Alma Mater Studiorum · Università di Bologna

Dipartimento di Fisica e Astronomia "Augusto Righi"
Corso di Laurea in Fisica

Tomografia dello Stato Quantistico

Relatore:
Prof. Tommaso Calarco

Presentata da:
Enrico Barbuio

Correlatore:
Dott.ssa Alessia Camuti Borani

Anno Accademico 2025/2026

Abstract

La tomografia dello stato quantistico, ovvero il processo di ricostruzione dello stato di un sistema a partire da misure ripetute su di esso, è un elemento fondamentale nell'ambito delle tecnologie quantistiche. L'obiettivo di questa tesi è presentare, analizzare e confrontare empiricamente i principali protocolli di tomografia dello stato quantistico per sistemi a variabili discrete. I metodi presi in esame sono l'inversione lineare (LI), la stima di massima verosimiglianza (MLE), un approccio basato sulle reti neurali (NN) e le Classical Shadows (CS). L'introduzione delle ultime due metodologie risponde alla necessità di mitigare il problema della dimensionalità, che rende i metodi tradizionali intrinsecamente inefficienti all'aumentare del numero di componenti.

Nella prima parte del lavoro si introducono i concetti teorici volti alla comprensione del problema, mentre nella seconda si presenta un confronto fra le performance dei metodi di ricostruzione dello stato completo (LI, MLE e NN). Le prestazioni sono valutate e confrontate tramite simulazioni numeriche in ambiente Python su sistemi a due e tre qubit, analizzandone l'accuratezza in funzione del numero di misure eseguite.

Indice

Introduzione

Nell'ambito delle tecnologie quantistiche, è fondamentale poter misurare lo stato di un sistema sia per verificare la correttezza delle operazioni eseguite, sia al termine di protocolli di computazione, di simulazione o di *sensing*. Si pensi come esempio agli algoritmi quantistici come quelli di Shor o di Grover, nei quali il risultato del calcolo è codificato nello stato finale del sistema.

La ricostruzione dello stato quantistico a partire da misure sperimentali è detta *tomografia dello stato quantistico*. Il problema centrale di questo processo deriva dai postulati stessi della meccanica quantistica: una singola misura estrae solo un'informazione parziale sul sistema e ne altera irreversibilmente lo stato, causandone il collasso. Poiché le misure sono intrinsecamente distruttive e stocastiche, la tomografia richiede la preparazione e la misurazione ripetuta di un elevato numero di copie identiche del medesimo stato per poterne ricostruire le proprietà statistiche. A questo si aggiunge la forte criticità legata al problema della dimensionalità: al crescere del numero di componenti del sistema, lo spazio degli stati e, di conseguenza, il numero di parametri da stimare scalano in maniera esponenziale. Preparare e misurare il sistema un numero esponenziale di volte diventa rapidamente proibitivo dal punto di vista sperimentale e computazionale. Nasce quindi la necessità di adottare metodi di ricostruzione efficienti. La scelta del metodo ottimale dipende dalle specifiche caratteristiche che si desidera ricostruire, dal livello di approssimazione tollerato e da informazioni che si hanno a priori sul sistema. I sistemi di interesse di questo lavoro sono quelli a variabili discrete, tra cui rientrano i computer quantistici e i loro elementi costituenti, i qubit.

Per comprendere appieno il problema, il capitolo 1 introdurrà tutte le nozioni di meccanica quantistica necessarie per descrivere un insieme di qubit e le varie operazioni di manipolazione e di misura. Con il capitolo 2 si entrerà nel vivo della tesi, descrivendo i principali metodi di ricostruzione dello stato di un sistema. Si affronteranno due metodi classici: inversione lineare (LI) e stima di massima verosimiglianza (MLE); e due metodi più avanzati: reti neurali (NN) e classical shadows (CS). Nel capitolo 3 si svolgeranno delle simulazioni al computer per verificare l'accuratezza dei primi tre metodi in relazione alle risorse computazionali.

Questa tesi si rivolge ad un lettore con solide basi di algebra lineare, si assumerà la conoscenza del formalismo *Bra-Ket* di Dirac. Seppur utili, non sono necessarie conoscenze di meccanica quantistica.

Capitolo 1

Fondamenti Teorici dell'Informazione Quantistica

1.1 I postulati della meccanica quantistica

In questa prima sezione si vogliono dare alcune nozioni elementari di meccanica quantistica. Non sarà una trattazione completa ma strettamente mirata alla comprensione del problema della tomografia dello stato quantistico.

1.1.1 Stato ed evoluzione del sistema

Si inizia definendo lo stato di un sistema.

Postulato 1. *Ad un sistema fisico isolato è associato uno spazio di Hilbert $\mathcal{H}$ detto* spazio degli stati. *Lo stato del sistema è completamente determinato da un* vettore di stato*, ovvero un vettore di modulo unitario $|\psi\rangle \in \mathcal{H}$.*

Uno spazio di Hilbert $\mathcal{H}$ è uno spazio lineare, finito o infinito-dimensionale, con prodotto interno e completo rispetto alla norma indotta dal prodotto interno.

Per motivi che verranno approfonditi in sezione 1.1.2, due stati che differiscono per una costante di modulo unitario (detta fase globale) sono indistinguibili ed equivalenti:

$$e^{i\varphi} |\psi\rangle \equiv |\psi\rangle , \quad \varphi \in \mathbb{R}. \tag{1.1}$$

Di conseguenza, si può dire che una definizione più precisa di stato quantistico sia la classe di equivalenza dei vettori d'onda uguali a $|\psi\rangle$ a meno di una fase globale.

Questa definizione di stato è estremamente versatile, in quanto si presta alla descrizione di varie tipologie di sistemi: da quelli a più livelli a particelle in campi di potenziale. In questa tesi, e in generale nella teoria dell'informazione quantistica, si è interessati a sistemi in cui $\dim(\mathcal{H}) < \infty$.

Il seguente postulato definisce lo stato di un sistema composto da più sottosistemi. Questa nozione sarà fondamentale in sezione 1.2 per comprendere la struttura matematica con cui descrivere i dispositivi quantistici.

Postulato 2. *Lo spazio degli stati di un sistema fisico composto è il prodotto tensoriale*[1] *degli spazi degli stati dei singoli sistemi componenti. Numerati con $i = 1, \ldots, n$*

i sottosistemi e indicati con $\mathcal{H}_1, \dots \mathcal{H}_n$ i rispettivi spazi degli stati, vale:

$$\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n. \tag{1.2}$$

Inoltre, se i singoli sottosistemi si trovano negli stati $|\psi_1\rangle_1, \dots, |\psi_n\rangle_n$, lo stato complessivo del sistema è lo stato prodotto*:*

$$|\psi\rangle = |\psi_1\rangle_1 \otimes \cdots \otimes |\psi_n\rangle_n . \tag{1.3}$$

Si noti che il postulato non garantisce, in generale, che uno stato qualsiasi possa essere scritto come uno stato prodotto.

Definizione (Stati separabili ed entangled)**.** Preso un vettore di stato $|\psi\rangle_{12} \in \mathcal{H}_{12} = \mathcal{H}_1 \otimes \mathcal{H}_2$, se esistono $|\psi_1\rangle_1 \in \mathcal{H}_1$ e $|\psi_2\rangle_2 \in \mathcal{H}_2$ tali che $|\psi\rangle_{12} = |\psi_1\rangle_1 \otimes |\psi_2\rangle_2$ allora $|\psi\rangle_{12}$ è detto *stato prodotto* o *stato separabile*, altrimenti è uno *stato entangled*.

Definito lo stato come rappresentazione del sistema quantistico, si è ora interessati all'evoluzione del sistema nel tempo.

Postulato 3. *L'evoluzione temporale di un sistema quantistico chiuso è descritta da un operatore unitaro U. Ovvero, indicando con $|\psi(t_1)\rangle$ lo stato al tempo t_1 e con $|\psi(t_2)\rangle$ lo stato al tempo t_2, vale*

$$|\psi(t_2)\rangle = \widehat{U} \, |\psi(t_1)\rangle \tag{1.4}$$

con $\widehat{U}$ operatore unitario dipendente esclusivamente da t_1 e t_2.

Lo studio della dinamica dei sistemi quantistici esula dai fini di questa tesi. Si noti però che l'operatore di evoluzione deve essere invertibile e ciò ha implicazioni sul tipo di operazioni realizzabili all'interno di un computer quantistico.

1.1.2 Osservabili e Misure

Si introdurrà il concetto di misura partendo dalla nozione intuitiva di osservabile introdotta da Dirac [1] per poi passare alla trattazione rigorosa delle Projection Valued Measures (PVM) delineata da Von Neumann [2]. Infine, tale formalismo verrà generalizzato nella definizione moderna di POVM (Positive Operator Valued Measures), seguendo l'impostazione di Nielsen e Chuang [3]. Dal punto di vista fisico, un'*osservabile* rappresenta una grandezza fisica misurabile del sistema. Sul piano matematico ad ogni osservabile corrisponde un operatore $\widehat{A}$ autoaggiunto, definito sullo spazio di Hilbert del sistema. Per autostati dell'osservabile $\widehat{A}$, ovvero $|\psi_j\rangle$ tali che

$$\widehat{A} \, |\psi_j\rangle = a_j \, |\psi_j\rangle , \tag{1.5}$$

la misura di $\widehat{A}$ restituisce con certezza $a_j \in \mathbb{R}$ e lo stato del sistema dopo la misura coincide con quello iniziale. Qundi, in generale, l'esito di una misura dell'osservabile

[1] Si ricorda la definizione di prodotto tensoriale. Presi $\mathcal{H}_A$ e $\mathcal{H}_B$ con le rispettive basi ortonormali: $\{|e_1^A\rangle, \dots, |e_{d_A}^A\rangle\}$ e $\{|e_1^B\rangle, \dots, |e_{d_B}^B\rangle\}$, lo spazio $\mathcal{H}_A \otimes \mathcal{H}_B$ è uno spazio di Hilbert $d_A \cdot d_B$ dimensionale che ha come base $\{|e_i^A\rangle \otimes |e_j^B\rangle \mid i = 1, \dots d_A, j = 1, \dots d_B\}$. Questo spazio è ben diverso da quello dato dal prodotto cartesiano: $\mathcal{H}_A \times \mathcal{H}_b$: la dimensione di quest'ultimo è $d_A + d_B$.

$\widehat{A}$ è un suo autovalore e lo stato dopo la misura è sempre il corrispondente autostato dell'osservabile. L'insieme di tutti i possibili esiti è detto *spettro* dell'osservabile: $\sigma(\widehat{A}) = \{a \mid a \text{ autovalore di } \widehat{A}\}$. Nel caso contrario in cui $|\psi\rangle$ non sia un autostato di $\widehat{A}$, una misura non ha esito certo, ma è possibile fare analisi di carattere probabilistico. Si consideri, a scopo illustrativo, il caso di un'osservabile $\widehat{A}$ con spettro *non degenere*. Ad ogni autovalore a_j è associato un unico autostato $|\psi_j\rangle$ e lo stato $|\psi\rangle$ può essere decomposto nella base $\{|\psi_j\rangle\}_{j=1}^{d}$ di autostati di $\widehat{A}$:

$$|\psi\rangle = \sum_{j=1}^{d} c_j \, |\psi_j\rangle \, . \tag{1.6}$$

Una misura di $\widehat{A}$ su $|\psi\rangle$ restituisce l'autovalore a_j con probabilità $p_j = |c_j|^2$. Tale esito determina il collasso dello stato nel corrispondente autostato $|\psi_j\rangle$. Di conseguenza, una ripetizione della stessa misura restituirà sempre lo stesso esito a_j perché il sistema si trova già in un autostato dell'osservabile. Questa interpretazione probabilistica della funzione d'onda è formalizzata dalla *legge di Born* [4].

Quanto visto nell'esempio può essere interpretato in questo modo: la probabilità di ottenere a_j è data dalla sovrapposizione dello stato iniziale con l'autospazio di a_j e lo stato finale è la proiezione di quello iniziale sull'autospazio dell'autovalore. L'utilizzo dei proiettori[2]permette di rimuovere l'ipotesi di spettro non degenere e dare la definizione di PVM (Projection Valued Measure).

Definizione (PVM)**.** Si consideri un'osservabile descritta dall'operatore autoaggiunto $\widehat{A}$. $\widehat{A}$ ammette la decomposizione spettrale

$$\widehat{A} = \sum_{j} a_j \widehat{P}_j \tag{1.7}$$

dove ogni $\widehat{P}_j$ è il *proiettore* sull'autospazio di a_j, uno degli autovalori distinti di $\widehat{A}$. $\{\widehat{P}_j\}_j$ costituisce un insieme completo di proiettori ortogonali, caratterizzato dalle proprietà:

$$\widehat{P}_i \widehat{P}_j = \delta_{ij} \widehat{P}_i, \quad \sum_{j} \widehat{P}_j = \widehat{1}. \tag{1.8}$$

Una misura PVM restituisce come risultato a_j con probabilità

$$p_j = \langle\psi|\widehat{P}_j|\psi\rangle \tag{1.9}$$

e lo stato del sistema dopo la misura è

$$\frac{\widehat{P}_j \, |\psi\rangle}{\sqrt{\langle\psi|\widehat{P}_j|\psi\rangle}}. \tag{1.10}$$

Data la natura probabilistica della misura è utile definire il *valore d'aspettazione* dell'osservabile $\widehat{A}$. Immaginando di preparare N copie identiche del sistema e di

[2]In generale un proiettore è un operatore hermitiano ($\widehat{P}^\dagger = \widehat{P}$) e idempotente ($\widehat{P}^2 = \widehat{P}$).

misurare su ciascuna l'osservabile, il valore d'aspettazione si calcola come media degli esiti $\{a_i\}_{i=1}^N$:

$$\langle \widehat{A} \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} a_i. \tag{1.11}$$

Ciò equivale a calcolare la media dei singoli elementi dello spettro pesata dalle rispettive probabilità di esito:

$$\langle \widehat{A} \rangle = \sum_j p_j a_j. \tag{1.12}$$

Ricordando eq. (1.9) e eq. (1.7) si ottiene:

$$\langle \widehat{A} \rangle = \sum_j \langle \psi | \widehat{P}_j | \psi \rangle \, a_j = \langle \psi | \left(\sum_j a_j \widehat{P}_j \right) | \psi \rangle = \langle \psi | \widehat{A} | \psi \rangle \,. \tag{1.13}$$

Tralasciando momentaneamente l'associazione con una specifica osservabile, le PVM possono essere definite direttamente a partire da un insieme completo di proiettori ortogonali $\{\widehat{P}_j\}_j$ e di corrispondenti esiti distinti $\{a_j\}_j$. In questo modo, risultano essere un caso particolare della classe più generale di misure definita nel seguente postulato.

Postulato 4. *Le misure quantistiche sono descritte da un insieme di operatori* $\{\widehat{M}_m\}_m$ *che soddisfano la relazione di completezza*

$$\sum_m \widehat{M}_m^\dagger \widehat{M}_m = \widehat{1}. \tag{1.14}$$

Se il sistema è preparato nello stato $|\psi\rangle$ *una misura avrà esito* m *con probabilità*

$$p_m = \langle \psi | \widehat{M}_m^\dagger \widehat{M}_m | \psi \rangle \tag{1.15}$$

e lo stato del sistema dopo la misura sarà

$$\frac{\widehat{M}_m \, |\psi\rangle}{\sqrt{\langle \psi | \widehat{M}_m^\dagger \widehat{M}_m | \psi \rangle}}. \tag{1.16}$$

Si noti che m indicizza i possibili esiti e non il valore numerico del singolo esito, che invece è un parametro sperimentale. Analogamente al caso dei proiettori, la richiesta di completezza del postulato, eq. (1.14), garantisce che le probabilità sommino a 1:

$$\sum_m p_m = \sum_m \langle \psi | \widehat{M}_m^\dagger \widehat{M}_m | \psi \rangle = \langle \psi | (\sum_m \widehat{M}_m^\dagger \widehat{M}_m) | \psi \rangle = \langle \psi | \psi \rangle = 1. \tag{1.17}$$

Nella maggior parte dei casi si è interessati esclusivamente all'esito della misura e non allo stato finale del sistema, in quanto questo viene scartato al termine della misura. Dato che la probabilità dell'esito m dipende solo da $\widehat{M}_m^\dagger \widehat{M}_m$ è sufficiente considerare gli operatori $\widehat{E}_m = \widehat{M}_m^\dagger \widehat{M}_m$. Per costruzione questi sono positivi[3]e

da eq. (1.14), $\{\widehat{E}_m\}_m$ è un insieme completo. Un insieme di operatori con queste caratteristiche prende il nome di POVM (Positive Operator-Valued Measure).

Alla luce di tutto ciò, è possibile giustificare perché vettori di stato che differiscono per una fase globale siano indistinguibili. Presi genericamente gli stati $|\psi\rangle$, $|\psi'\rangle = e^{i\varphi}|\psi\rangle$ e una POVM $\{\widehat{E}_m\}_m$:

$$p'_m = \langle\psi'|\widehat{E}_m|\psi'\rangle = e^{-i\varphi}\langle\psi|\widehat{E}_m|\psi\rangle\, e^{i\varphi} = \langle\psi|\widehat{E}_m|\psi\rangle = p_m. \tag{1.18}$$

Da questo, è possibile concludere che la fase globale tra i due stati non influenza l'esito della misura, risultando dunque irrilevante.

1.1.3 Operatore densità

In questa sezione, verrà introdotto un modo alternativo per definire lo stato di un sistema tramite operatori. Per approfondire i singoli passaggi logici si consiglia di consultare [5].

Stati puri

Gli stati incontrati finora, definiti tramite vettori di stato, sono detti *stati puri*. Dato $|\psi\rangle$, si definisce l'*operatore densità* (o, spesso, *matrice densità*) come l'operatore

$$\widehat{\rho}_\psi = |\psi\rangle\langle\psi|\,. \tag{1.19}$$

Questa costruzione, a prima vista artificiosa, porta con sè diversi vantaggi e sarà di importanza fondamentale per le prossime sezioni.

L'operatore densità soddisfa le seguenti proprietà:

- L'associazione di eq. (1.19) è univoca ma non iniettiva. Infatti, presi $|\psi\rangle$ e $|\psi'\rangle = |\psi\rangle\, e^{i\varphi}$ si ha:

$$\widehat{\rho}_\psi = |\psi\rangle\langle\psi| \tag{1.20}$$

$$\widehat{\rho}_{\psi'} = |\psi'\rangle\langle\psi'| = |\psi\rangle\, e^{i\varphi}\big(|\psi\rangle\, e^{i\varphi}\big)^\dagger = |\psi\rangle\, e^{i\varphi}e^{-i\varphi}\langle\psi| = |\psi\rangle\langle\psi| = \widehat{\rho}_\psi. \tag{1.21}$$

  Ciò significa che l'operatore densità non risente del fattore di fase globale, proprio come lo stato fisico.

