  $\nabla_\alpha \langle \hat{O} \rangle$ . Es decir, se modifican los parámetros variacionales a través de la *regla de optimización*,

$$\alpha \rightarrow \alpha - \eta \nabla_\alpha \langle \hat{O} \rangle, \quad (\text{B.5})$$

donde  $\nabla_\alpha$  denota el gradiente en el espacio de parámetros, y  $\eta$  es el parámetro de *razón de aprendizaje*.

La derivada parcial del observable  $\langle \hat{O} \rangle$  está dada por,

$$\begin{aligned} \frac{\partial}{\partial \alpha_k} \langle \hat{O} \rangle &= \frac{\partial}{\partial \alpha_k} \sum_x \frac{|\psi_\alpha(x)|^2}{\langle \Psi_\alpha | \Psi_\alpha \rangle} \tilde{O}(x) \\ &= \frac{\partial}{\partial \alpha_k} \sum_x \frac{|\psi_\alpha(x)|^2}{\langle \Psi_\alpha | \Psi_\alpha \rangle} \sum_{x'} \frac{\psi_\alpha(x')}{\psi_\alpha(x)} \langle x | \hat{O} | x' \rangle, \end{aligned} \quad (\text{B.6})$$

y definiendo  $\partial_{\alpha_k} = \partial/\partial\alpha_k$ , resulta,

$$\begin{aligned}
 \partial_{\alpha_k} \langle \hat{A} \rangle &= \sum_{x,x'} \langle x | \hat{O} | x' \rangle \partial_{\alpha_k} \left[ \frac{\psi_\alpha(x) \psi_{\alpha^*}^*(x)}{\sum_{x''} \psi_\alpha(x'') \psi_{\alpha^*}^*(x'')} \frac{\psi_\alpha(x')}{\psi_\alpha(x)} \right] \\
 &= \sum_{x,x'} \langle x | \hat{O} | x' \rangle \partial_{\alpha_k} \left[ \frac{\psi_{\alpha^*}^*(x) \psi_\alpha(x')}{\sum_{x''} \psi_\alpha(x'') \psi_{\alpha^*}^*(x'')} \right] \\
 &= \sum_{x,x'} \langle x | \hat{O} | x' \rangle \left\{ \frac{\psi_{\alpha^*}^*(s) [\partial_{\alpha_k} \psi_\alpha(x')]}{\sum_{x''} \psi_\alpha(x'') \psi_{\alpha^*}^*(x'')} - \frac{\sum_{x''} \psi_{\alpha^*}^*(x'') [\partial_{\alpha_k} \psi_\alpha(x'')] \psi_{\alpha^*}^*(x) \psi_\alpha(x')}{[\sum_{x''} \psi_\alpha(x'') \psi_{\alpha^*}^*(x'')]^2} \right\} \\
 &= \sum_x \frac{\psi_{\alpha^*}^*(x) \psi_\alpha(x)}{\langle \Psi_\alpha | \Psi_\alpha \rangle} \left[ \left( \sum_{x''} \frac{\psi_\alpha(x'')}{\psi_\alpha(x)} \langle x | \hat{O} | x'' \rangle \right. \right. \\
 &\quad \left. \left. - \sum_{x'} \frac{\psi_\alpha(x') \psi_{\alpha^*}^*(x')}{\langle \Psi_\alpha | \Psi_\alpha \rangle} \sum_{x''} \frac{\psi_\alpha(x'')}{\psi_\alpha(x)} \langle x'' | \hat{O} | x' \rangle \right) \frac{\partial_{\alpha_k} \psi_\alpha(x)}{\psi_\alpha(x)} \right] \\
 &= \sum_x p_\alpha(x) \left[ \left( \tilde{O}(x) - \sum_{x'} p_\alpha(x') \tilde{O}(x') \right)^* \partial_{\alpha_k} \ln \psi_\alpha(x) \right] \\
 &= \sum_x p(x) \left[ \left( \tilde{O}(x) - \mathbb{E}[\tilde{O}] \right)^* \mathcal{O}_k \right] \\
 &= \mathbb{E}[(\tilde{A} - \mathbb{E}[\tilde{A}])^* \mathcal{O}_k], \tag{B.7}
 \end{aligned}$$

donde se ha supuesto que el operador  $\hat{O}$  es hermitiano en el espacio de Hilbert del sistema, entonces  $\langle x | \hat{O} | x' \rangle = \langle x' | \hat{O} | x \rangle$  y  $\hat{O}^\dagger = \hat{O}$ , y se ha definido a la *derivada logarítmica* sobre el parámetro variacional  $\alpha_k$  de la función de onda variacional como,

$$\mathcal{O}_k(x) \equiv \frac{\partial}{\partial\alpha_k} \ln \psi_\alpha(x), \tag{B.8}$$

respecto a la configuración específica del sistema  $x$ .

### B.1.1. Técnicas de proyección

La principal idea detrás de estas técnicas es proyectar varias veces a lo largo de un operador específico la función de onda que se desea aproximar, de tal forma que todas las contribuciones de los estados excitados se van disminuyendo hasta que sean despreciables, y sólo sea apreciable la contribución del estado base [37].

Sea  $|\Psi_\alpha^0\rangle$  una función de onda inicial del sistema, que puede ser la función ya optimizada

por el principio variacional como en la Ec. (1.18), dada por,

$$\begin{aligned} |\Psi_\alpha^0\rangle &= \sum_j \sum_i \psi_\alpha(x_i) \varphi_j(x_i) |\varphi_j\rangle \\ &= \sum_j \varphi_{\alpha,j}(\{x_i\}) |\varphi_j\rangle, \end{aligned} \quad (\text{B.9})$$

donde  $\varphi_{\alpha,j}(\{x_i\}) = \sum_i \psi_\alpha(x_i) \varphi_j(x_i)$ , y se ha expresado la función de onda variacional en la base de estados propios del Hamiltoniano  $\{|\varphi_i\rangle\}$ , con energías  $E_i$ , y donde  $|\psi_0\rangle$  es el estado base con energía  $E_0$ . La técnica de proyección se puede implementar iterativamente de la forma,

$$|\Psi_\alpha^{n+1}\rangle = (\Lambda - \hat{H}) |\Psi_\alpha^n\rangle, \quad (\text{B.10})$$

donde  $\hat{H}$  es el Hamiltoniano del sistema y  $\Lambda$  es un operador lineal con el número real  $\lambda$  en todas sus entradas diagonales. Entonces, usando la forma de la Ec. (1.17),

$$(\Lambda - \hat{H})^n |\Psi_\alpha^0\rangle = (\lambda - E_0)^n \left[ \varphi_{\alpha,0}(\{x_i\}) |\varphi_0\rangle + \sum_{i \neq 0} \varphi_{\alpha,i}(\{x_i\}) \left( \frac{\lambda - E_i}{\lambda - E_0} \right)^n |\varphi_i\rangle \right], \quad (\text{B.11})$$

y eligiendo  $\lambda$  tal que  $|\lambda - E_i| > |\lambda - E_0|$ , es decir,  $\lambda > (E_{max} + E_0)/2$  donde se define  $E_{max} = \max_i E_i$  es el eigenvalor máximo del Hamiltoniano  $\hat{H}$ . Por lo tanto, la componente del estado base en la Ec. (B.11) crece mucho más rápido, mientras que las componentes de los estados excitados disminuyen conforme  $n \rightarrow \infty$ , dejando sólo la componente  $|\psi_0\rangle$  aparte con un factor  $\varphi_0(\alpha)(\lambda - E_0)^n$ . Por lo tanto, se puede concluir que,

$$\lim_{n \rightarrow \infty} (\Lambda - \hat{H})^n |\Psi_\alpha^0\rangle \propto |\varphi_0\rangle. \quad (\text{B.12})$$

### B.1.2. Reconfiguración estocástica

Sea  $|\Psi_\alpha\rangle = \sum_x \psi_\alpha(x) |x\rangle$  una función de onda de prueba, que depende del conjunto de parámetros variacionales  $\alpha = \{\alpha_1, \alpha_2, \dots\}$ , con  $\psi_\alpha(x) = \langle x | \Psi_\alpha \rangle$  la amplitud de probabilidad respecto al conjunto de estados accesibles  $\{x\}$ . Entonces, para un  $\Lambda$  lo suficientemente

grande, podemos escribir,

$$(\Lambda - \hat{H}) |\Psi_\alpha\rangle = \delta\alpha_0 + \sum_{j \neq 0} \delta\alpha_j \frac{\partial}{\partial \alpha_j} |\Psi_\alpha\rangle + c^\perp |\psi_\alpha^\perp\rangle, \quad (\text{B.13})$$

donde  $\{\delta\alpha_j\}$  son el cambio variacional de los parámetros  $\alpha_j$  y  $|\Psi_\alpha^\perp\rangle$  es un estado en el subespacio ortogonal de la función variacional  $|\Psi_\alpha\rangle$ , con un peso  $c^\perp$ . Ahora, notando que,