- $\widehat{\rho}_\psi$ è un proiettore[4]e, di conseguenza, è definito positivo.

- Ha traccia[5]unitaria:

$$\mathrm{Tr}(\widehat{\rho}_\psi) = 1. \tag{1.22}$$

[3]Un operatore $\widehat{O}$ è positivo se: è hermitiano e $\forall\,|\psi\rangle\,,\ \langle\psi|\widehat{O}|\psi\rangle \geq 0$.

[4]Si veda la definizione nella nota [2]; la verifica è banale.

[5]Si ricorda la definizione di traccia di un operatore: data la base ortonormale $\{|n\rangle\}_n$,

$$\mathrm{Tr}(\widehat{A}) = \sum_n \langle n|\widehat{A}|n\rangle\,. \tag{1.23}$$

Dimostrazione. Il caso di nostro interesse è quello finito-dimensionale. Presa la base ortonormale $\{|n\rangle\}_{n=1,\dots,N}$,

$$\mathrm{Tr}(\widehat{\rho}_\psi) = \sum_{n=1}^{N} \langle n| \left(|\psi\rangle \langle\psi|\right) |n\rangle = \sum_{n=1}^{N} \langle\psi|n\rangle \langle n|\psi\rangle = \langle\psi| \left(\sum_{n=1}^{N} |n\rangle \langle n|\right) |\psi\rangle = 1, \tag{1.24}$$

dove sono state usate la proprietà associativa delle operazioni tra *bra* e *ket* e la relazione di completezza della base $|n\rangle$. □

Inoltre, quanto visto in sezione 1.1.2 può essere riscritto in funzione dell'operatore densità. Seguendo passaggi analoghi alla dimostrazione precedente, si ottiene che il valore d'aspettazione di un operatore $\widehat{A}$ vale:

$$\langle\widehat{A}\rangle = \langle\psi|\widehat{A}|\psi\rangle = \mathrm{Tr}(\widehat{\rho}_\psi \widehat{A}). \tag{1.25}$$

Analogamente, per una POVM $\{\widehat{E}_m\}_m$ eq. (1.15) diventa

$$p_m = \mathrm{Tr}(\widehat{\rho}_\psi \widehat{E}_m). \tag{1.26}$$

Stati misti

Si consideri il caso più generale in cui una procedura sperimentale prepara il sistema in uno di N stati $|\psi_k\rangle$ secondo le probabilità p_k. Ciò è descritto da un *ensemble di stati puri* $\{(p_k, |\psi_k\rangle)\}_{k=1}^{N}$, per il quale si definisce l'operatore densità come:

$$\widehat{\rho} = \sum_{k=1}^{N} p_k |\psi_k\rangle \langle\psi_k| . \tag{1.27}$$

Quando $N = 1$, eq. (1.27) si riduce al caso degli stati puri trattati precedentemente; mentre per $N > 1$, $\widehat{\rho}$ descrive uno *stato misto*. Si noti che $\{p_k\}$ sono probabilità classiche: il sistema non si trova in una sovrapposizione quantistica di $|\psi_k\rangle$, ma è sempre in uno solo dei possibili stati e l'osservatore ignora quale. In questa interpretazione è necessario prestare attenzione, in quanto dato un operatore $\widehat{\rho}$ il set di stati puri che lo genera non è unico. Ci si potrebbe chiedere se sia possibile descrivere uno stato misto tramite un vettore di stato; la risposta è negativa. In primo luogo, le ampiezze che compaiono nei vettori di stato sono numeri complessi la cui fase relativa ha un significato fisico. Volendo esprimere una miscela classica di due stati puri ortogonali $|\psi_1\rangle$ e $|\psi_2\rangle$ con un vettore, sia $(|\psi_1\rangle + |\psi_2\rangle)/\sqrt{2}$ che $(|\psi_1\rangle - |\psi_2\rangle)/\sqrt{2}$ sarebbero rappresentazioni valide, pur essendo stati fisicamente diversi. Un'ulteriore verifica si ottiene osservando che il valore d'aspettazione per una generica osservabile $\widehat{A}$ calcolato tramite eq. (1.13) non fornirebbe l'espressione corretta per un ensemble. Infatti, definendo $|\psi\rangle$ come:

$$|\psi\rangle = \sum_{k=1}^{N} \sqrt{p_k} |\psi_k\rangle \tag{1.28}$$

il valore di aspettazione di $\widehat{A}$ risulterebbe:

$$\begin{aligned}\langle\psi|\widehat{A}|\psi\rangle &= \left(\sum_{k=1}^{N}\sqrt{p_k}\,\langle\psi_k|\right)\widehat{A}\left(\sum_{l=1}^{N}\sqrt{p_l}\,|\psi_l\rangle\right) = \\ &= \sum_{k=1}^{N} p_k\,\langle\psi_k|\widehat{A}|\psi_k\rangle + \sum_{k\neq l}\sqrt{p_k p_l}\,\langle\psi_k|\widehat{A}|\psi_l\rangle \neq \sum_{k=1}^{N} p_k\,\langle\psi_k|\widehat{A}|\psi_k\rangle\end{aligned} \tag{1.29}$$

La presenza dei termini non diagonali nella seconda sommatoria rende il risultato incoerente con la media pesata delle aspettazioni dei singoli stati. Tutte le proprietà dell'operatore densità viste prima restano valide anche nel caso degli stati misti; ma la comodità principale è data dal fatto che continua a valere eq. (1.25) per il calcolo del valore di aspettazione di un'osservabile:

$$\langle\widehat{A}\rangle = \sum_{k=1}^{N} p_k\,\langle\psi_k|\widehat{A}|\psi_k\rangle = \mathrm{Tr}(\widehat{\rho}\widehat{A}) \tag{1.30}$$

e eq. (1.26) che fornisce la probabilità dell'm-esimo esito per la POVM $\{\widehat{E}_m\}_m$

$$p_m = \mathrm{Tr}(\widehat{\rho}\widehat{E}_m). \tag{1.31}$$

Generalizzazione dei postulati della meccanica quantistica

Gli operatori densità permettono di riformulare in maniera equivalente ma più generale i postulati enunciati nelle sezioni 1.1.1 e 1.1.2. Si inizia enunciando in maniera rigorosa quanto detto finora sugli operatori densità.

Definizione (Spazio degli operatori lineari)**.** Preso uno spazio di Hilbert $\mathcal{H}$, $\mathcal{L}(\mathcal{H})$ è lo spazio degli operatori lineari su $\mathcal{H}$:

$$\mathcal{L}(\mathcal{H}) = \{\widehat{O} \mid \widehat{O} : \mathcal{H} \to \mathcal{H} \text{ lineare}\}. \tag{1.32}$$

Esso è a sua volta uno spazio di Hilbert rispetto al prodotto interno di Hilbert-Schmidt $\langle\widehat{A},\widehat{B}\rangle = \mathrm{Tr}(\widehat{A}^\dagger\widehat{B})$.

Teorema (Caratterizzazione dell'operatore densità)**.** *Un operatore $\widehat{\rho} \in \mathcal{L}(\mathcal{H})$ è l'operatore densità di un ensemble di stati quantistici se e solo se:*

1. *ha traccia unitaria:* $\mathrm{Tr}(\widehat{\rho}) = 1$,
2. *è positivo.*

Inoltre ogni operatore densità può essere descritto da un ensemble di stati ortogonali fra loro.

Dimostrazione. Sia $\widehat{\rho}$ l'operatore densità dell'ensemble $\{(p_k, |\psi_k\rangle)\}_{k=1}^N$.

$$\mathrm{Tr}(\widehat{\rho}) = \sum_{k=1}^{N} p_k\,\mathrm{Tr}(|\psi_k\rangle\langle\psi_k|) = \sum_{k=1}^{N} p_k = 1. \tag{1.33}$$

Dato un vettore di stato arbitrario $|\psi\rangle$:

$$\langle\psi|\widehat{\rho}|\psi\rangle = \sum_{k=1}^{N} p_k \langle\psi|\psi_k\rangle \langle\psi_k|\psi\rangle = \sum_{k=1}^{N} p_k |\langle\psi|\psi_k\rangle|^2 \geq 0. \tag{1.34}$$

Sia $\widehat{\rho}$ un operatore positivo. Allora $\widehat{\rho}$ è reale, quindi ammette la decomposizione

$$\widehat{\rho} = \sum_{j=1}^{d} \lambda_j |j\rangle\langle j|, \tag{1.35}$$

con $\lambda_j \geq 0$ e $|j\rangle$ ortogonali. La condizione di traccia unitaria implica

$$1 = \mathrm{Tr}(\widehat{\rho}) = \sum_{j=1}^{d} \lambda_j. \tag{1.36}$$

Le λ_j possono essere interpretate come probabilità, $\widehat{\rho}$ è l'operatore dell'ensemble $\{(\lambda_j, |j\rangle)\}_{j=1}^{d}$. Quest'ultima affermazione prova anche che per ogni operatore densità, esiste un ensemble di stati ortogonali che lo genera. □

Questo teorema permette di dare una definizione rigorosa dello spazio degli operatori densità.

Definizione (Spazio degli operatori densità)**.** $\mathcal{S}(\mathcal{H}) \subset \mathcal{L}(\mathcal{H})$ è lo spazio degli operatori densità, definito come l'insieme degli operatori positivi a traccia unitaria:

$$\mathcal{S}(\mathcal{H}) = \{\widehat{\rho} \in \mathcal{L}(\mathcal{H}) \mid \widehat{\rho} \geq 0,\ \mathrm{Tr}(\widehat{\rho}) = 1\}. \tag{1.37}$$

Si mostra ora un utile criterio per la distinzione degli stati puri dagli stati misti.

Teorema. *Sia $\widehat{\rho}$ un operatore densità, allora*

$$\mathrm{Tr}(\widehat{\rho}^2) \leq 1, \tag{1.38}$$

inoltre,

$$\mathrm{Tr}(\widehat{\rho}^2) = 1 \tag{1.39}$$

se e solo $\widehat{\rho}$ rappresenta uno stato puro*.*

Dimostrazione. Per quanto dimostrato prima, ogni $\widehat{\rho}$ operatore densità può essere descritto da un ensemble di stati $|j\rangle$ ortogonali

$$\widehat{\rho} = \sum_{j=1}^{d} p_j |j\rangle\langle j| \quad \text{con} \quad \langle j|j'\rangle = \delta_{j,j'}. \tag{1.40}$$

Quindi

$$\mathrm{Tr}(\widehat{\rho}^2) = \mathrm{Tr}(\sum_{j,j'=1}^{d} p_j p_{j'} |j\rangle \langle j|j'\rangle \langle j'|) = \mathrm{Tr}(\sum_{j,j'=1}^{d} p_j^2 |j\rangle\langle j|) \leq \mathrm{Tr}(\sum_{j,j'=1}^{d} p_j |j\rangle\langle j|) = 1. \tag{1.41}$$

$\mathrm{Tr}(\widehat{\rho}^2) = 1$ se e solo se

$$\sum_{j=1}^{d} p_j^2 = \sum_{j=1}^{d} p_j, \tag{1.42}$$

la condizione si verifica solo quando un solo p_j vale 1 e gli altri 0, ovvero l'ensemble è composto da un solo stato. □

Si hanno ora tutti gli strumenti per rivisitare i vecchi postulati nel formalismo degli operatori densità.

Postulato 1. *Ad un sistema fisico isolato è associato uno* spazio degli stati *di operatori sullo spazio di Hilbert $\mathcal{H}$: $\mathcal{S}(\mathcal{H})$. Lo stato del sistema è completamente determinato dall'*operatore densità $\widehat{\rho} \in \mathcal{S}(\mathcal{H})$. *Se il sistema si trova nello stato $\widehat{\rho}_k$ con probabilità p_k, l'operatore densità che descrive il sistema è $\widehat{\rho} = \sum_k p_k \widehat{\rho}_k$.*

Postulato 2. *L'evoluzione temporale di un sistema quantistico chiuso è descritta da un operatore unitaro $\widehat{U}$. Ovvero, indicando con $\widehat{\rho}(t_1)$ lo stato al tempo t_1 e con $\widehat{\rho}(t_2)$ lo stato al tempo t_2, vale*

$$\widehat{\rho}(t_2) = \widehat{U}\widehat{\rho}(t_1)\widehat{U}^\dagger \tag{1.43}$$

con $\widehat{U}$ operatore unitario dipendente esclusivamente da t_1 e t_2.

Postulato 3. *Lo spazio degli stati di un sistema fisico composto è il prodotto tensoriale degli spazi degli stati dei singoli sistemi componenti. Numerati con $i = 1, \ldots, n$ i sottosistemi, se l'i-esimo sistema si trova nello stato $\widehat{\rho}_i$ lo stato complessivo del sistema è:*

$$\widehat{\rho} = \widehat{\rho}_1 \otimes \cdots \otimes \widehat{\rho}_n. \tag{1.44}$$

Postulato 4. *Le misure quantistiche sono descritte da un insieme di operatori $\{\widehat{M}_m\}_m \subset \mathcal{L}(\mathcal{H})$ che soddisfa la relazione di completezza*

$$\sum_m \widehat{M}_m^\dagger \widehat{M}_m = \widehat{1}. \tag{1.45}$$

dove m indicizza i possibili esiti della misura. Se il sistema è preparato nello stato $\widehat{\rho}$ una misura avrà esito m con probabilità

$$p_m = \mathrm{Tr}(\widehat{M}_m^\dagger \widehat{M}_m \widehat{\rho}) \tag{1.46}$$

e lo stato del sistema dopo la misura sarà

$$\frac{\widehat{M}_m \widehat{\rho} \widehat{M}_m^\dagger}{\mathrm{Tr}(\widehat{M}_m^\dagger \widehat{M}_m \widehat{\rho})}. \tag{1.47}$$

In conclusione, il formalismo delle misure può essere completato con la seguente definizione, fondamentale nell'ambito della tomografia.

Definizione. $\mathcal{P}$ è un *insieme tomograficamente completo* se per qualsiasi coppia di stati $\widehat{\rho} \neq \widehat{\sigma}$ esiste un operatore $\widehat{P}$ in $\mathcal{P}$, il cui valore d'aspettazione sia diverso per i due stati:

$$\mathrm{Tr}(\widehat{\rho}\widehat{P}) \neq \mathrm{Tr}(\widehat{\sigma}\widehat{P}). \tag{1.48}$$

Se l'insieme $\mathcal{P}$ tomograficamente completo è una POVM allora questa viene detta *IC-POVM (POVM informazionalmente completa).*

1.1.4 Notazione e isomorfismo tra $\mathcal{H}$ e $\mathbb{C}^{2^n}$

Formalmente, per un sistema di n qubit, gli stati puri vivono nello spazio di Hilbert astratto $\mathcal{H}$. Nella teoria dell'informazione quantistica si fissa un isomorfismo che identifica $\mathcal{H}$ con lo spazio dei vettori colonna complessi di dimensione $d = 2^n$, $\mathbb{C}^{2^n}$. Di conseguenza, lo spazio degli operatori lineari $\mathcal{L}(\mathcal{H})$ viene identificato con l'algebra delle matrici quadrate $\mathbb{C}^{2^n \times 2^n}$. Seppur spesso trattati come sinonimi, in questo testo si manterrà la distinzione formale tra un operatore astratto e la sua rappresentazione matriciale, indicando il primo con il simbolo di cappellino ($\widehat{\ }$).

Per completezza si elencano le varie operazioni dello spazio di Hilbert con la loro controparte nello spazio $\mathbb{C}^d$. I *ket* di stato sono rappresentati da vettori colonna complessi

$$|\psi\rangle \longleftrightarrow \boldsymbol{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_d \end{pmatrix}. \tag{1.49}$$

L'operazione di aggiunzione tra *bra* e *ket* è equivalente alla trasposizione seguita da coniugazione complessa

$$\langle\psi| = (|\psi\rangle)^\dagger \longleftrightarrow \boldsymbol{z}^\dagger = \begin{pmatrix} z_1^* & \cdots & z_d^* \end{pmatrix}. \tag{1.50}$$

Il prodotto *bra-ket* equivale al prodotto scalare, calcolato come prodotto righe per colonne

$$\langle\psi|\phi\rangle \longleftrightarrow \boldsymbol{z}^\dagger \boldsymbol{w} = \begin{pmatrix} z_1^* & \cdots & z_d^* \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_d \end{pmatrix}. \tag{1.51}$$

Il prodotto *ket-bra* è il prodotto esterno di $\mathbb{C}^d$

$$|\psi\rangle\langle\phi| \longleftrightarrow \boldsymbol{z}\boldsymbol{w}^\dagger = \begin{pmatrix} z_1 \\ \vdots \\ z_d \end{pmatrix} \begin{pmatrix} w_1^* & \cdots & w_d^* \end{pmatrix} = \begin{pmatrix} z_1 w_1 & z_1 w_2 & \cdots & z_1 w_d \\ z_2 w_1 & z_2 w_2 & \cdots & z_2 w_d \\ \vdots & \vdots & \ddots & \vdots \\ z_d w_1 & z_d w_2 & \cdots & z_d w_d \end{pmatrix}. \tag{1.52}$$

Identificata una base $\{|n\rangle\}$ con la base ortonormale canonica $\{\boldsymbol{e}_n\}$, gli operatori sono rappresentati da matrici:

$$\widehat{A} \longleftrightarrow A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{dd} \end{pmatrix} \tag{1.53}$$

dove gli elementi di matrice sono $a_{ij} = \langle i|\,\widehat{A}\,|j\rangle$. Il prodotto tra operatori lineari è il prodotto righe per colonne fra matrici. Il prodotto tensoriale si traduce in prodotto

di Kronecker

$$|\psi\rangle \otimes |\phi\rangle \;\longleftrightarrow\; \boldsymbol{z} \otimes \boldsymbol{w} = \begin{pmatrix} z_1 \boldsymbol{w} \\ \vdots \\ z_d \boldsymbol{w} \end{pmatrix} = \begin{pmatrix} z_1 w_1 \\ \vdots \\ z_1 w_d \\ \vdots \\ z_d w_1 \\ \vdots \\ z_d w_d \end{pmatrix} \tag{1.54}$$

$$\widehat{A} \otimes \widehat{B} \;\longleftrightarrow\; A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1d}B \\ a_{21}B & a_{22}B & \cdots & a_{2d}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{d1}B & a_{d2}B & \cdots & a_{dd}B \end{pmatrix}. \tag{1.55}$$

1.2 Il computer quantistico

Alla base dell'architettura di un computer classico risiede il *bit*: una variabile booleana che può assumere i due valori 0 o 1. In completa analogia, l'unità fondamentale della computazione quantistica è il *qubit*, abbreviazione di quantum bit, che è un sistema quantistico a due livelli. Classicamente i processi computazionali vengono eseguiti tramite porte logiche classiche, che trovano la loro controparte nelle *porte logiche quantistiche* o *gate*. Lo stato di un computer classico è sempre ben determinato, può essere copiato, cancellato e misurato. Per un computer quantistico la situazione non è così semplice: come visto nei postulati precedenti, ad essere determinate sono proprietà statistiche del sistema, le quali sono misurabili solo parzialmente per via del collasso della funzione d'onda. Questo fatto rende molto più complessa la realizzazione di operazioni su qubit e l'estrazione di informazioni da essi.