$$\begin{aligned} \frac{\partial}{\partial \alpha_j} |\Psi_\alpha\rangle &= \sum_x \frac{\partial}{\partial \alpha_j} \langle x | \Psi_\alpha \rangle |x\rangle \\ &= \sum_x \frac{\partial}{\partial \alpha_j} (\langle x | \Psi_\alpha \rangle) \cdot \frac{\langle x | \Psi_\alpha \rangle}{\langle x | \Psi_\alpha \rangle} |x\rangle \\ &= \sum_x \frac{\partial}{\partial \alpha_j} [\ln(\langle x | \Psi_\alpha \rangle)] \langle x | \Psi_\alpha \rangle |x\rangle \\ &= \sum_x \frac{\partial}{\partial \alpha_j} [\ln \psi_\alpha(x)] \psi_\alpha(x) |\psi\rangle, \end{aligned} \quad (\text{B.14})$$

y definiendo al operador  $\hat{O}_j$  como,

$$\hat{O}_j |x\rangle = \frac{\partial}{\partial \alpha_j} [\ln(\psi_\alpha(x))], \quad (\text{B.15})$$

es posible reescribir la Ec. (B.13) como,

$$(\Lambda - \hat{H}) |\Psi_\alpha\rangle = \sum_{j=0} \delta\alpha_j \hat{O}_j |\Psi_\alpha\rangle + |\Psi_\alpha^\perp\rangle. \quad (\text{B.16})$$

donde  $\hat{O}_0 = \hat{I}$  es el operador unitario. Por como vimos del método de proyección (sección B.1.1), la solución para  $\delta\alpha_j$  de la Ec. (B.16) marca un camino en el espacio de parámetros variacionales por el cual la energía asociada a la función de onda variacional  $|\Psi_\alpha\rangle$  se minimiza. Por lo tanto, nos interesa resolver el sistema de ecuaciones lineales para optimizar los parámetros variacionales.

Primero, haciendo producto interior a la Ec. (B.16) con  $\langle \Psi_\alpha |$  se obtiene,

$$\Lambda - \langle \hat{H} \rangle = \delta\alpha_0 + \sum_{j=1} \delta\alpha_j \langle \hat{O}_j \rangle, \quad (\text{B.17})$$

donde se ha supuesto que la función de onda variacional no está normalizada. Ahora, tomando el producto interior de la Ec. (B.16) con  $\langle \Psi_\alpha | \left( \frac{\partial}{\partial \alpha_j} \right)^\dagger = \langle \Psi_\alpha | \hat{O}_j^\dagger$ , la ecuación resulta,

$$\Lambda \langle \hat{O}_j^\dagger \rangle - \langle \hat{O}_j^\dagger \hat{H} \rangle = \delta\alpha_0 \langle \hat{O}_j^\dagger \rangle + \sum_k \delta\alpha_k \langle \hat{O}_j^\dagger \hat{O}_k \rangle, \quad (\text{B.18})$$

donde se ha dividido por la norma de la función de onda variacional  $|\Psi_\alpha\rangle$  para obtener los valores esperados sobre los operadores  $\hat{O}_j$ . Ahora, resolviendo para el parámetro auxiliar  $\delta\alpha_0$  en las Ecs. (B.17) y (B.18) se obtiene,

$$\langle \hat{H} \rangle \langle \hat{O}_j^\dagger \rangle - \langle \hat{O}_j^\dagger \hat{H} \rangle = \sum_k \delta\alpha_k \left( \langle \hat{O}_j^\dagger \hat{O}_k \rangle - \langle \hat{O}_k \hat{O}_j^\dagger \rangle \right), \quad (\text{B.19})$$

y definiendo la matriz  $\mathbf{S}$  y el vector  $\vec{f}$  con elementos,

$$S_{j,k} = \langle \hat{O}_j^\dagger \hat{O}_k \rangle - \langle \hat{O}_k \hat{O}_j^\dagger \rangle, \quad (\text{B.20})$$

$$f_j = \langle \hat{O}_j^\dagger \hat{H} \rangle - \langle \hat{H} \rangle \langle \hat{O}_j^\dagger \rangle, \quad (\text{B.21})$$

la Ec. (B.19) se reduce a,

$$\sum_k \delta\alpha_k S_{j,k} = -R_j. \quad (\text{B.22})$$

A la matriz  $\mathbf{S}$  se le identifica como la *matriz de covarianza*, y al vector  $\vec{f}$  como el conjunto de *fuerzas generalizadas*. Entonces la Ec. (B.22) puede escribirse como un producto de matrices y vectores de la forma,