Si tratterà il qubit in maniera puramente astratta, ma è bene introdurre brevemente la sua realizzazione fisica. Qualsiasi sistema fisico a due livelli, sufficientemente isolato e controllabile, può fungere da qubit. In pratica, sono state proposte e realizzate diverse piattaforme sperimentali per l'implementazione di computer quantistici, ciascuna con vantaggi e svantaggi in termini di scalabilità, tempi di coerenza, fedeltà delle operazioni e facilità di lettura dello stato. Tra le architetture più mature si trovano i *qubit superconduttori* [6], in cui l'informazione è codificata nei livelli energetici di un circuito elettrico non lineare (giunzione Josephson) raffreddato a temperature criogeniche; gli *ioni intrappolati* [7], in cui il qubit è realizzato tramite due livelli elettronici (o iperfini) di ioni confinati da campi elettromagnetici e manipolati otticamente; e i *qubit fotonici* [8], in cui l'informazione è codificata in gradi di libertà della luce, per esempio la polarizzazione di singoli fotoni. Fra altri approcci si trovano: atomi neutri intrappolati otticamente [9] e spin nucleari o elettronici in stato solido (ad esempio i centri di vacanza dell'azoto nel diamante [10]).

Ognuna di queste implementazioni richiede protocolli specifici di preparazione, manipolazione e misura dello stato, con rumore e canali di errore caratteristici della piattaforma. In questa tesi si studiano le tecniche di tomografia dello stato quanti-

stico a livello del formalismo matematico, la trattazione che segue sarà volutamente *agnostica* rispetto alla piattaforma fisica: si assumerà semplicemente di poter preparare, manipolare e misurare qubit nella base computazionale, senza approfondire ulteriormente i dettagli implementativi delle singole architetture.

1.2.1 Qubit singolo e visualizzazione

Il qubit è un sistema quantistico a due livelli, descritto da un vettore d'onda normalizzato $|\psi\rangle$ appartenente ad uno spazio di Hilbert bidimensionale $\mathcal{H}$. Dalla linearità di $\mathcal{H}$ segue la principale differenza tra bit classico e qubit quantistico: un bit può assumere un singolo stato tra 0 e 1; un qubit può trovarsi nello stato $|0\rangle$, $|1\rangle$ e in una qualsiasi combinazione lineare dei due. Convenzionalmente vengono indicati con $|0\rangle$ e $|1\rangle$ i vettori ortonormali che formano la base computazionale. In questo modo si può scrivere lo stato del qubit come

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \tag{1.56}$$

dove $\alpha, \beta \in \mathbb{C}$ sono i coefficienti, o i "pesi", degli elementi di base, tali per cui $|\alpha|^2 + |\beta|^2 = 1$.

Lo stato scritto come in eq. 1.56 necessita di 4 variabili reali. Raccogliendo un fattore di fase globale e tenendo conto della condizione di normalizzazione i gradi di libertà scendono a due. Graficamente, questo significa restringere lo spazio degli stati ad una sfera di raggio unitario, detta *sfera di Bloch*. Impiegando la seguente parametrizzazione

$$|\psi\rangle = \cancel{e^{i\gamma}}(\cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\varphi}\sin\left(\frac{\theta}{2}\right)|1\rangle), \tag{1.57}$$

con $\theta \in [0, \pi]$ e $\varphi \in [0, 2\pi]$, e identificando θ con l'angolo polare e φ con l'angolo azimutale, si giunge alla rappresentazione di figura 1.1.

La matrice densità per un sistema a due livelli può essere parametrizzata nel seguente modo

$$\widehat{\rho} = \frac{\widehat{1} + \mathbf{r}\cdot\widehat{\boldsymbol{\sigma}}}{2} \tag{1.58}$$

con $\mathbf{r} = (r_x, r_y, r_z)$, $|\mathbf{r}| \leq 1$ e $\widehat{\boldsymbol{\sigma}} = (\widehat{\sigma}_x, \widehat{\sigma}_y, \widehat{\sigma}_z)$ vettore di operatori di Pauli. Si ricorda che, fissando $|0\rangle = \begin{pmatrix}1\\0\end{pmatrix}$, $|1\rangle = \begin{pmatrix}0\\1\end{pmatrix}$, gli operatori di Pauli sono rappresentati dalle matrici:

$$\widehat{\sigma}_x = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \quad \widehat{\sigma}_y = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, \quad \widehat{\sigma}_z = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \tag{1.59}$$

Dimostrazione. $\widehat{\rho}$ è hermitiana per cui

$$\widehat{\rho} = \begin{pmatrix}a & b+ic\\ b-ic & d\end{pmatrix} = \frac{a+d}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} + b\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} + c\begin{pmatrix}0 & i\\ -i & 0\end{pmatrix} + \frac{a-d}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \tag{1.60}$$

con $a, b, c, d \in \mathbb{R}$. Si identificano

$$r_x = 2b, \quad r_y = -2c, \quad r_z = a - d, \tag{1.61}$$

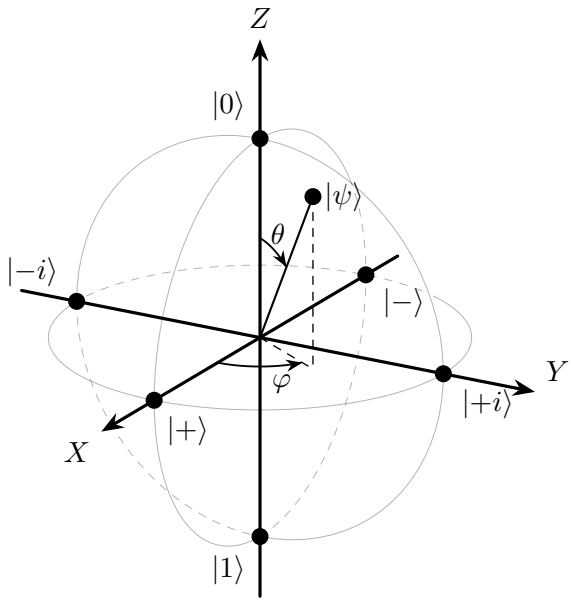

Figura 1.1: Sfera di Bloch. Sono evidenziati gli stati delle basi: $X:\ \{|+\rangle, |-\rangle\}$. $Y:\ \{|+i\rangle, |-i\rangle\}$, $Z:\ \{|0\rangle, |1\rangle\}$, È rappresentato lo stato con $\theta = 30°$, $\varphi = 60°$.

e dalla condizione di traccia unitaria segue

$$\frac{a+d}{2} = \frac{1}{2}, \tag{1.62}$$

in quanto $\widehat{\sigma}_x, \widehat{\sigma}_y, \widehat{\sigma}_z$ hanno traccia nulla. Per cui

$$\widehat{\rho} = \frac{1}{2}\begin{pmatrix} 1+r_z & r_x - ir_y \\ r_x + ir_y & 1-r_z \end{pmatrix}. \tag{1.63}$$

La condizione di positività di $\widehat{\rho}$ equivale alla positività dei suoi autovalori. L'equazione caratteristica è risolta da:

$$\lambda_\pm = \frac{1 \pm \sqrt{r_x^2 + r_y^2 + r_z^2}}{2} = \frac{1 \pm |\mathbf{r}|}{2}. \tag{1.64}$$

$\lambda_+ \geq 0$ sempre, mentre $\lambda_- \geq 0$ se e solo se $|\mathbf{r}| \leq 1$. □

Lo stato è quindi determinato da $\mathbf{r}$, in base ai valori che questo assume si distinguono i casi:

- $|\mathbf{r}| = 1$, stato puro;
- $|\mathbf{r}| < 1$, stato misto;
- $|\mathbf{r}| = 0$, stato massimalmente misto.

Rappresentando $\mathbf{r}$ nel piano cartesiano si ottiene la generalizzazione della sfera di Bloch agli stati misti, come mostrato in fig. 1.2.

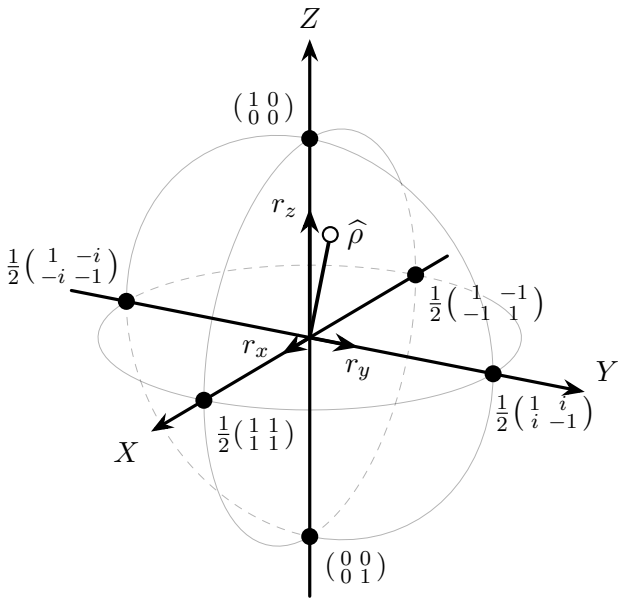

Figura 1.2: Sfera di Bloch nel formalismo degli stati misti. Matrici densità esplicitate nella base $|0\rangle = \begin{pmatrix}1\\0\end{pmatrix}$, $|1\rangle = \begin{pmatrix}0\\1\end{pmatrix}$.

1.2.2 Sistemi a più qubit

Come visto in sezione 1.1.1, per sistemi composti lo spazio degli stati corrisponde al prodotto tensoriale degli spazi degli stati dei singoli sottosistemi. Per esempio, per un sistema di due qubit la base computazionale è

$$\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}, \tag{1.65}$$

con $|ij\rangle = |i\rangle_1 \otimes |j\rangle_2$.

Come si è visto in precedenza, non tutti gli stati di più qubit sono stati separabili. Esistono, infatti, gli stati entangled, tra cui si riportano come esempio i componenti della base di Bell:

$$\left|\Phi^+\right\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}},\ \left|\Psi^+\right\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}},\ \left|\Phi^-\right\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}},\ \left|\Psi^-\right\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}. \tag{1.66}$$

1.2.3 Misura di qubit

Come discusso in sezione 1.1.2, la misura di un sistema quantistico è descritta in generale da una POVM. Nel contesto della computazione quantistica, per convenzione, si assume che la misura fisica disponibile sui qubit sia sempre la *misura in base computazionale*, ovvero la PVM data dai proiettori sugli stati $|0\rangle$ e $|1\rangle$:

$$\mathcal{P}_Z = \{|0\rangle\langle 0|,\ |1\rangle\langle 1|\}. \tag{1.67}$$

Per un sistema di n qubit, la misura in base computazionale si esegue qubit per qubit (o simultaneamente su tutti i qubit) e restituisce una stringa di bit $b \in \{0,1\}^n$,

con probabilità data dalla regola di Born eq. (1.46) applicata al proiettore $|b\rangle\langle b| = |b_1\rangle\langle b_1| \otimes \cdots \otimes |b_n\rangle\langle b_n|$.

Questa scelta non è restrittiva quanto potrebbe sembrare. Se si è interessati a misurare un'osservabile $\widehat{A}$ diagonale in una base diversa da quella computazionale, ad esempio $\widehat{\sigma}_x$, è sufficiente applicare al sistema, prima della misura, la rotazione unitaria $\widehat{U}$ che mappa la base degli autostati di $\widehat{A}$ nella base computazionale, e quindi eseguire la misura standard. Formalmente, se $\widehat{A} = \widehat{U}^\dagger \widehat{Z} \widehat{U}$ con $\widehat{Z}$ diagonale nella base computazionale, misurare $\widehat{A}$ su $\widehat{\rho}$ equivale a misurare $\widehat{Z}$ su $\widehat{U}\widehat{\rho}\widehat{U}^\dagger$:

$$\mathrm{Tr}[\widehat{A}\widehat{\rho}] = \mathrm{Tr}[\widehat{U}^\dagger \widehat{Z} \widehat{U} \widehat{\rho}] = \mathrm{Tr}[\widehat{Z}\, \widehat{U}\widehat{\rho}\widehat{U}^\dagger]. \tag{1.68}$$

In questo modo, ogni misura in una base arbitraria si riconduce a una rotazione seguita da una misura in base computazionale: questo è precisamente il meccanismo su cui si basa il protocollo delle *Classical Shadows*, discusso in sezione 2.5.

Nella pratica sperimentale, la misura di un singolo qubit alla volta è generalmente distruttiva: dopo la misura lo stato collassa nell'autostato corrispondente all'esito ottenuto, per cui la ricostruzione di più proprietà dello stesso stato richiede la preparazione di copie multiple del sistema.

Un'ultima precisazione riguarda la relazione tra la misura di un operatore e di un proiettore. Misurare $\widehat{\sigma}_z$ significa misurare i proiettori di eq. (1.67) e ricostruire il valore d'aspettazione con il teorema spettrale:

$$\langle \widehat{\sigma}_z \rangle = +1 \cdot p(0) - 1 \cdot p(1), \tag{1.69}$$

in alcune implementazioni misurare entrambi i proiettori contemporaneamente può essere complicato quindi si ricava una delle due probabilità dall'altra $p(1) = 1 - p(0)$ conoscendo il numero totale di conteggi

$$\langle \widehat{\sigma}_z \rangle = 2p(0) - 1. \tag{1.70}$$

1.2.4 Porte logiche

Le *porte logiche quantistiche* (o *gate*) sono l'analogo delle porte logiche classiche: operatori che agiscono sullo stato del sistema implementando l'evoluzione unitaria descritta nei postulati di sezioni 1.1.1 e 1.1.3. A differenza del caso classico, l'unitarietà impone che ogni gate sia invertibile: non esiste un analogo quantistico diretto di porte irreversibili come AND o OR.

Gate a singolo qubit

Un gate a singolo qubit è descritto da una matrice unitaria 2×2. Tra i più utilizzati si trovano le tre matrici di Pauli, già introdotte in sezione 1.2.1, il *gate di Hadamard*

$$\widehat{H} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \tag{1.71}$$

che mappa la base computazionale nella base X ($\widehat{H}\,|0\rangle = |+\rangle$, $\widehat{H}\,|1\rangle = |-\rangle$), e il *gate di fase*

$$\widehat{S} = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \tag{1.72}$$

che, applicato dopo $\widehat{H}$, permette di ruotare nella base Y ($\widehat{S}\widehat{H}\,|0\rangle = |+i\rangle$, a meno di normalizzazione). Questi due gate, insieme all'identità, sono esattamente quelli che compaiono nello schema di misura *random Pauli* descritto in sezione 2.5.

Più in generale, qualsiasi trasformazione unitaria a singolo qubit può essere vista come una rotazione della sfera di Bloch (fig. 1.1) e parametrizzata tramite angoli di Eulero. Si definiscono le rotazioni attorno ai tre assi

$$\widehat{R}_x(\theta) = e^{-i\theta\widehat{\sigma}_x/2}, \qquad \widehat{R}_y(\theta) = e^{-i\theta\widehat{\sigma}_y/2}, \qquad \widehat{R}_z(\theta) = e^{-i\theta\widehat{\sigma}_z/2}, \tag{1.73}$$

e ogni gate a singolo qubit $\widehat{U}$ può essere scritto, a meno di una fase globale irrilevante, come composizione di tre di queste rotazioni:

$$\widehat{U} = e^{i\alpha}\widehat{R}_z(\beta)\widehat{R}_y(\gamma)\widehat{R}_z(\delta), \tag{1.74}$$

con $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ opportuni.

Gate a più qubit

Per costruire computazioni universali sono necessari anche gate che agiscono su più qubit simultaneamente, in modo da generare entanglement tra i sottosistemi. L'esempio più noto è il *CNOT* (*controlled-NOT*), che applica una $\widehat{\sigma}_x$ su un qubit bersaglio condizionatamente allo stato di un qubit di controllo. La trattazione dettagliata dei gate multi-qubit, dei set universali di porte logiche e della loro sintesi non è rilevante ai fini di questa tesi; per un approfondimento si rimanda a [3, Cap. 4].

[5] È immediato verificare che $\widehat{S}\widehat{H}\,|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i\,|1\rangle) = |+i\rangle$.

Capitolo 2

Tomografia dello Stato Quantistico

La *tomografia dello stato quantistico* consiste nella ricostruzione dello stato di un sistema quantistico a partire da misure fisiche su di esso. Come evidenziato nell'ampia panoramica di [11], questa branca nasce negli anni '70 dallo studio della distribuzione di probabilità nello spazio delle fasi per sistemi a *variabili continue*. L'esempio tipico è la ricostruzione della distribuzione di probabilità dello stato stazionario di una singola particella. In questo caso si studiano funzioni a valori reali nello spazio delle fasi, come la funzione di Wigner $W(\boldsymbol{q}, \boldsymbol{p})$. La sfida risiede nella natura infinito-dimensionale dello spazio di Hilbert che contiene queste funzioni.

Nel contesto di interesse della presente tesi, che è quello delle tecnologie quantistiche, l'approccio si è declinato verso lo studio di sistemi descritti da *variabili discrete*, come nel caso di sistemi di più qubit. La differenza fondamentale risiede nel fatto che lo spazio di Hilbert ora è finito-dimensionale. La tomografia non mira a ricostruire una funzione a variabili continue ma gli elementi di una matrice densità, attraverso la misura di un insieme finito di osservabili.

La misura di sistemi con un alto numero di componenti si scontra con il *problema della dimensionalità*. Una ricostruzione totale tramite metodi classici è intrinsecamente inefficiente, in quanto richiede un numero esponenziale di misure e risorse computazionali. Per questo motivo verranno introdotti due protocolli alternativi volti ad aggirare questo problema.

I metodi classici che si affronteranno nel capitolo sono: inversione lineare (sezione 2.2) e stima di massima verosimiglianza (sezione 2.3), mentre i protocolli più recenti che verranno introdotti consistono in: appilcazione delle reti neurali alla tomografia dello stato completo (sezione 2.4) e le Classical Shadows (sezione 2.5) il cui obiettivo è ricostruire proprietà specifiche dello stato al posto della matrice densità completa.

2.1 Generalità sul problema

2.1.1 Esempio a 1 qubit

Per comprendere il problema può essere utile considerare il semplice esempio di un sistema a singolo qubit [12, Cap. 4]. Modificando leggermente quanto detto in sezione 1.2.1, la matrice densità può essere rappresentata tramite la parametrizzazione

di Stokes

$$\widehat{\rho} = \frac{1}{2}\sum_{i=0}^{3} S_i \widehat{\sigma}_i \tag{2.1}$$

con

$$\widehat{\sigma}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \widehat{1} \quad \widehat{\sigma}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \widehat{\sigma}_x \quad \widehat{\sigma}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \widehat{\sigma}_y \quad \widehat{\sigma}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \widehat{\sigma}_z \tag{2.2}$$

e S_i parametri reali che rispettano:

$$S_0 = 1, \tag{2.3}$$

$$\sum_{i=1}^{3} S_i^2 \in [0,1]. \tag{2.4}$$

Questi parametri, che determinano univocamente lo stato, sono legati al valore di aspettazione dei corrispondenti operatori $\widehat{\sigma}_i$

$$\langle \widehat{\sigma}_i \rangle = \mathrm{Tr}(\widehat{\sigma}_i \widehat{\rho}) = \sum_{j=0}^{3} S_j \,\mathrm{Tr}(\widehat{\sigma}_i \widehat{\sigma}_j) = \sum_{j=0}^{3} S_j \delta_{ij} = S_i. \tag{2.5}$$

In questo modo si è ridotto il problema della tomografia dello stato alla misura dei valori di aspettazione di tre operatori linearmente indipendenti $\{\widehat{\sigma}_i\}_{i=1}^{3}$.

2.1.2 Criteri di valutazione

Per dare una definizione rigorosa del problema e, in seguito, valutare i risultati delle simulazioni, sono necessarie alcune nozioni di norme e distanze sullo spazio degli operatori.

Definizione (Norma di traccia)**.** Dato un operatore $\widehat{X}$ si definisce la *norma di traccia* come

$$\|\widehat{X}\|_{\mathrm{Tr}} = \mathrm{Tr}\,|\widehat{X}|\,, \tag{2.6}$$

con $|\widehat{X}| = \sqrt{\widehat{X}^\dagger \widehat{X}}$.

La norma di traccia equivale alla somma dei valori singolari dell'operatore. Per matrici hermitiane i valori singolari coincidono con i valori assoluti degli autovalori, per cui eq. (2.6) si riduce a

$$\|\widehat{X}\|_{\mathrm{Tr}} = \mathrm{Tr}\,\sqrt{\widehat{X}^2} = \sum_i |\lambda_i| \tag{2.7}$$

con λ_i autovalori di $\widehat{X}$. Questa norma induce, sullo spazio degli operatori densità, una metrica: la *distanza di traccia*.

Definizione (Distanza di traccia)**.** Dati due operatori densità si definisce la loro *distanza di traccia* come

$$d_{\mathrm{Tr}}(\widehat{\rho}_1, \widehat{\rho}_2) = \frac{1}{2}\|\widehat{\rho}_1 - \widehat{\rho}_2\|_{\mathrm{Tr}}. \tag{2.8}$$

È convenzione inserire il fattore 1/2 per fare si che la distanza sia compresa tra 0 e 1.

In sezione 1.1.3 è stato introdotto il prodotto interno di Hilbert-Schmidt, questo induce un'omonima norma.

Definizione (Norma di Hilbert-Schmidt)**.** Dato un operatore $\widehat{X} \in \mathcal{L}(\mathcal{H})$ si definisce la *norma di Hilbert-Schmidt* come

$$\|\widehat{X}\|_{\mathrm{HS}} = \left(\langle \widehat{X}, \widehat{X}\rangle_{\mathrm{HS}}\right)^{1/2} = \left(\mathrm{Tr}\,|\widehat{X}|^{2}\right)^{1/2}. \tag{2.9}$$

Dalla quale si ricava la distanza di Hilbert-Schmidt.

Definizione (Distanza di Hilbert-Schmidt)**.** Dati due operatori $\widehat{A}, \widehat{B} \in \mathcal{L}(\mathcal{H})$, si definisce la loro *distanza di Hilbert-Schmidt* come

$$d_{\mathrm{HS}}(\widehat{A}, \widehat{B}) = \|\widehat{\rho}_1 - \widehat{\rho}_2\|_{\mathrm{HS}}. \tag{2.10}$$

Sebbene la distanza di traccia sia una metrica naturale, per misurare la somiglianza tra stati quantistici spesso si ricorre alla fedeltà di Uhlmann.

Definizione (Fedeltà di Uhlmann)**.** Per due stati quantistici la *fedeltà* è definita come

$$F(\widehat{\rho}_1, \widehat{\rho}_2) = \|\sqrt{\widehat{\rho}_1}\sqrt{\widehat{\rho}_2}\|_{\mathrm{Tr}}^{2} = \left(\mathrm{Tr}(\sqrt{\sqrt{\widehat{\rho}_1}\widehat{\rho}_2\sqrt{\widehat{\rho}_1}})\right)^{2}. \tag{2.11}$$

Questa quantifica la sovrapposizione tra due stati: vale 1 per stati identici e 0 per stati ortogonali. È utile osservare che per due stati puri si semplifica in $F = |\langle\psi_1|\psi_2\rangle|^{2}$.

Tra queste non esiste la metrica perfetta e ciascuna di queste nozioni ha i propri vantaggi e svantaggi:

- La **distanza di Hilbert-Schmidt** è la più semplice da calcolare. Richiede solo operazioni elementari su matrici e il calcolo di tracce, non è necessaria la diagonalizzazione dello stato. Tuttavia, all'aumentare delle dimensioni dello spazio di Hilbert, tende a contrarre le distanze facendo apparire vicini stati fisicamente ben distinti.

- La **distanza di traccia** possiede invece un significato fisico profondo: quantifica esattamente la massima probabilità teorica di distinguere due stati quantistici tramite una singola misura ottimale. A livello numerico, però, richiede l'estrazione degli autovalori della matrice differenza.

- La **fedeltà di Uhlmann**, pur non essendo una vera e propria metrica spaziale, rappresenta lo standard nella teoria dell'informazione quantistica, in quanto ha un'interpretazione fisica diretta e intuitiva.

In particolare, nel resto di questa tesi e nell'analisi delle simulazioni, si utilizzerà la fedeltà come figura di merito principale per quantificare il successo della ricostruzione tomografica, in accordo con gli standard della letteratura. Per garantire il confronto con altri articoli (es. [13]) si valuterà anche la distanza di Hilbert-Schmidt.

In seguito, per affrontare il problema della *shadow tomography*, serviranno alcune definizioni più generali.

Definizione (Semi-norma indotta)**.** Dato un insieme di osservabili $\mathcal{A}$, si definisce la *semi-norma indotta da* $\mathcal{A}$ come

$$\|\widehat{X}\|_{\mathcal{A}} = \sup_{\widehat{A}\in\mathcal{A}} |\mathrm{Tr}(\widehat{X}\widehat{A})| . \tag{2.12}$$

Ancora una volta si può derivare una nozione di distanza (anche se non propriamente una metrica) prendendo la norma della differenza di due operatori:

$$d_{\mathcal{A}}(\widehat{\rho}_1, \widehat{\rho}_2) = \|\widehat{\rho}_1 - \widehat{\rho}_2\|_{\mathcal{A}}. \tag{2.13}$$

Il fatto che $\|\cdot\|_{\mathcal{A}}$ sia una semi norma fa si che, in generale $\|\widehat{X}\|_{\mathcal{A}} = 0$ non implichi $\widehat{X} = 0$. Di conseguenza, possono esistere due operatori densità diversi che rispetto al set di operatori $\mathcal{A}$ appaiono uguali (con distanza nulla).

2.1.3 Definizione del problema

Innanzitutto, è utile definire alcune variabili ricorrenti nel resto del capitolo:

- si indica con n il numero di qubit,
- la dimensione dello spazio di Hilbert è $d = 2^n$,
- gli operatori densità $\widehat{\rho} \in \mathcal{L}(\mathcal{H})$ necessitano di $(2^n)^2 = 4^n$ variabili reali per essere descriti.

Come anticipato, la tomografia si basa sulla misura ripetuta di sistemi preparati allo stesso modo. Il processo di preparazione dello stato e successiva misura è detto *shot*. Un parametro importante della ricostruzione è il *numero di shots* totali N.

Per completezza si dà una definizione rigorosa del problema. Una prima proposta è quella di [14], questa esposizione si basa sulla formulazione di [15].

Definizione (Procedura tomografica)**.** Si consideri un sistema a n componenti nello stato $\widehat{\rho}_0$. Si scelga una precisione ϵ e una probabilità di fallimento $\delta \in (0, 1)$. Una *procedura tomografica* relativa al set di osservabili $\mathcal{A}$ è una procedura sperimentale che usa $N(\epsilon, \delta, n)$ copie identiche del sistema per ricavare $\widehat{\rho}$, una stima di $\widehat{\rho}_0$, che rispetti

$$\mathbb{P}(\|\widehat{\rho} - \widehat{\rho}_0\|_{\mathcal{A}} \leq \epsilon) \geq 1 - \delta. \tag{2.14}$$

Ciò significa che la procedura, usando N copie del sistema, ricostruisce uno stato ϵ-vicino allo stato reale con probabilità $1 - \delta$. Ottenere un limite esatto per N in funzione di ϵ, δ e n, tenendo conto delle incertezze sperimentali è piuttosto complesso. In questa tesi ci si limiterà a valutare empiricamente le performance nel capitolo 3.

Si noti che sotto questa definizione rientrano sia la tomografia completa che quella parziale. Infatti, considerando $\mathcal{A} = \{\widehat{A} \ : \ \|\widehat{A}\|_\infty \leq 1\}$, la norma indotta da $\mathcal{A}$ coincide con la norma di traccia:

$$\|\widehat{X}\|_1 = \|\widehat{X}\|_{\{\widehat{A} \,:\, \|\widehat{A}\|_\infty \leq 1\}} = \sup_{\|\widehat{A}\|_\infty \leq 1} \mathrm{Tr}(\widehat{X}\widehat{A}). \tag{2.15}$$

2.2 Inversione lineare

Il primo metodo affrontato è l'inversione lineare [12, Cap. 4.2], [16, Sez. II]. Innanzitutto, si trasforma la matrice densità $\widehat{\rho}$ in un vettore $(2^n)^2$ dimensionale. Si usa la stessa tecnica vista nell'esempio di sezione 2.1.1, ovvero la parametrizzazione di Stokes, generalizzata a un sistema di n qubit. Si sceglie un set matrici $\{\widehat{\Gamma}_\nu\}_{\nu=1}^{4^n}$, che formano una base ortonormale rispetto al prodotto interno di Hilbert-Schmidt:

$$\forall \nu, \mu, \quad \langle \widehat{\Gamma}_\nu, \widehat{\Gamma}_\mu \rangle = \mathrm{Tr}[\widehat{\Gamma}_\nu^\dagger \widehat{\Gamma}_\mu] = \delta_{\nu,\mu}. \tag{2.16}$$

Questo permette di espandere un qualsiasi operatore $\widehat{A}$ nella base $\{\widehat{\Gamma}_\nu\}$

$$\forall \widehat{A}, \quad \widehat{A} = \sum_{\nu}^{4^n} \widehat{\Gamma}_\nu \, \mathrm{Tr}[\widehat{\Gamma}_\nu^\dagger \widehat{A}]. \tag{2.17}$$

Scelte $\widehat{\Gamma}_\nu$ tutte hermitiane, essendo anche $\widehat{\rho}$ hermitiana, i coefficienti del suo sviluppo sono numeri reali ed eq. (2.17) si riscrive come

$$\widehat{\rho} = \sum_{\nu} \widehat{\Gamma}_\nu r_\nu \quad \text{con} \quad r_\nu = \mathrm{Tr}[\widehat{\Gamma}_\nu \widehat{\rho}] \in \mathbb{R}. \tag{2.18}$$

In questo modo si associa lo stato $\widehat{\rho}$ ad un vettore $\boldsymbol{r} = (r_1, \ldots, r_{4^n})$ di numeri reali.

Si misura il sistema con l'insieme di proiettori $\mathcal{P} = \{\widehat{P}_\mu = |\psi_\mu\rangle\langle\psi_\mu|\}_{\mu=1}^{4^n}$. L'esito μ si ottiene con probabilità

$$p_\mu = \mathrm{Tr}[\widehat{P}_\mu \widehat{\rho}], \tag{2.19}$$

che, ricordando eq. (2.18), si può scrivere in funzione di $\boldsymbol{r}$:

$$p_\mu = \mathrm{Tr}[\widehat{P}_\mu \widehat{\rho}] = \mathrm{Tr}[\widehat{P}_\mu \left(\sum_{\nu} \widehat{\Gamma}_\nu r_\nu \right)] = \sum_{\nu} \mathrm{Tr}[\widehat{P}_\mu \widehat{\Gamma}_\nu] r_\nu. \tag{2.20}$$

Definendo la mappa lineare $B: \ \boldsymbol{r} \mapsto \boldsymbol{p}$

$$B_{\mu,\nu} = \mathrm{Tr}[\widehat{P}_\mu \widehat{\Gamma}_\nu], \tag{2.21}$$

si ottiene il sistema lineare

$$p_\mu = \sum_{\nu} B_{\mu,\nu} r_\nu. \tag{2.22}$$

Per risolvere il problema non resta che invertire la relazione (2.22) e ricostruire $\widehat{\rho}$ da eq. (2.18):

$$\widehat{\rho} = \sum_{\nu} \widehat{\Gamma}_\nu r_\nu = \sum_{\nu} \widehat{\Gamma}_\nu \sum_{\mu} (B^{-1})_{\nu,\mu} p_\mu = \sum_{\mu} \left(\sum_{\nu} \widehat{\Gamma}_\nu (B^{-1})_{\nu,\mu} \right) p_\mu. \tag{2.23}$$

Chiamando

$$\widehat{M}_\mu = \sum_{\nu} \widehat{\Gamma}_\nu (B^{-1})_{\nu,\mu}, \tag{2.24}$$

si ottiene una formula chiusa per $\widehat{\rho}$ in funzione delle misure $\boldsymbol{p}$:

$$\widehat{\rho} = \sum_{\mu} \widehat{M}_{\mu} p_{\mu}. \tag{2.25}$$

Nella risoluzione è stata assunta l'invertibilità della matrice B. Per garantirla, l'insieme $\mathcal{P}$ deve essere composto da 4^n proiettori linearmente indipendenti[1]. Questa condizione è equivalente alla richiesta che l'insieme di misura sia *tomograficamente completo*. L'insieme utilizzato è composto da proiettori su stati presi da basi diverse, ma verrà definito con più precisione nel capitolo successivo.

Nell'inversione lineare standard, viene richiesto un insieme minimale, ovvero che contiene esattamente 4^n proiettori. In caso contrario, B sarebbe una matrice $M \times 4^n$ con $M > 4^n$. La risoluzione è comunque possibile ricorrendo alla tecnica dei minimi quadrati e alla pseudo-inversa di Moore-Penrose, ma la sua trattazione esula dagli scopi di questa tesi.

In laboratorio ciò che si misura non è direttamente il vettore di probabilità $\boldsymbol{p}$ ma piuttosto dei conteggi $\boldsymbol{n}$. Per ciascun proiettore $\widehat{P}_{\nu}$ si misurano N copie indipendenti del sistema, conteggiando il numero di volte n_{ν} in cui la misura dà esito positivo. Le variabili n_{ν} seguono una distribuzione binomiale con valore d'aspettazione $\mathbb{E}(n_{\nu}) = Np_{\nu}$ e varianza $\mathrm{Var}(n_{\nu}) = Np_{\nu}(1-p_{\nu})$. La miglior stima delle probabilità è quindi data dalle frequenze di esiti positivi:

$$f_{\nu} = \frac{n_{\nu}}{N} \tag{2.26}$$

$$\mathbb{E}(f_{\nu}) = p_{\nu} \tag{2.27}$$

$$\mathrm{Var}(f_{\nu}) = \frac{p_{\nu}(1-p_{\nu})}{N}. \tag{2.28}$$

Per una ricostruzione da dati sperimentali eq. (2.25) diventa

$$\widehat{\rho} = \sum_{\mu} \widehat{M}_{\mu} f_{\mu}. \tag{2.29}$$

Sebbene la scelta delle basi di misura possa apparire arbitraria, essa ha un impatto drastico sulla precisione della ricostruzione tomografica. Le fluttuazioni statistiche del vettore di frequenze misurato si propagano sulla ricostruzione, in maniera proporzionale al numero di condizionamento della matrice B. Se si scelgono insiemi di proiettori quasi paralleli, il sistema lineare risulta mal condizionato, amplificando il rumore statistico. Per ottenere la massima precisione a parità di misure, si deve minimizzare la sovrapposizione tra i proiettori.

Per semplicità e coerenza con [16], è stato usato un insieme di misure composto da proiettori, ma questa scelta non è l'unica, nè la più efficiente. Esistono altri schemi che utilizzano le POVM, per esempio le *SIC-POVM* (POVM informazionalmente complete e simmetriche) [17], o la *square root measurement* [18].

Concludiamo rimarcando i principali problemi dell'inversione lineare. Innanzitutto, non garantisce che la matrice ricostruita sia una matrice densità valida. A

[1] È bene non confondere gli elementi $\widehat{P}_{\mu}$ con gli stati che li generano $|\psi_{\mu}\rangle$. Una base di ket è composta da 2^n elementi mentre un insieme massimale di proiettori linearmente indipendenti ne contiene 4^n.

causa di errori sperimentali o anche solo per la natura stocastica delle misure, può accadere che $\mathrm{Tr}[\widehat{\rho}] \neq 1$ o $\mathrm{Tr}[\widehat{\rho}^2] > 1$. Sebbene la prima condizione sia risolvibile con una semplice normalizzazione, la seconda non può essere risolta senza distorcere significativamente lo stato. Tuttavia, il problema principale di questo metodo rimane, come anticipato, la quantità di misure necessarie per una ricostruzione accurata.

2.3 Stima di massima verosimiglianza

Le criticità dell'inversione lineare sono risolte nella tecnica della *stima di massima verosimiglianza* (MLE, dall'inglese *Maximum Likelihood Estimation*). La miglior stima dello stato viene ricavata massimizzando la verosimiglianza, ovvero la probabilità di osservare le frequenze sperimentali, con il vincolo che la matrice sia uno stato fisico, quindi non solo hermitiano ma anche definito positivo e a traccia unitaria. A tal fine è necessario adottare una parametrizzazione più complessa della matrice densità. Come conseguenza la regola di Born eq. (2.19) non è più facilmente invertibile e la soluzione deve essere ottenuta tramite ottimizzazione numerica. A seconda dell'insieme di misura si possono utilizzare diverse funzioni di verosimiglianza e metodi di minimizzazione. Qui si presenta la versione generale descritta in [16, Sez. IV].