$$\delta\vec{\alpha} \cdot \mathbf{S} = -\vec{f},$$

donde  $\delta\vec{\alpha} = \{\delta\alpha_1, \delta\alpha_2, \dots\}$  es el vector en el espacio de parámetros que apunta hacia donde  $\langle \hat{H} \rangle$  se minimiza. Por lo tanto, se tiene  $\delta\vec{\alpha} = -\vec{f} \cdot \mathbf{S}^{-1}$  donde  $\mathbf{S}^{-1}$  denota la inversa de la matriz de covarianza. Entonces, el sistema de ecuaciones se resuelve para los cambios

diferenciales de los parámetros variacionales como,

$$\delta\alpha_j = - \sum_k S_{j,k}^{-1} f_k. \quad (\text{B.23})$$

Note que la matriz de covarianza  $\mathbf{S}$  no siempre es invertible, por lo que la matriz inversa se lleva a cabo mediante la *pseudo-inversa de Moore-Penrose*.

Entonces, la actualización de parámetros variacionales que propone el método de la reconfiguración estocástica para la optimización de la energía variacional  $\langle \hat{H} \rangle$  está dada por,

$$\alpha_j \rightarrow \alpha_j - \Delta t \sum_k S_{j,k}^{-1} f_k, \quad (\text{B.24})$$

donde  $\Delta t$  es un paso pequeño que asegura una mejor precisión en la convergencia, pero que aumenta el número de iteraciones de optimización necesarios.

# Bibliografía

- [1] Eduard Prugovecki. *Quantum mechanics in Hilbert space*. Academic Press, 1982.
- [2] Lev Davidovich Landau, Evgenii Mikhaïlovich Lifshits, and Evgenij Mihajlovič Lifšic. *Mechanics*, volume 1. CUP Archive, 1960.
- [3] Stephen Gasiorowicz. *Quantum physics*. John Wiley & Sons, 2007.
- [4] Timothy Gowers, June Barrow-Green, and Imre Leader. *The Princeton companion to mathematics*. Princeton University Press, 2008.
- [5] Wolfgang Demtröder. *Atoms, molecules and photons*, volume 3. Springer, 2010.
- [6] Robert G Watts. Introduction to perturbation methods. In *Essentials of Applied Mathematics for Engineers and Scientists*, pages 153–162. Springer, 2012.
- [7] Ivar Ekeland. On the variational principle. *Journal of Mathematical Analysis and Applications*, 47(2):324–353, 1974.
- [8] Max Born and W Heisenberg. Zur quantentheorie der molekeln. In *Original Scientific Papers Wissenschaftliche Originalarbeiten*, pages 216–246. Springer, 1985.
- [9] John Clarke Slater. The self consistent field and the structure of atoms. *Physical Review*, 32(3):339, 1928.
- [10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning (adaptive computation and machine learning series). *Cambridge Massachusetts*, pages 321–359, 2017.
- [11] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

- [12] Adnan Khashman. Application of an emotional neural network to facial recognition. *Neural Computing and Applications*, 18(4):309–320, 2009.
- [13] Ioannis Kanellopoulos and Graeme G Wilkinson. Strategies and best practice for neural network image classification. *International Journal of Remote Sensing*, 18(4):711–725, 1997.
- [14] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.
- [15] Gary Marcus, Ernest Davis, and Scott Aaronson. A very preliminary analysis of dall-e 2. *arXiv preprint arXiv:2204.13807*, 2022.
- [16] Sebastian Thrun. Learning to play the game of chess. *Advances in neural information processing systems*, 7, 1994.
- [17] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.
- [18] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [19] Daniel González-Cuadra, Alexandre Dauphin, Przemysław R Grzybowski, Paweł Wójcik, Maciej Lewenstein, and Alejandro Bermudez. Symmetry-breaking topological insulators in the  $z_2$  bose-hubbard model. *Physical Review B*, 99(4):045139, 2019.
- [20] John S Bell. Speakable and unspeakable in quantum mechanics, 1989.
- [21] John Bardeen, Leon N Cooper, and John Robert Schrieffer. Theory of superconductivity. *Physical review*, 108(5):1175, 1957.
- [22] Patrik Fazekas. *Lecture notes on electron correlation and magnetism*, volume 5. World scientific, 1999.