Per parametrizzare la matrice densità affinché questa sia sempre uno stato fisico valido si usa la *parametrizzazione di Cholesky*:

$$\widehat{\rho} = \frac{TT^\dagger}{\mathrm{Tr}(TT^\dagger)} \tag{2.30}$$

con T matrice triangolare inferiore con elementi reali sulla diagonale.

Dimostrazione. La verifica dell'hermiticità è immediata: $(TT^\dagger)^\dagger = TT^\dagger$. Per qualsiasi vettore $|\psi\rangle$, $\langle\psi|TT^\dagger|\psi\rangle = \langle\psi'|\psi'\rangle \geq 0$, per cui $\widehat{\rho}$ è semidefinita positiva. La traccia unitaria si impone semplicemente dividendo per $\mathrm{Tr}(TT^\dagger)$. □

Ogni matrice semidefinita positiva ammette decomposizione di Cholesky, che è unica solo se la matrice è strettamente definita positiva. Le variabili che definiscono $\widehat{\rho}$ sono i 2^{2n} parametri reali della matrice di Cholesky. Ad esempio per un qubit:

$$T(\boldsymbol{t}) = \begin{pmatrix} t_1 & 0 \\ t_3 + it_4 & t_2 \end{pmatrix}, \tag{2.31}$$

$$\widehat{\rho}(\boldsymbol{t}) = \frac{T(\boldsymbol{t})T^\dagger(\boldsymbol{t})}{\mathrm{Tr}(T(\boldsymbol{t})T^\dagger(\boldsymbol{t}))}. \tag{2.32}$$

Per trovare i parametri $\boldsymbol{t}$ ottimali si definisce una *funzione di verosimiglianza* che valuta la probabilità che lo stato $\widehat{\rho}(\boldsymbol{t})$ riproduca gli esiti sperimentali $\boldsymbol{f}$ definiti in eq. (2.26). Ricordando che una singola misura è un evento binomiale, per un alto numero di misure le variabili f_μ seguono approssimativamente delle distribuzioni gaussiane con media p_μ e varianza $\sigma_\mu = \sqrt{p_\mu/N}$. La probabilità che uno stato $\widehat{\rho}(\boldsymbol{t})$ generi gli esiti $\boldsymbol{f}$ è quindi

$$P(\boldsymbol{f};\boldsymbol{t}) = \prod_{\mu=1}^{4^n} P(f_\mu;\boldsymbol{t}) = \prod_{\mu=1}^{4^n} \frac{1}{\sqrt{2\pi}} \sqrt{\frac{N}{p_\mu(\boldsymbol{t})}} \exp\left(-\frac{N(f_\mu - p_\mu(\boldsymbol{t}))^2}{2p_\mu(\boldsymbol{t})}\right). \tag{2.33}$$

con p_μ dati da eq. (2.19):

$$p_\mu(\boldsymbol{t}) = \mathrm{Tr}[\widehat{P}_\mu \widehat{\rho}(\boldsymbol{t})]. \tag{2.34}$$

Un modo più semplice e numericamente stabile per massimizzare $P(\boldsymbol{f})$ è minimizzare l'inverso del suo logaritmo. Trascurando la dipendenza da p_μ fuori dall'esponenziale questo vale:

$$\mathcal{L}_{\chi^2}(\boldsymbol{t}; \boldsymbol{f}) = \sum_{\mu=1}^{4^n} \frac{N(f_\mu - p_\mu(\boldsymbol{t}))^2}{2p_\mu(\boldsymbol{t})} \tag{2.35}$$

ovvero la funzione χ^2.

I parametri $\boldsymbol{t}$ ottimali sono quelli che realizzano:

$$\boldsymbol{t}_{\mathrm{MLE}} = \arg\min(\mathcal{L}(\boldsymbol{t}; \boldsymbol{f})), \tag{2.36}$$

la minimizzazione avviene per via numerica. Per avere dei parametri $\boldsymbol{t}$ di partenza da cui svolgere l'ottimizzazione, tipicamente si svolge prima un'inversione lineare.

Sia l'inversione lineare che la stima di massima verosimiglianza sono stimatori asintoticamente non distorti (*unbiased*), quindi il loro valore atteso coincide esattamente con la matrice densità vera: $\mathbb{E}[\widehat{\rho}_{\mathrm{LI}}] = \mathbb{E}[\widehat{\rho}_{\mathrm{MLE}}] = \widehat{\rho}_0$. Ciò significa che, nel limite di un numero infinito di misurazioni ($N \to \infty$), le ricostruzioni tendono allo stato vero $\widehat{\rho}_{\mathrm{LI}} \to \widehat{\rho}_0$ e $\widehat{\rho}_{\mathrm{MLE}} \to \widehat{\rho}_0$. Tuttavia, all'aumentare delle dimensioni del sistema, mantenere un'incertezza statistica costante richiede un numero totale di *shots* che scala con la dimensione della POVM, ovvero 4^n, il che rende presto impraticabile l'approccio standard. Per aggirare questo limite sono state sviluppate principalmente due strategie:

1. **Tomografia completa:** se si desidera ricostruire l'intero stato, si possono integrare informazioni a priori sulle sue proprietà geometriche o fisiche (ad esempio, assumendo che lo stato sia puro, quasi puro, o che appartenga a una specifica distribuzione di stati misti).

2. **Tomografia parziale:** ci si può limitare a ricostruire esclusivamente alcune proprietà specifiche del sistema, come i valori d'aspettazione di un set ridotto di osservabili.

Nell'ultimo decennio si è potuto assistere a progressi significativi in entrambe le direzioni, grazie a due metodologie che verranno approfondite nelle sezioni successive: l'applicazione delle *reti neurali* alla tomografia quantistica e l'introduzione della tecnica delle *Classical Shadows*.

2.4 Tomografia basata su reti neurali

2.4.1 Introduzione alle reti neurali

Le *reti neurali* sono una classe di modelli computazionali appartenenti al campo del *machine learning*. L'obiettivo di una rete neurale è mappare dei dati in input $\boldsymbol{x}$ nell'output desiderato $\boldsymbol{y}$, approssimando una funzione tipicamente ignota o molto complicata $f_{\mathrm{true}}(\boldsymbol{x})$. L'output della rete neurale $\boldsymbol{y} = f(\boldsymbol{x}; \boldsymbol{\theta})$ è funzione di un set di

parametri $\boldsymbol{\theta}$, che vengono ottimizzati nel processo di *addestramento* affinché $f(\boldsymbol{x};\boldsymbol{\theta})$ riproduca il più fedelmente possibile $f_{\text{true}}(\boldsymbol{x})$. L'architettura su cui ci si concentra in questa tesi è detta *feedforward neural network* [19, Cap. 6], in cui il flusso di dati è unidirezionale dall'input verso l'output. La vera forza di queste reti risiede nel *teorema di approssimazione universale*: a grandi linee, per qualsiasi funzione tra due spazi finito-dimensionali, è garantita l'esistenza di una feedforward neural network in grado di approssimarla con precisione arbitraria.

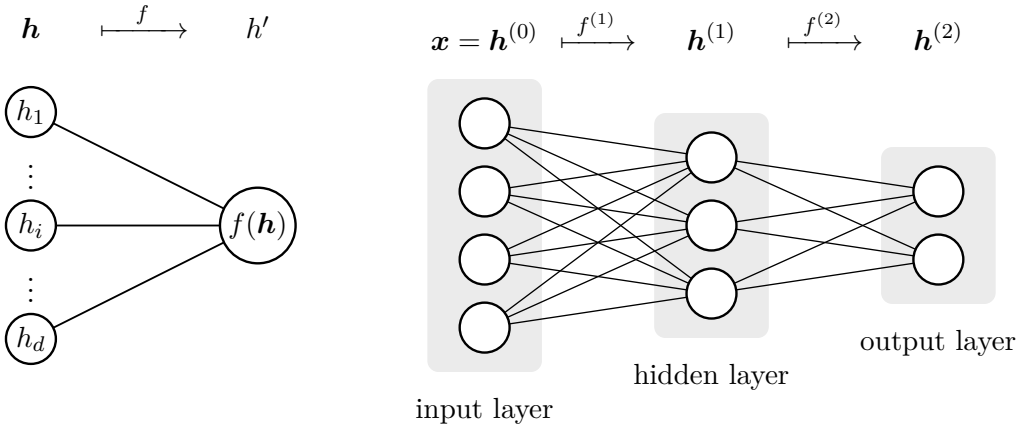

Figura 2.1: A sinistra il diagramma di un singolo neurone, a destra un esempio di rete neurale a tre strati.

L'unità fondamentale della rete è il *neurone* (si veda fig. 2.1, a sinistra): una funzione non lineare $f : \mathbb{R}^d \to \mathbb{R}$ data dalla composizione di una trasformazione affine z seguita da una funzione non lineare g, detta *funzione di attivazione*:

$$z(\boldsymbol{h}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{h} + b \tag{2.37}$$

$$f(\boldsymbol{h}) = g(z(\boldsymbol{h})). \tag{2.38}$$

Più neuroni organizzati in parallelo e alimentati dallo stesso vettore di ingresso definiscono un *layer* (o strato). Gli output di un layer vengono utilizzati come input per lo strato successivo, l'output finale si ottiene come composizione di funzioni (si veda fig. 2.1, a destra):

$$f(\boldsymbol{x};\boldsymbol{\theta}) = f^{(K)}(f^{(K-1)}(\dots(f^{(1)}(\boldsymbol{x})))). \tag{2.39}$$

Ciascun layer k è caratterizzato da un vettore di dati $\boldsymbol{h}^{(k)}$ e da una funzione $f^{(k)} : \mathbb{R}^{d_{k-1}} \to \mathbb{R}^{d_k}$ definita da:

$$\boldsymbol{z}^{(k)}(\boldsymbol{h}^{(k-1)}) = W^{(k)}\boldsymbol{h}^{(k-1)} + \boldsymbol{b}^{(k)} \tag{2.40}$$

$$\boldsymbol{h}^{(k)} = f^{(k)}(\boldsymbol{h}^{(k-1)}) = g^{(k)}(\boldsymbol{z}^{(k)}(\boldsymbol{h}^{(k-1)})) \tag{2.41}$$

dove $W^{(k)} \in \mathbb{R}^{d_k \times d_{k-1}}$ definisce la trasformazione lineare e $\boldsymbol{b}^{(k)} \in \mathbb{R}^{d_k}$ è il vettore di *bias*, sono entrambi parametri allenabili. In questa catena di calcolo, $\boldsymbol{h}^{(0)} = \boldsymbol{x}$ rappresenta il layer di input, mentre i vettori intermedi $\boldsymbol{h}^{(1)}, \dots, \boldsymbol{h}^{(K-1)}$ costituiscono i cosiddetti *layer nascosti*, fino a raggiungere lo strato finale di output $\boldsymbol{h}^{(K)}$. Il

numero di layer K determina la *profondità* del modello, mentre la dimensionalità d_k (ovvero il numero di neuroni) di ciascun livello ne determina la *larghezza*. La funzione di attivazione $g^{(k)}$ è fissata e in comune a tutti i neuroni di uno stesso strato. Per i layer nascosti la scelta ricade solitamente sulla ReLU (Rectified Linear Unit):

$$g_i^{\text{ReLU}}(\boldsymbol{z}) = \max\{0, z_i\} \in \mathbb{R}^+. \tag{2.42}$$

Per l'ultimo strato, l'attivazione viene scelta in base al tipo di variabili richieste in output. Le più utilizzate sono la tangente iperbolica

$$g_i^{\tanh}(\boldsymbol{z}) = \tanh(z_i) \in [-1, 1], \tag{2.43}$$

la sigmoide logistica

$$g_i^{\text{sigmoid}}(\boldsymbol{z}) = \frac{1}{1 + e^{-z_i}} \in [0, 1], \tag{2.44}$$

la *softmax*

$$g_i^{\text{softmax}}(\boldsymbol{z}) = \frac{e^{z_i}}{\sum_j e^{z_j}} \in [0, 1] \quad \text{t.c.} \quad \sum_i g_i(\boldsymbol{z}) = 1, \tag{2.45}$$

oppure lo strato finale può anche essere privo di funzione di attivazione.

A seconda della natura dei dati disponibili nel dataset e della formulazione della funzione di costo, l'addestramento di una rete neurale può procedere secondo due paradigmi:

- *Apprendimento supervisionato*: il dataset di calibrazione X è composto da coppie del tipo $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, dove a ogni vettore di ingresso $\boldsymbol{x}_i$ è associata un'etichetta o un target ideale $\boldsymbol{y}_i$ che la rete dovrà imparare a riprodurre.

- *Apprendimento non supervisionato*: il dataset contiene esclusivamente i vettori di input $X = \{\boldsymbol{x}_i\}$ privi di target associati. Il compito della rete si sposta verso l'estrazione di regolarità intrinseche, la riduzione della dimensionalità dello spazio dei dati o la modellizzazione della distribuzione di probabilità sottostante.

La risoluzione del problema tomografico si colloca nel contesto dell'apprendimento supervisionato, in cui l'obiettivo dell'addestramento è minimizzare una determinata funzione costo (*loss function*).

La *funzione costo* quantifica la discrepanza tra la predizione del modello $f(\boldsymbol{x}_i; \boldsymbol{\theta})$ e il valore reale $\boldsymbol{y}_i$. Come la funzione di attivazione, anche questa dipende dal contesto. Una scelta molto comune è lo scarto quadratico medio (MSE):

$$\mathcal{L}_{\text{MSE}} = \frac{1}{|X|} \sum_{(\boldsymbol{x}, \boldsymbol{y}) \in X} \|f(\boldsymbol{x}) - \boldsymbol{y}\|^2. \tag{2.46}$$

Questa nasce per problemi di regressione lineare ed è ottimizzata per lavorare senza funzione di attivazione nell'ultimo layer, ma viene spesso utilizzata quando non è disponibile una *loss* più adeguata.

Definito il modello e la funzione costo $\mathcal{L}(\boldsymbol{\theta})$ il passo successivo è l'addestramento, ovvero la minimizzazione di $\mathcal{L}(\boldsymbol{\theta})$ rispetto ai parametri del modello $\boldsymbol{\theta}$ su un dataset di allenamento. Per trovare

$$\boldsymbol{\theta}_{\mathrm{ML}} = \arg\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) \tag{2.47}$$

si utilizzano algoritmi iterativi di minimizzazione numerica che sfruttano gradienti analitici calcolati automaticamente. Il calcolo dei gradienti si basa sulla regola della catena dell'analisi. Per questo tipo di reti l'algoritmo utilizzato è detto di *back-propagation*, poiché il gradiente rispetto ai parametri di uno strato dipende dai gradienti calcolati negli strati successivi. Di conseguenza, i gradienti vengono propagati a ritroso, partendo dall'ultimo strato fino a raggiungere il primo. Gli algoritmi per svolgere la minimizzazione sono molteplici, dai più basici come la discesa del gradiente, fino ai più avanzati come il cosiddetto *Adam* [20], che sarà quello utilizzato successivamente.

Spesso i dataset sono molto grandi e calcolare l'intera funzione costo ad ogni iterazione diventa dispendioso: per questo si divide il dataset X in dataset più piccoli X_i detti *batch*. Ad ogni iterazione, dunque, l'ottimmizzazione viene effettuata su un singolo batch. Una serie di iterazioni che coinvolge tutti i batch definisce un *epoca*. Solitamente, l'addestramento termina dopo un certo numero di epoche totali, o dopo che la funzione costo rimane circa costante per un certo numero di epoche consecutive. È buona prassi riservare una parte del dataset per la *validazione*. Questi dati non vengono usati nell'addestramento, ma solo per verificare che la rete stia estraendo caratteristiche generali al posto di memorizzare semplicemente il dataset.

Le caratteristiche di una rete neurale sono molteplici ed interconnesse, è necessario scegliere: il tipo di input e output, il numero di *layers*, il numero di neuroni in ciascun *layer*, le funzioni di attivazione, la funzione costo e l'algoritmo di minimizzazione con i suoi parametri. La progettazione di una rete neurale è quindi un processo complesso, basato su tentativi e successive ottimizzazioni.

Come nota finale si osservi che anche la funzione χ^2 usata in sezione 2.3 è una funzione costo e il problema della stima di massima verosimiglianza, proprio come l'addestramento di una rete neurale, è un problema di machine learning. Ad esempio, al posto del χ^2 si sarebbe potuta utilizzare la cross-entropy (o log-likelihood)

$$\mathcal{L}_{\mathrm{CE}} = -\sum_{i=1}^{N} \sum_{\mu} p_{\mu,i} \log(f_\mu(x_i)), \tag{2.48}$$

usata tipicamente nei problemi di classificazione con reti neurali, ma sarebbe stata una scelta meno specifica.

2.4.2 Reti neurali per la tomografia dello stato quantistico

Descritta a grandi linee una rete neurale, si entra ora nei dettagli implementativi della rete che è stata realizzata. Il progetto si ispira a quello dell'articolo [13].

L'input della rete $\boldsymbol{x}$ è il vettore di frequenze sperimentali, $\boldsymbol{f}$ in eq. (2.26). Per l'output si adotta la stessa parametrizzazione della matrice densità usata in sezione 2.3, ovvero la parametrizzazione di Cholesky, per cui $f(\boldsymbol{x})$ sono i coefficienti reali

($\boldsymbol{t}$) della matrice T di eq. (2.31). La matrice $\widehat{\rho}$ è ricostruita secondo eq. (2.32):

$$\widehat{\rho}(f(\boldsymbol{x})) = \frac{T(f(\boldsymbol{x}))T^{\dagger}(f(\boldsymbol{x}))}{\mathrm{Tr}[T(f(\boldsymbol{x}))T^{\dagger}(f(\boldsymbol{x}))]}. \tag{2.49}$$

Il dataset di allenamento contiene come target $\boldsymbol{y}$ la decomposizione di Cholesky di stati generati randomicamente e come dati $\boldsymbol{x}$ in parte dei vettori di probabilità reali e in parte delle frequenze campionate (più dettagli nella sezione 3.1.4).

Sperimentando con un diverso numero di strati e di neuroni per strato, si è giunti alla conclusione che, per sistemi fino a tre qubit, due layer nascosti, con almeno il doppio dei neuroni rispetto al numero di parametri (4^n) sono sufficienti. Per sistemi di 1, 2 o 3 qubit è stata usata una rete con 4^n neuroni di input, due layer nascosti con 128 neuroni ciascuno e 4^n neuroni di output.

Le funzioni di attivazione sono: ReLU (eq. (2.42)) per strati interni e tangente iperbolica eq. (2.43) per lo strato finale. La tangente iperbolica è stata scelta perché produce simmetricamente valori sia positivi che negativi in $[-1, 1]$. Il range non è importante al fine della normalizzazione della matrice di Cholesky, in quanto calcolando successivamente la matrice densità, si divide per la traccia, ma una funzione di questo tipo garantisce maggior stabilità numerica impedendo ai singoli parametri di assumere valori fuori scala.

Per funzione costo sono state valutate due opzioni:

1. Lo scarto quadratico medio, calcolato direttamente sui parametri della matrice di Cholesky
$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{N}\sum_{i=1}^{N} \|f(\boldsymbol{x}_i) - \boldsymbol{y}_i\|^2. \tag{2.50}$$

2. Il quadrato della distanza di Hilbert-Schmidt tra la matrice densità vera $\widehat{\rho}(\boldsymbol{y})$ e quella predetta $\widehat{\rho}(f(\boldsymbol{x}))$
$$\mathcal{L}_{\mathrm{HS}} = [d_{\mathrm{HS}}(\widehat{\rho}(\boldsymbol{y}), \widehat{\rho}(f(\boldsymbol{x})))]^2 = \mathrm{Tr}\left[(\widehat{\rho}(\boldsymbol{y}) - \widehat{\rho}(f(\boldsymbol{x})))^2\right]. \tag{2.51}$$

La prima si comporta discretamente per stati misti, in quanto la decomposizione di Cholesky è unica. Per stati puri, o comunque non strettamente definiti positivi, la decomposizione di Cholesky non è unica quindi la MSE non è ottimale. La seconda è computazionalmente più complessa, in quanto richiede la ricostruzione di $\widehat{\rho}$ durante l'allenamento. Tuttavia, confrontando direttamente le matrici densità al posto dei parametri, si permette alla rete di estrarre le correlazioni corrette, ottenendo risultati migliori in tutti i regimi testati. Si precisa che per stati puri è stata usata una versione leggermente più semplice di eq. (2.51):

$$\mathcal{L}'_{\mathrm{HS}} = -\mathrm{Tr}\left[\widehat{\rho}(\boldsymbol{y})\widehat{\rho}(f(\boldsymbol{x}))\right]. \tag{2.52}$$

Questa si ricava da eq. (2.51) sviluppando il quadrato e ricordando che per stati puri $\mathrm{Tr}(\widehat{\rho}^2) = 1$. La fedeltà non viene usata direttamente come funzione costo perché necessita del calcolo delle radici quadrate di matrici: non solo l'operazione è estremamente costosa, necessitando della decomposizione spettrale, ma il gradiente della funzione presenta delle singolarità per autovalori nulli.

2.5 Classical Shadows

Come accennato in sezione 2.3, per ricostruire uno stato completo il numero di misure da effettuare sul sistema cresce esponenzialmente con il numero di qubit. Qualora sia sufficiente stimare solo alcune proprietà dello stato, il protocollo delle *Classical Shadows* [21] offre una valida alternativa ai metodi classici. Questo protocollo non è progettato per ricostruire direttamente la matrice densità, ma è ottimizzato per predire i valori di aspettazione di un dato insieme di osservabili, ovvero funzioni lineari[2]nello stato. In questo modo, per opportune osservabili, è possibile aggirare il problema della dipendenza esponenziale del numero di misure dal numero di qubit. Al contrario di quanto visto in precedenza, si otterrà una formula per il numero di *shots* N che non dipende direttamente dal numero di qubit n, ma dal "peso" delle osservabili, quantificato mediante un'opportuna norma.

Il protocollo inizia dalla scelta di $\mathcal{U}$, un insieme finito di trasformazioni unitarie $\widehat{U}$. Indicati con $|b\rangle$ gli elementi della base computazionale: $\{|b\rangle\}_{b\in\{0,1\}^n}$, ciascuna $\widehat{U}\in\mathcal{U}$ definisce la base di misura $\{|u\rangle_{\widehat{U}} = \widehat{U}^\dagger|b\rangle\}_{\{b\in 0,1^n\}}$.

Ad ogni *shot* si seleziona randomicamente un'evoluzione $\widehat{U}\in\mathcal{U}$ e si misura sulla base corrispondente. Per fare ciò si applica $\widehat{U}$, si misura lo stato ottenuto $\widehat{U}\widehat{\rho}\widehat{U}^\dagger$ nella base computazionale, si ottiene un certo esito b, infine viene memorizzato lo stato nella base $\{|u\rangle_{\widehat{U}}\}$ applicando la trasformazione inversa $\widehat{U}^\dagger$. Quindi, fissato $\widehat{U}$, l'esito b corrisponde a

$$\widehat{U}^\dagger|b\rangle\langle b|\,\widehat{U} \tag{2.53}$$

e si ottiene con probabilità

$$P(b) = \langle b|\widehat{U}\widehat{\rho}\widehat{U}^\dagger|b\rangle\,. \tag{2.54}$$

Il valore d'aspettazione di tutta questa operazione, tenendo conto della scelta randomica di $\widehat{U}$ e dell'esito aleatorio della misura b, è:

$$\mathbb{E}\left[\widehat{U}^\dagger|b\rangle\langle b|\,\widehat{U}\right] = \mathbb{E}_{\widehat{U}\sim\mathcal{U}}\left[\mathbb{E}_b\left[\widehat{U}^\dagger|b\rangle\langle b|\,\widehat{U}\right]\right] \tag{2.55}$$

$$= \frac{1}{|\mathcal{U}|}\sum_{\widehat{U}\in\mathcal{U}}\sum_{b\in\{0,1\}^n} P(b)\widehat{U}^\dagger|b\rangle\langle b|\,\widehat{U} \tag{2.56}$$

$$= \frac{1}{|\mathcal{U}|}\sum_{\widehat{U}\in\mathcal{U}}\sum_{b\in\{0,1\}^n} \langle b|\widehat{U}\widehat{\rho}\widehat{U}^\dagger|b\rangle\,\widehat{U}^\dagger|b\rangle\langle b|\,\widehat{U} = \mathcal{M}(\widehat{\rho}). \tag{2.57}$$

$\mathcal{M}:\ \mathcal{S}(\mathcal{H})\to\mathcal{S}(\mathcal{H})$ è un operazione lineare in $\widehat{\rho}$ anche detto *canale quantistico*[3], invertibile per opportune scelte di $\mathcal{U}$. A questo punto si definisce *classical shadow*

$$\widehat{\rho}_{\text{cs}} = \mathcal{M}^{-1}(\widehat{U}^\dagger|b\rangle\langle b|\,\widehat{U}), \tag{2.58}$$

ovvero lo stimatore di $\widehat{\rho}$ dato da una singola misura, il suo valore atteso equivale al vero operatore densità:

$$\mathbb{E}[\widehat{\rho}_{\text{cs}}] = \widehat{\rho}. \tag{2.59}$$

[2]Approfondendo [21], si può trovare anche un esempio di Classical Shadows per funzioni quadratiche e il concetto può essere generalizzato a polinomi di vario ordine.

[3]Un canale quantistico è una mappa lineare, completamente positiva e che preserva la traccia (CPTP), che mappa operatori densità in peratori densità.

Sia ora $\widehat{A}$ una funzione lineare nello stato, ad esempio un operatore associato ad un'osservabile, della quale si vuole predire il valore di aspettazione:

$$a = \mathrm{Tr}[\widehat{\rho}\widehat{A}]. \tag{2.60}$$

La variabile aleatoria

$$a_{\mathrm{cs}} = \mathrm{Tr}\left[\widehat{\rho}_{\mathrm{cs}}\widehat{A}\right] \tag{2.61}$$

riproduce il corretto valore d'aspettazione dell'osservabile $\widehat{A}$:

$$\mathbb{E}[a_{\mathrm{cs}}] = \mathrm{Tr}\left[\mathbb{E}[\widehat{\rho}_{\mathrm{cs}}]\widehat{A}\right] = \mathrm{Tr}[\widehat{\rho}\widehat{A}] = a. \tag{2.62}$$

Inoltre la sua varianza segue:

$$\mathrm{Var}(a_{\mathrm{cs}}) = \mathbb{E}(a_{\mathrm{cs}} - \mathbb{E}(a_{\mathrm{cs}}))^2 \leq \|\widehat{A} - \frac{\mathrm{Tr}[\widehat{A}]}{2^n}\widehat{1}\|^2_{\mathrm{shadow}}, \tag{2.63}$$

dove la *norma shadow* è definita come

$$\|\widehat{A}\|_{\mathrm{shadow}} = \max_{\widehat{\rho}\ \text{stato}} \left[\mathbb{E}_{\widehat{U}\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\widehat{U}\widehat{\rho}\widehat{U}^\dagger|b\rangle \left(\langle b|\widehat{U}\mathcal{M}^{-1}(\widehat{A})\widehat{U}^\dagger|b\rangle\right)^2\right]^{1/2}. \tag{2.64}$$

La *norma shadow* quantifica la varianza massima, sul peggior stato possibile, dello stimatore a_{cs} associato all'osservabile $\widehat{A}$, per un dato schema di misura $\mathcal{U}$. È questa quantità, e non la dimensione dello spazio di Hilbert, a determinare il numero di *shot* necessari. Per osservabili locali e scelte opportune di $\mathcal{U}$, $\|\widehat{A}\|^2_{\mathrm{shadow}}$ resta limitata da una costante indipendente da n, il che è precisamente ciò che rende il protocollo efficiente.

A questo punto ci si interroga su quale sia il modo più efficiente di valutare il valore vero di un'osservabile eq. (2.60) a partire da N Classical Shadows indipendenti $(\widehat{\rho}_{\mathrm{cs},1}, \ldots \widehat{\rho}_{\mathrm{cs},N})$. Il modo più ovvio è quello di valutare direttamente il valore d'aspettazione come media:

$$a(N,1) = \frac{1}{N}\sum_{n=1}^{N} \mathrm{Tr}[\widehat{A}\widehat{\rho}_{\mathrm{cs},j}], \tag{2.65}$$

il cui problema, però, è che il numero di misure necessarie scala come $1/\delta$ con una dipendenza $N = \frac{\mathrm{Var}(a)}{\delta\epsilon^2}$. In alternativa, si può utilizzare il metodo della *mediana delle medie*. Dividendo il campione in N gruppi da K elementi, per ciascun gruppo si fa la media dei suoi elementi, per poi prendere la mediana di tutte queste medie. Quindi, usando $N_{\mathrm{tot}} = NK$ elementi:

$$a(N,K) = \underset{k\in\{1,\ldots,K\}}{\mathrm{mediana}}\{a^{(k)}(N,1)\} \quad \text{con} \quad a^{(k)}(N,1) = \frac{1}{N}\sum_{1+N(k-1)}^{Nk} \mathrm{Tr}[\widehat{A}\widehat{\rho}_{\mathrm{cs},j}]. \tag{2.66}$$

Ciò permette di avere un miglioramento esponenziale nella dipendenza dalla probabilità di fallimento.

Il risultato principale di [21] è il seguente.

Teorema. *Scelte M osservabili di interesse, $\widehat{A}_1, \ldots, \widehat{A}_M$, e fissati*

$$K = 2\ln\left(\frac{2M}{\delta}\right) \quad e \quad N = \frac{34}{\epsilon^2} \max_{i=1,\ldots,M} \|\widehat{A}_i - \frac{\mathrm{Tr}[\widehat{A}_i]}{2^n}\widehat{1}\|^2_{\mathrm{shadow}}, \tag{2.67}$$

il metodo delle Classical Shadows, con lo stimatore mediana delle medie*, predice tutti i valori d'aspettazione con errore minore di ϵ con probabilità maggiore di $1-\delta$. Il problema è quindi risolto con un numero di misure:*

$$N_{\mathrm{tot}} = \mathcal{O}\left(\frac{\ln(M)}{\epsilon^2} \max_{i=1,\ldots,M} \|\widehat{A}_i - \frac{\mathrm{Tr}[\widehat{A}_i]}{2^n}\widehat{1}\|^2_{\mathrm{shadow}}\right). \tag{2.68}$$

Le scelte più comuni per l'insieme $\mathcal{U}$ sono due. La prima è il gruppo di Clifford su n qubit campionato uniformemente. Questa scelta rende la *norma shadow* limitata per osservabili a basso rango (ad esempio proiettori su stati puri o fedeltà), ma richiede circuiti difficili da implementare. La seconda è $\mathcal{U} = \{\widehat{H}, \widehat{S}\widehat{H}, \widehat{1}\}^{\otimes n}$, ovvero rotazioni di Pauli casuali indipendenti applicate a ciascun qubit (*random Pauli measurements*). Questo schema è facile da realizzare su circuiti quantistici ed è particolarmente efficiente per osservabili locali, a scapito però delle prestazioni su osservabili globali. Per entrambe il canale $\mathcal{M}$ ha una forma chiusa facilmente invertibile e la *norma shadow* può essere calcolata semplicemente.

Ad ogni modo, in questo lavoro di tesi, il metodo delle *Classical Shadows* non verrà approfondito ulteriormente né confrontato con gli altri metodi in quanto non è ottimale per la tomografia completa dello stato. Per chi fosse interessato ad approfondire, si segnalano alcuni sviluppi successivi al lavoro originale: [22] per le *randomized measurements*, che inquadra le Classical Shadows in un contesto più ampio di tecniche basate su misure casuali; il lavoro di [23] che studia il protocollo in presenza di rumore sperimentale; e la tecnica di *derandomizzazione* di [24], che permette di scegliere $\mathcal{U}$ in modo deterministico e adattivo, riducendo ulteriormente il numero di misure necessarie quando le osservabili di interesse sono note in anticipo.

Capitolo 3

Simulazioni e analisi comparativa

Dopo aver descritto e confrontato i principali metodi per la tomografia quantistica, in questo capitolo si riporteranno delle simulazioni numeriche. L'analisi riguarderà esclusivamente la tomografia completa di stati quantistici e i metodi che verranno comparati consistono in:

- Inversione Lineare (LI),
- Stima di Massima Verosimiglianza (MLE),
- Rete Neurale (NN).

Le Classical Shadows, seppur applicabili anche alla ricostruzione completa di stati quantistici, non verranno analizzate.

L'obiettivo del capitolo è comparare le performance dei diversi metodi, tenendo in considerazione vantaggi e svantaggi di ognuno

3.1 Implementazione

Il codice per le simulazioni numeriche è stato implementato in `Python` ed è reperibile su github `https://github.com/enro284/quantum-state-tomography-sim.git`. Di seguito si approfondiscono i vari dettagli.

3.1.1 Generazione di stati casuali

Nelle simulazioni si usano tre tipi di stati:

- **Puri:** derivati da un ensemble di un solo stato, hanno purezza unitaria $\mathrm{Tr}(\widehat{\rho}^2) = 1$
- **Misti:** generati secondo l'ensemble di Hilbert-Schmidt [25]. La loro purezza può assumere qualsiasi valore $\frac{1}{2^n} \leq \mathrm{Tr}(\widehat{\rho}^2) \leq 1$ ma sono molto più probabili valori tendenti al limite inferiore.
- **Semi-puri:** stati misti derivati da un ensemble di due stati con $p_1 = 1 - \epsilon$, $p_2 = \epsilon$. Si fissa $\epsilon = 0.1$, la purezza dello stato è $0.82 \leq \mathrm{Tr}(\widehat{\rho}^2) \leq 1$.

Uno stato puro è generato estraendo parte reale e parte immaginaria dei coefficienti del vettore d'onda da distribuzioni gaussiane indipendenti, a media nulla e varianza unitaria. Il vettore d'onda viene normalizzato e convertito in matrice densità.

$$b = 1, \ldots, 2^n, \quad c_b = \mathcal{N}(0,1) + i\mathcal{N}(0,1), \tag{3.1}$$

$$|\psi\rangle = \sum_{b=1}^{2^n} c_b \, |b\rangle \, , \quad |b\rangle \text{ base computazionale}, \tag{3.2}$$

$$\widehat{\rho} = |\psi\rangle\langle\psi| \, . \tag{3.3}$$

Per stati semi-puri si generano due stati puri $\widehat{\rho}_1$ e $\widehat{\rho}_2$, la matrice densità è

$$\widehat{\rho} = (1-\epsilon)\widehat{\rho}_1 + \epsilon\widehat{\rho}_2 \quad \text{con} \quad \epsilon = 0.1 \tag{3.4}$$

Gli stati misti si ottengono come

$$\widehat{\rho} = \frac{GG^\dagger}{\mathrm{Tr}(GG^\dagger)} \tag{3.5}$$

dove G è una matrice dell'ensemble di Ginibre, quindi con parte reale e parte immaginaria di ciascun coefficiente estratte come variabili normali indipendenti.

3.1.2 Inversione Lineare

L'algoritmo è esattamente quello descritto in sezione 2.2, implementato tramite una specifica base $\{\widehat{\Gamma}_\nu\}_{\nu=0}^{4^n-1}$ e uno specifico insieme di misura $\mathcal{P}$.

Come visto in sezione 2.1.1 le matrici di Pauli con l'identità definiscono una base a un qubit:

$$\widehat{\Gamma}_0 = \frac{1}{\sqrt{2}}\widehat{1}, \quad \widehat{\Gamma}_1 = \frac{1}{\sqrt{2}}\widehat{\sigma}_x, \quad \widehat{\Gamma}_2 = \frac{1}{\sqrt{2}}\widehat{\sigma}_y, \quad \widehat{\Gamma}_3 = \frac{1}{\sqrt{2}}\widehat{\sigma}_z. \tag{3.6}$$

La generalizzazione a n qubit segue da quest'ultima, con la base data dal prodotto tensoriale di tutte le combinazioni a n elementi della base a un qubit:

$$\widehat{\Gamma}_\nu = \widehat{\Gamma}_{\nu_1} \otimes \cdots \otimes \widehat{\Gamma}_{\nu_n}. \tag{3.7}$$

Per generare tutte le combinazioni al variare di $\nu = 0, \ldots, 4^n - 1$, si scelgono come indici $\nu_1, \ldots, \nu_n$ le cifre di ν in base 4. La verifica di eq. (2.16) e (2.17) è immediata.