- [23] Rudolf Grimm, Matthias Weidemüller, and Yurii B Ovchinnikov. Optical dipole traps for neutral atoms. In *Advances in atomic, molecular, and optical physics*, volume 42, pages 95–170. Elsevier, 2000.
- [24] Nikolai Borisovich Delone and Vladimir P Krainov. Ac stark shift of atomic energy levels. *Physics-Uspexhi*, 42(7):669, 1999.
- [25] Immanuel Bloch, Jean Dalibard, and Wilhelm Zwerger. Many-body physics with ultracold gases. *Reviews of modern physics*, 80(3):885, 2008.
- [26] RB Stinchcombe. Ising model in a transverse field. i. basic theory. *Journal of Physics C: Solid State Physics*, 6(15):2459, 1973.
- [27] HA Gersch and GC Knollman. Quantum cell model for bosons. *Physical Review*, 129(2):959, 1963.
- [28] Pierre Pfeuty. The one-dimensional ising model with a transverse field. *ANNALS of Physics*, 57(1):79–90, 1970.
- [29] Glen Bigan Mbeng, Angelo Russomanno, and Giuseppe E Santoro. The quantum ising chain for beginners. *arXiv preprint arXiv:2009.09208*, 59, 2020.
- [30] Dieter Jaksch and Peter Zoller. The cold atom hubbard toolbox. *Annals of physics*, 315(1):52–79, 2005.
- [31] Thierry Giamarchi, Christian Rüegg, and Oleg Tchernyshyov. Bose–einstein condensation in magnetic insulators. *Nature Physics*, 4(3):198–204, 2008.
- [32] Dieter Jaksch, Christoph Bruder, Juan Ignacio Cirac, Crispin W Gardiner, and Peter Zoller. Cold bosonic atoms in optical lattices. *Physical Review Letters*, 81(15):3108, 1998.
- [33] Neil W Ashcroft and N David Mermin. *Solid state physics*. Cengage Learning, 2022.
- [34] Gregory H Wannier. The structure of electronic excitation levels in insulating crystals. *Physical Review*, 52(3):191, 1937.

- [35] JK Freericks and H Monien. Phase diagram of the bose-hubbard model. *EPL (Europhysics Letters)*, 26(7):545, 1994.
- [36] Paul M Chaikin, Tom C Lubensky, and Thomas A Witten. *Principles of condensed matter physics*, volume 10. Cambridge university press Cambridge, 1995.
- [37] Federico Becca and Sandro Sorella. *Quantum Monte Carlo approaches for correlated systems*. Cambridge University Press, 2017.
- [38] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.
- [39] Sandro Sorella. Generalized lanczos algorithm for variational quantum monte carlo. *Physical Review B*, 64(2):024512, 2001.
- [40] Sandro Sorella. Green function monte carlo with stochastic reconfiguration. *Physical review letters*, 80(20):4558, 1998.
- [41] Louis Komzsisik. *The Lanczos method: evolution and application*. SIAM, 2003.
- [42] Robert Jastrow. Many-body problem with strong forces. *Physical Review*, 98(5):1479, 1955.
- [43] Paul Smolensky. Information processing in dynamical systems: Foundations of harmony theory. Technical report, Colorado Univ at Boulder Dept of Computer Science, 1986.
- [44] Nicolas Le Roux and Yoshua Bengio. Representational power of restricted boltzmann machines and deep belief networks. *Neural computation*, 20(6):1631–1649, 2008.
- [45] Roger G Melko, Giuseppe Carleo, Juan Carrasquilla, and J Ignacio Cirac. Restricted boltzmann machines in quantum physics. *Nature Physics*, 15(9):887–892, 2019.
- [46] Luciano Loris Viteritti, Francesco Ferrari, and Federico Becca. Accuracy of restricted boltzmann machines for the one-dimensional  $j_{-1} - j_{-2}$  heisenberg model. *SciPost Physics*, 12(5):166, 2022.

- [47] Douglas Hendry and Adrian E Feiguin. Machine learning approach to dynamical properties of quantum many-body systems. *Physical Review B*, 100(24):245123, 2019.
- [48] Kenny Choo, Giuseppe Carleo, Nicolas Regnault, and Titus Neupert. Symmetries and many-body excitations with neural-network quantum states. *Physical review letters*, 121(16):167204, 2018.
- [49] Filippo Vicentini, Damian Hofmann, Attila Szabó, Dian Wu, Christopher Roth, Clemens Giuliani, Gabriel Pescia, Jannes Nys, Vladimir Vargas-Calderon, Nikita Astrakhansev, and Giuseppe Carleo. Netket 3: Machine learning toolbox for many-body quantum systems, 2021.
- [50] Pascual Jordan and Eugene P Wigner. About the pauli exclusion principle. *Z. Phys*, 47(631):14–75, 1928.
- [51] Quantentheoretische Beiträge zum Benzolproblem and Quantentheoretische Beiträge zum Problem. Quantentheoretische beiträge zum benzolproblem. *Z. Phys. Chem*, 70:204–286, 1931.
- [52] William Barford. *Electronic and optical properties of conjugated polymers*. Oxford University Press, 2013.