L'insieme di misura ad un qubit è composto dai proiettori:

$$\widehat{P}_1 = |0\rangle\langle 0| \, , \quad \widehat{P}_2 = |-i\rangle\langle -i| \, , \quad \widehat{P}_3 = |+\rangle\langle +| \, , \quad \widehat{P}_4 = |1\rangle\langle 1| \, . \tag{3.8}$$

La generalizzazione a n qubit segue lo stesso schema usato per le basi:

$$\widehat{P}_\mu = \widehat{P}_{\mu_1} \otimes \cdots \otimes \widehat{P}_{\mu_n}. \tag{3.9}$$

Seguendo le eq. (2.21) e (2.24) si può facilmente calcolare la matrice B, la sua inversa e le matrici $\widehat{M}_\mu$. Il metodo `linear_inversion(frequency_vector)` realizza eq. (2.29).

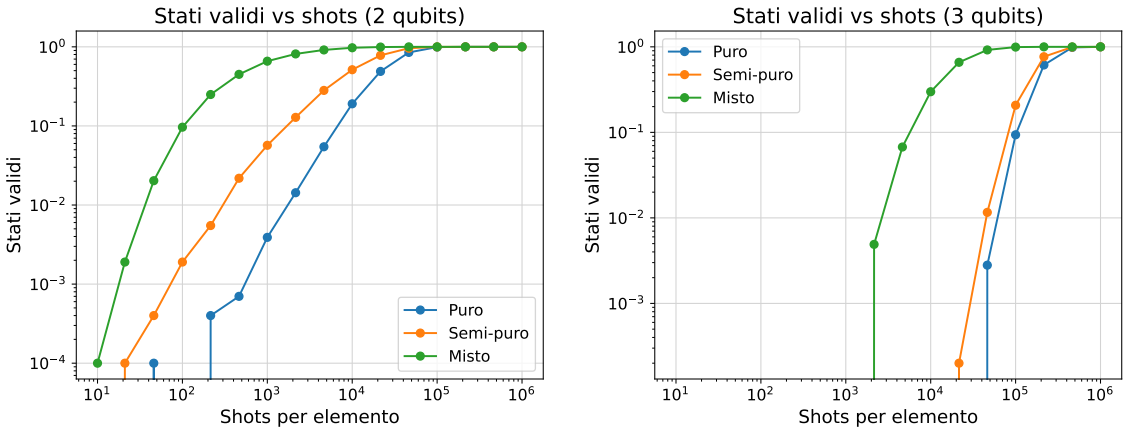

Figura 3.1: Frazione di stati validi sul totale delle ricostruzioni per inversione lineare. A sinista il grafico per 2 qubit, a destra per 3. I punti di colore diverso rappresentano i tre tipi di stati: puri, semipuri o misti.

Come accennato in sezione 2.2 l'inversione lineare non implementa il vincolo che la matrice densità sia definita positiva, perciò, a causa delle incertezze statistiche, è possibile che venga ricostruita una matrice negativa. Si ricorda che una matrice è definita positiva se e solo se tutti i suoi autovalori sono maggiori o uguali a zero. In fig. 3.1 è riportata la frazione di stati (su un totale di 1000 ricostruzioni per punto) senza autovalori negativi (entro una tolleranza di 10^{-2}).

I risultati sono coerenti con le aspettative: all'aumentare del numero di misure (e quindi della precisione dei dati) il numero di stati fisici aumenta. Tuttavia per avere buone probabilità di ottenere stati validi ($\sim 90\%$) il numero di misure è elevatissimo: $\sim 10^5$ shot per elemento per stati di 2 qubit (quindi $\sim 4 \times 10^5$ shot totali), $\sim 10^6$ per 3 qubit (quindi $\sim 2 \times 10^7$ shot totali). Stati puri e semi-puri sono caratterizzati da un alto numero di autovalori nulli, per questi ottenere una stima positiva è più difficile.

Questi risultati giustificano la necessità di forzare la positività delle matrici stimate prima di confrontarle con lo stato vero o di utilizzarle per la stima di massima verosimiglianza.

3.1.3 Stima di massima verosimiglianza

L'algoritmo descritto in sezione 2.3 presenta tre importanti criticità:

- Per avere dei parametri $\boldsymbol{t}$ di partenza dall'inversione lineare è necessario svolgere la decomposizione di Cholesky di stati potenzialmente non definiti positivi.
- La singolarità della funzione costo $\mathcal{L}_{\chi^2}$ per $p_\mu \to 0$.
- La minimizzazione di una funzione non lineare ad alta dimensionalità.

La prima è facilmente risolvibile: trattandosi solamente di parametri di partenza non serve che questi siano esatti. La $\widehat{\rho}$ ricavata viene resa strettamente positiva svolgendo una decomposizione spettrale, per poi troncare gli autovalori nel range

$[\epsilon, 1]$ e ricomporre la matrice. La decomposizione di Cholesky viene affidata ad il metodo `numpy.linalg.cholesky()` e T viene convertita nel vettore di parametri da ottimizzare $\boldsymbol{t}$. Per ovviare al secondo problema, in eq. (2.35) sono state poste delle condizioni sul denominatore. Se $p_\mu > \delta$ si calcola il termine $\frac{(f_\mu - p_\mu(\boldsymbol{t}))^2}{2p_\mu(\boldsymbol{t})}$; se $p_\mu < \delta$ e $f_\mu > \delta$ si calcola $\frac{(f_\mu - p_\mu(\boldsymbol{t}))^2}{2f_\mu(\boldsymbol{t})}$; se entrambi sono minori di δ il dato viene escluso dalla sommatoria. Come tolleranza si è scelto $\delta = 10^{-5}$.

Il terzo problema è più complesso, poiché nella scelta dell'algoritmo di ottimizzazione è necessario tenere conto di molti fattori tra cui il tipo di misura utilizzato e la quantità di qubit in relazione alle risorse computazionali. Per mantenere la trattazione semplice, si è scelto di adottare il medesimo approccio degli autori originali [16], sfruttando un algoritmo di ottimizzazione multidimensionale privo di gradienti analitici (nello specifico, il metodo dei set di direzioni di Powell [26]). Questo è implementato dalla funzione `scipy.optimize.minimize(method='Powell')`.

3.1.4 Rete Neurale

L'architettura della rete neurale definita in sezione 2.4.2 è stata implementata in `python` con la libreria `PyTorch` [27].

Sono state addestrate sei reti diverse, una per ogni tipo di stato a 2 e a 3 qubit. Per ciascuna il dataset di allenamento è composto da matrici di Cholesky con le probabilità di misura della matrice densità corrispondente. In realtà solo l'80% degli input della rete sono probabilità pure, il resto sono frequenze campionate da una distribuzione binomiale con numero di shots per proiettore variabile tra 10 e 10^3. Come notato da [13] questo accorgimento migliora le prestazioni della rete per bassi numeri di misure.

Il dataset (50000 elementi per due qubit e 200000 per tre qubit) è stato diviso in batch da 100 elementi e il 20% è stato tenuto per la validazione. L'allenamento è proseguito per un numero variabile di epoche tra le 100 e le 300, fino al raggiungimento del minimo della funzione costo sul dataset di validazione.

Gli stati che costituiscono il dataset di test verranno generati in un secondo momento ma seguiranno le stesse distribuzioni usate per addestramento e validazione.

3.2 Ricostruzione di stati a due qubit

3.2.1 Esempio: stato entangled $|\Psi^+\rangle$

Prima di addentrarsi nelle statistiche di confronto dei vari metodi si prova a ricostruire un singolo stato:

$$\left|\Psi^+\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \tag{3.10}$$

la cui matrice densità è:

$$\widehat{\rho} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.11}$$

Per la ricostruzione, il numero di shot è scelto per avere una fedeltà media di almeno 0.99 sul metodo più preciso. Le matrici densità stimate dai tre metodi con 100 shot per proiettore (quindi 1600 shots totali) sono:

$$\rho_{\mathrm{LI}} = \begin{pmatrix} -0.100 & 0.090+0.050i & 0.130+0.040i & 0.075-0.115i \\ 0.090-0.050i & 0.560 & 0.365+0.015i & -0.020+0.010i \\ 0.130-0.040i & 0.365-0.015i & 0.540 & -0.100-0.010i \\ 0.075+0.115i & -0.020-0.010i & -0.100+0.010i & 0.000 \end{pmatrix},$$

$$\rho_{\mathrm{MLE}} = \begin{pmatrix} 0.013 & 0.047-0.014i & 0.071-0.005i & -0.007 \\ 0.047+0.014i & 0.512 & 0.390-0.027i & -0.008+0.023i \\ 0.071+0.005i & 0.390+0.027i & 0.468 & -0.040+0.016i \\ -0.007 & -0.008-0.023i & -0.040-0.016i & 0.007 \end{pmatrix},$$

$$\rho_{\mathrm{NN}} = \begin{pmatrix} 0.003 & 0.038-0.017i & 0.036-0.015i & -0.001+0.004i \\ 0.038+0.017i & 0.528 & 0.495+0.007i & -0.028+0.039i \\ 0.036+0.015i & 0.495-0.007i & 0.465 & -0.026+0.037i \\ -0.001-0.004i & -0.028-0.039i & -0.026-0.037i & 0.004 \end{pmatrix}.$$

In tabella 3.1 sono mostrate le distanze di Hilbert-Schmidt e fedeltà di ciascun metodo. Come atteso la rete neurale (NN) restituisce la ricostruzione più accurata, seguita da stima di massima verosimiglianza (MLE) e infine inversione lineare (LI). L'inversione lineare ha restituito uno stato non valido: ha un autovalore negativo (-0.224), che è stato impostato manualmente a 0 per confrontare il prodotto della ricostruzione con gli altri metodi.

Metodo	**Distanza di H-S**	**Fedeltà**
LI (Positivo)	0.301	0.751
MLE	0.174	0.880
NN	0.094	0.991

Tabella 3.1: Confronto delle metriche di ricostruzione dello stato $|\Psi^+\rangle$ con 100 shots per proiettore. Le reti neurali hanno la performance migliore, seguite da stima di massima verosimiglianza e inversione lineare.

3.2.2 Confronto

Per il test dei tre metodi sono stati generati casualmente 1000 stati per ogni tipologia di stato come descritto in sezione 3.1.1. D'ora in poi la metrica di riferimento sarà la fedeltà, anche se per completezza sono riportati anche i grafici della distanza di Hilbert-Schmidt.

Per prima cosa si confrontano le distribuzioni delle fedeltà tra i tre metodi a parità di numero di shots. Il numero di shots è scelto per avere una fedeltà media di almeno 0.99 sul metodo più preciso. Si fa riferimento agli istogrammi delle fig. da 3.2 a 3.4.

Per stati puri la rete neurale è la più precisa, seguita dalla stima di massima verosimiglianza e dall'inversione lineare. La spiegazione è semplice: i due metodi

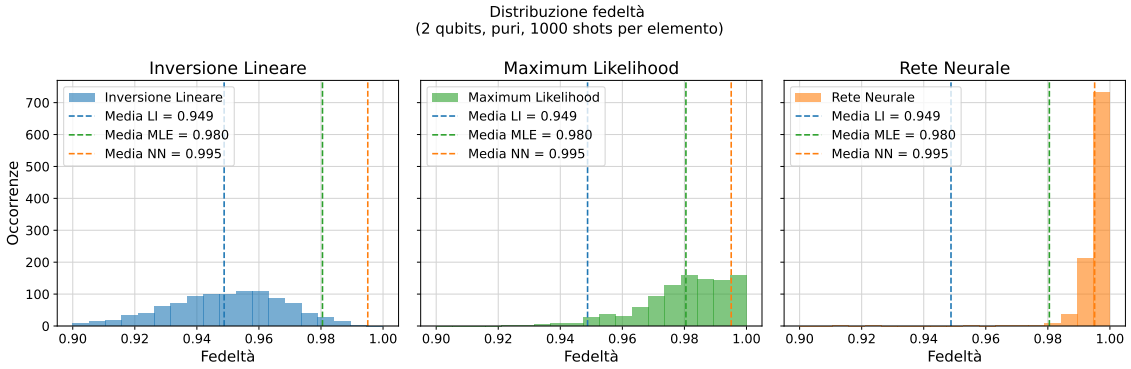

Figura 3.2: Istogrammi delle fedeltà delle ricostruzioni per stati puri di due qubit con 1000 shots per proiettore. La NN ha la performance migliore, sia in termini di fedeltà media che di dispersione, è seguita da MLE e LI.

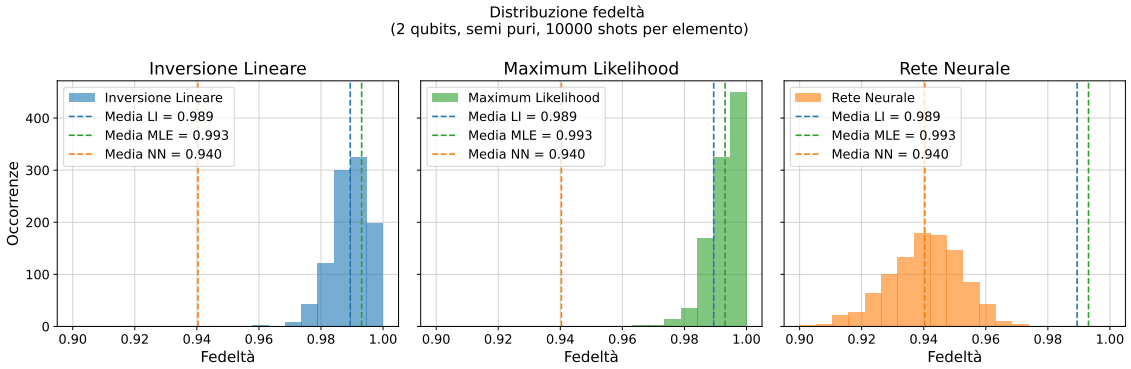

Figura 3.3: Istogrammi delle fedeltà delle ricostruzioni per stati semi-puri di due qubit con 10 000 shots per proiettore. La MLE ma la performance migliore, sia in termini di fedeltà media che di dispersione, seguita a poca distanza da LI, le fedeltà della NN sono le peggiori.

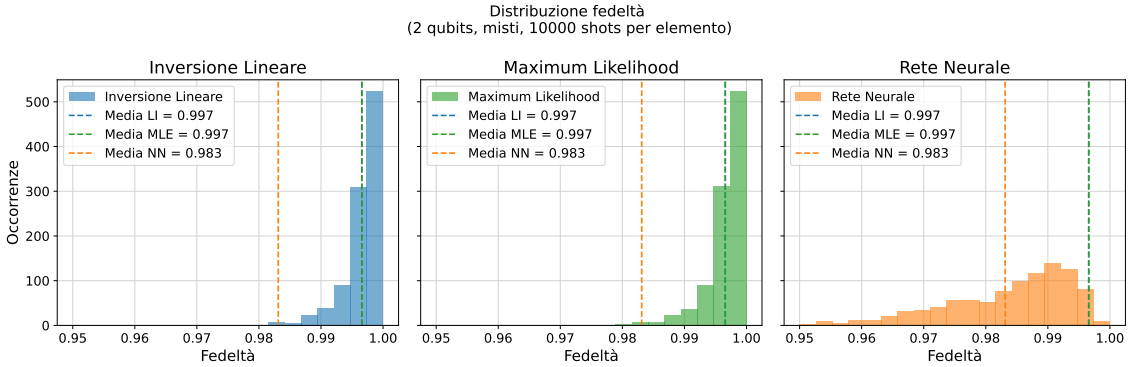

Figura 3.4: Istogrammi delle fedeltà delle ricostruzioni per stati misti di due qubit con 10 000 shots per proiettore. MLE e LI hanno performance identiche con fedeltà di 0.997, la NN ha una dispersione di fedeltà molto più ampia e media peggiore.

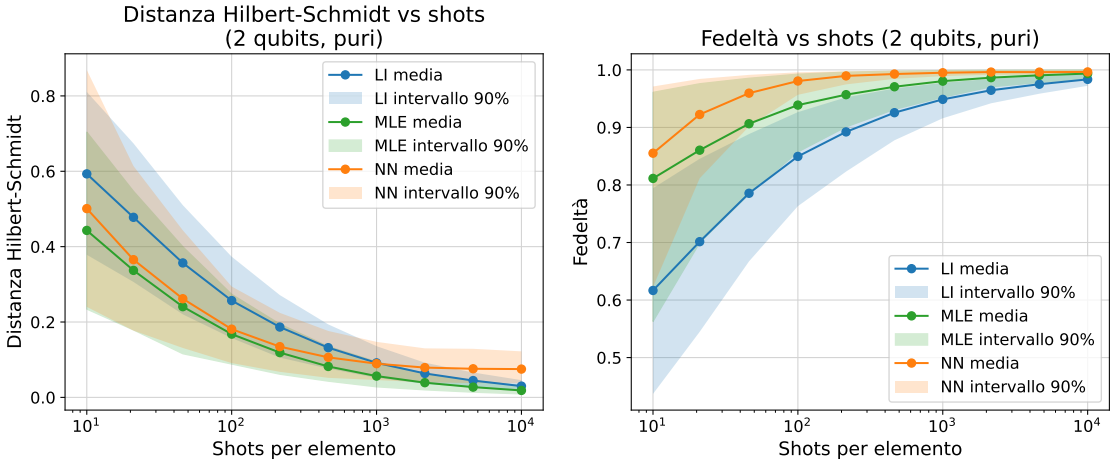

Figura 3.5: Qualità della ricostruzione per stati puri di due qubit rispetto al numero di misure, a sinistra è riportata la distanza di Hilbert-Schmidt, a destra la fedeltà.

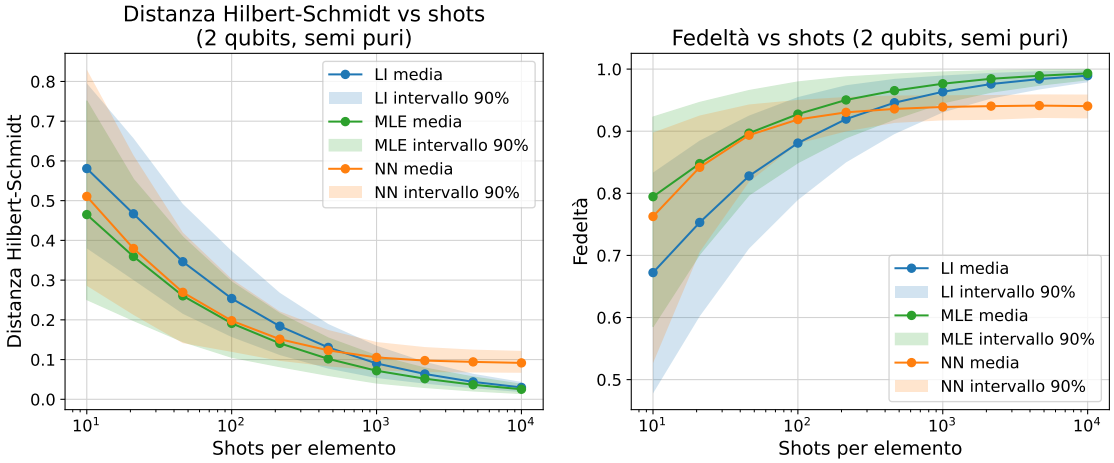

Figura 3.6: Qualità della ricostruzione per stati semi-puri di due qubit rispetto al numero di misure, a sinistra è riportata la distanza di Hilbert-Schmidt, a destra la fedeltà.

classici sono *unbiased*, non hanno informazione sul tipo di stato, mentre la rete è stata allenata esclusivamente su stati puri e di conseguenza tenderà a prevedere stati di quel tipo.

Per stati semi-puri e misti le reti neurali hanno la performance peggiore: seppur addestrate a riconoscere stati di questo tipo, l'alto numero di misure (160 000 shot) necessario a raggiungere una fedeltà di 0.99, avvantaggia i metodi classici in quanto asintoticamente *unbiased*.

Per avere un quadro più completo si confronta ora l'andamento della precisione delle tre tecniche al variare del numero di misure, i risultati sono in fig. da 3.5 a 3.7.

Per ogni tipo di stato la stima di massima verosimiglianza si comporta decisamente meglio dell'inversione lineare. Come atteso, per numero di shots molto elevati le performance di questi due metodi tendono ad eguagliarsi.

Si confrontano infine le reti neurali con i metodi classici al variare del numero di

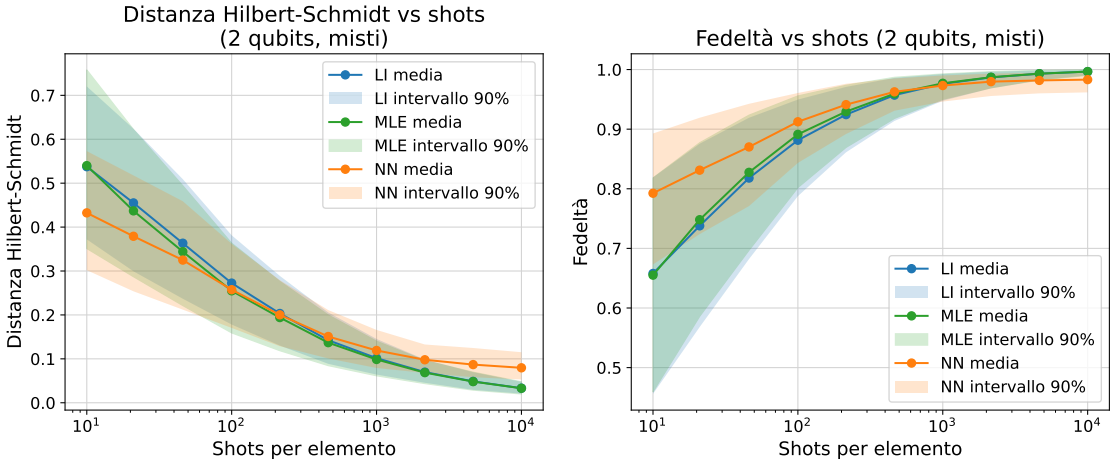

Figura 3.7: Qualità della ricostruzione per stati misti di due qubit rispetto al numero di misure, a sinistra è riportata la distanza di Hilbert-Schmidt, a destra la fedeltà.

misure. Per stati puri la rete neurale mantiene la fedeltà di ricostruzione più alta in tutti i regimi considerati. Per stati misti il divario si riduce sensibilmente: la MLE supera la rete neurale a partire da $\sim 10^3$ shot per proiettore. Per stati semi-puri, invece, la MLE è costantemente superiore alla rete neurale, la quale sembra raggiungere un plateau con fedeltà inferiore al 95%; in questo regime la rete viene superata anche dall'inversione lineare a partire da ~ 500 shot per proiettore.

Una possibile spiegazione di questo comportamento risiede nella definizione di stato semi-puro adottata (sezione 3.1.1). La miscela $\widehat{\rho} = (1-\epsilon)\,|\psi_1\rangle\langle\psi_1| + \epsilon\,|\psi_2\rangle\langle\psi_2|$ di due stati puri produce uno stato i cui autovalori effettivi dipendono dalla sovrapposizione casuale $\langle\psi_1|\psi_2\rangle$ e non sono fissati a $(1-\epsilon, \epsilon)$. Al variare della sovrapposizione si passa da stati quasi puri a stati genuinamente a rango 2, rendendo la classe degli stati semi-puri piuttosto eterogenea. Poiché la rete neurale deve apprendere un'unica funzione di ricostruzione valida su tutta la classe di addestramento, l'eterogeneità della definizione qui adottata potrebbe risultare ancor più difficile da modellare rispetto alla distribuzione di stati misti. Come possibile sviluppo futuro, sarebbe interessante ripetere addestramento e test della rete neurale su stati semi-puri costruiti secondo la definizione più convenzionale $\widehat{\rho} = (1-\epsilon)\,|\psi\rangle\langle\psi| + \epsilon\,\widehat{1}/d$. Il confronto con le prestazioni ottenute in questo caso permetterebbe di verificare se l'eterogeneità della distribuzione qui adottata sia effettivamente la causa del comportamento osservato.

3.3 Ricostruzione di stati a tre qubit

A causa del costo computazionale della stima di massima verosimiglianza (~ 1 s per ricostruzione), il numero di stati di test è stato ridotto a 100 per tipologia.

Su stati a tre qubit le reti neurali perdono parte del vantaggio rispetto alla stima di massima verosimiglianza. Questo risultato è inaspettato e in contrasto con [13]. Il problema può essere dovuto al fatto che la rete sia troppo semplice in relazione al numero di qubit o che i dati di addestramento siano insufficienti.

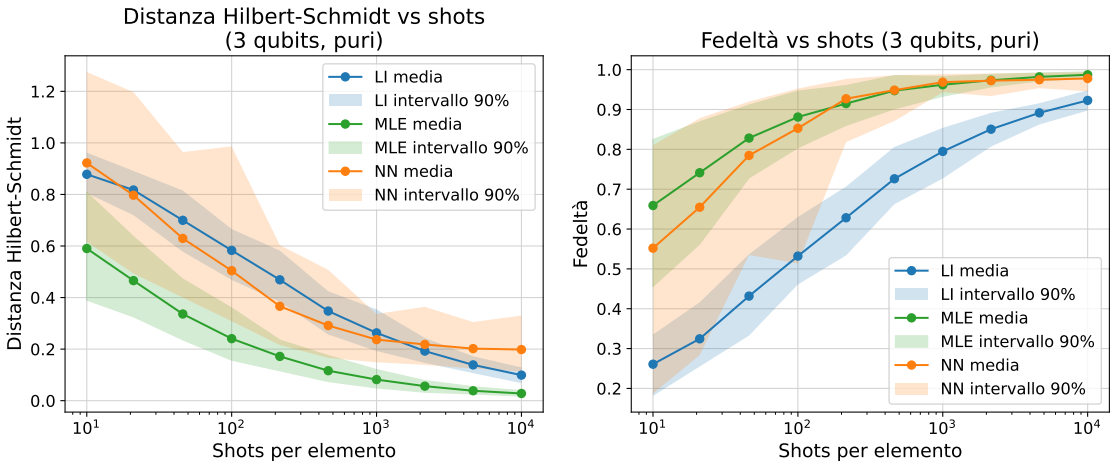

Figura 3.8: Qualità della ricostruzione per stati puri di tre qubit rispetto al numero di misure, a sinistra è riportata la distanza di Hilbert-Schmidt, a destra la fedeltà.

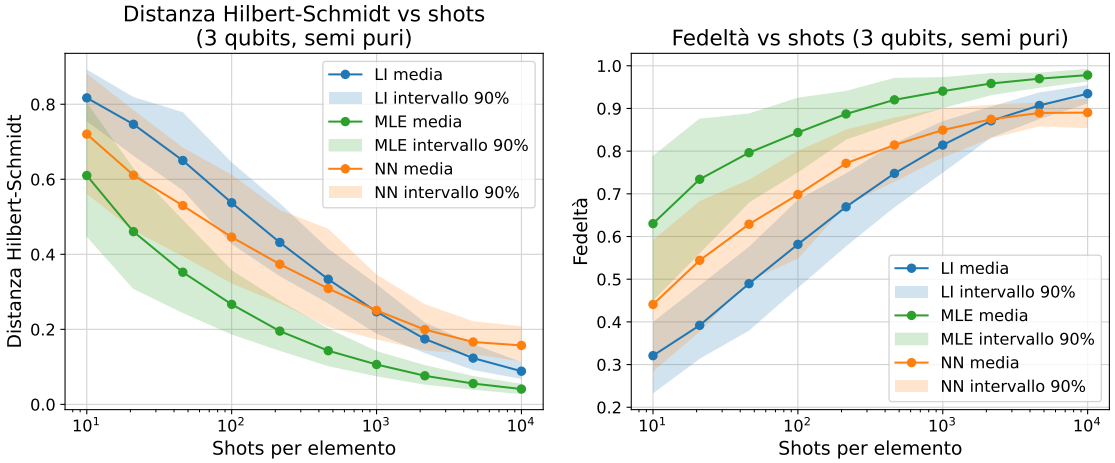

Figura 3.9: Qualità della ricostruzione per stati semi-puri di tre qubit rispetto al numero di misure, a sinistra è riportata la distanza di Hilbert-Schmidt, a destra la fedeltà.

Per stati puri la rete si comporta in maniera simile alla stima di massima verosimiglianza seppur con dispersione della fedeltà molto maggiore. Per stati semi-puri le performance della rete sono più vicine all'inversione lineare che alla MLE. Solo sugli stati misti si nota il vantaggio della rete rispetto ai metodi classici per basse misure (fino a ~ 500 shots per proiettore).

L'inaspettato vantaggio della rete sugli stati misti, sebbene inatteso, è spiegabile come segue. Il dataset di addestramento è generato nello stesso modo del dataset di test, ovvero con stati uniformemente distribuiti secondo l'ensemble di Hilbert-Schmidt. In questo caso è probabile che la rete abbia imparato a prevedere stati su quella distribuzione e quindi per un basso numero di misure batte i metodi classici.

Oltre alla qualità della ricostruzione si deve tenere conto anche di altri fattori, tra cui il costo computazionale e la quantità di dati.

L'ottimizzazione numerica usata nella stima di massima verosimiglianza è piut-

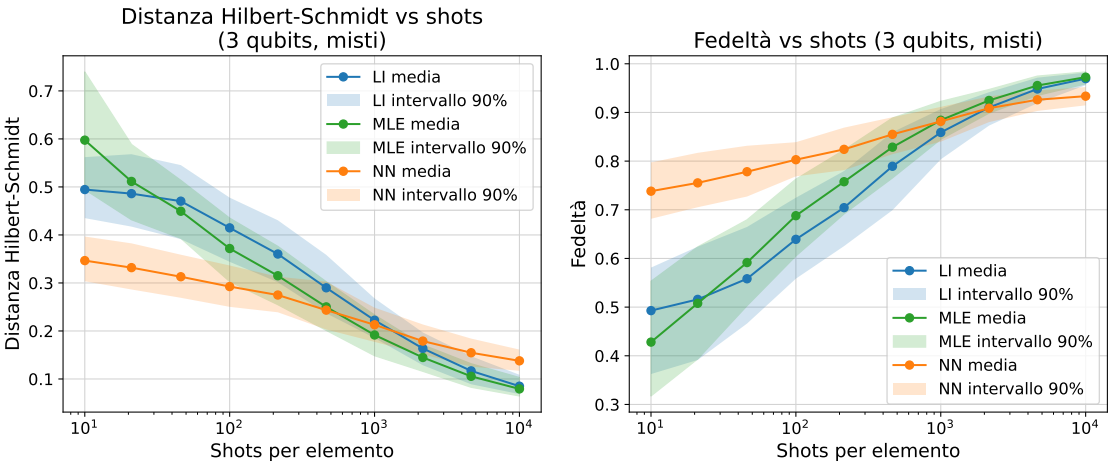

Figura 3.10: Qualità della ricostruzione per stati misti di tre qubit rispetto al numero di misure, a sinistra è riportata la distanza di Hilbert-Schmidt, a destra la fedeltà.

tosto dispendiosa, impiegando circa un secondo per ricostruzione. Anche le reti neurali seguono un processo di ottimizzazione, ma questo avviene solamente in allenamento. In compenso, il processo di inferenza è rapido, con tempi inferiori al millisecondo, confrontabili con quelli dell'inversione lineare. Sotto questo aspetto i dati hanno confermato la tendenza individuata da [13].

Un elemento critico nell'applicazione delle reti neurali risiede tuttavia nella quantità di dati necessaria all'addestramento. Per compensare l'espansione dello spazio dei parametri nel passaggio da due a tre qubit, la dimensione del dataset è stata quadruplicata, passando da 50 000 a 200 000 elementi. Pertanto, questa tendenza pone dubbi sulla scalabilità di tale metodo.

Conclusioni

Questa tesi si è sviluppata attorno al formalismo della teoria quantistica, finalizzata alla comprensione di alcuni algoritmi di tomografia dello stato quantistico, a partire dai più classici (inversione lineare e stima di massima verosimiglianza) e arrivando ai metodi di frontiera quali le reti neurali e le Classical Shadows. L'implementazione dei metodi per la ricostruzione completa dello stato svolta nel capitolo 3 è stata fondamentale per comprendere a fondo i punti di forza e le criticità di ciascuno.

L'inversione lineare, pur essendo il metodo più semplice ed economico, non garantisce la fisicità dello stato ricostruito: raggiungere una probabilità ragionevole ($\sim 90\%$) di ottenere una matrice densità valida richiede un numero di misure estremamente elevato, dell'ordine di 10^5–10^6 *shots* per proiettore.

Per costruzione, la stima di massima verosimiglianza risolve questo problema e si è dimostrata sistematicamente più accurata della LI, al prezzo però di un'ottimizzazione numerica non lineare relativamente costosa.

Riguardo alle reti neurali, le simulazioni nel caso di 2 qubit sono risultate piuttosto rappresentative e in linea con i risultati della letteratura, in particolare con l'articolo [13], al quale sono ispirate. Su stati puri e misti e nel regime di bassi *shots* è stata verificata la superiorità dell'approccio basato su reti neurali che sono riuscite a sfruttare conoscenze a priori sulla distribuzione degli stati per ottenere ricostruzioni migliori rispetto agli algoritmi classici. Il vantaggio della rete si è ridotto drasticamente passando a sistemi di tre qubit: le prestazioni sono rimaste competitive solo per stati misti e in un regime di poche misure, un risultato che è probabilmente da attribuire più alla somiglianza tra le distribuzioni di addestramento e di test che a un'effettiva capacità della rete di generalizzare. Questo suggerisce che l'architettura utilizzata, per quanto adeguata nel caso a due qubit, non scali automaticamente a sistemi di dimensione maggiore, necessitando di un'accurata progettazione e di una maggiore quantità di dati.

Le Classical Shadows, pur non incluse nel confronto numerico in quanto pensate per la stima efficiente di proprietà specifiche piuttosto che per la ricostruzione completa dello stato, restano una direzione promettente.

Possibili estensioni di questo lavoro riguardano: l'ottimizzazione dell'architettura e del regime di addestramento delle reti neurali per sistemi di dimensione maggiore; un'analisi più approfondita del comportamento delle reti neurali su stati semi-puri; l'approfondimento della scelta dell'insieme di misura e del suo effetto sul condizionamento della ricostruzione; e la validazione dei metodi su dati sperimentali reali.

Bibliografia

[1] P. A. M. Dirac. *The Principles of Quantum Mechanics*. 4. ed. (rev.), repr. International Series of Monographs on Physics 27. Oxford: Clarendon Press, Oxford University Press, 2010.

[2] John Von Neumann, Robert T. Beyer e Nicholas A. Wheeler. *Mathematical Foundations of Quantum Mechanics*. New edition. Princeton: Princeton University Press, 2018.

[3] Michael A. Nielsen e Isaac L. Chuang. *Quantum Computation and Quantum Information*. 10th anniversary ed. Cambridge ; New York: Cambridge University Press, 2010.

[4] Max Born. «Zur Quantenmechanik der StoSSvorgänge». In: *Zeitschrift für Physik* 37.12 (dic. 1926), pp. 863–867. DOI: 10.1007/BF01397477.

[5] Steven J. van Enk. *Mixed States and Pure States*. Lecture Notes. 2009. URL: https://pages.uoregon.edu/svanenk/solutions/Mixed_states.pdf.

[6] P. Krantz et al. «A Quantum Engineer's Guide to Superconducting Qubits». In: *Applied Physics Reviews* 6.2 (giu. 2019), p. 021318. DOI: 10.1063/1.5089550. arXiv: 1904.06560 [quant-ph].

[7] Colin D. Bruzewicz et al. «Trapped-Ion Quantum Computing: Progress and Challenges». In: *Applied Physics Reviews* 6.2 (mag. 2019), p. 021314. DOI: 10.1063/1.5088164. arXiv: 1904.04178 [quant-ph].

[8] Pieter Kok et al. «Linear Optical Quantum Computing with Photonic Qubits». In: *Reviews of Modern Physics* 79.1 (gen. 2007), pp. 135–174. DOI: 10.1103/RevModPhys.79.135. arXiv: quant-ph/0512071.

[9] Loïc Henriet et al. «Quantum Computing with Neutral Atoms». In: *Quantum* 4 (set. 2020), p. 327. DOI: 10.22331/q-2020-09-21-327.

[10] Marcus W. Doherty et al. «The Nitrogen-Vacancy Colour Centre in Diamond». In: *Physics Reports*. The Nitrogen-Vacancy Colour Centre in Diamond 528.1 (lug. 2013), pp. 1–45. DOI: 10.1016/j.physrep.2013.02.001. arXiv: 1302.3288 [cond-mat.mtrl-sci].

[11] K Banaszek, M Cramer e D Gross. «Focus on Quantum Tomography». In: *New Journal of Physics* 15.12 (dic. 2013), p. 125020. DOI: 10.1088/1367-2630/15/12/125020.

[12] Matteo Paris e Jaroslav eháek, cur. *Quantum State Estimation*. Vol. 649. Lecture Notes in Physics. Berlin, Heidelberg: Springer, 2004. DOI: 10.1007/b98673.

[13] Dominik Koutný et al. «Neural-Network Quantum State Tomography». In: *Physical Review A* 106.1 (lug. 2022), p. 012409. DOI: 10.1103/PhysRevA.106.012409.

[14] Scott Aaronson. *Shadow Tomography of Quantum States*. Nov. 2017. URL: https://arxiv.org/abs/1711.01053v2.

[15] Franz J. Schreiber, Jens Eisert e Johannes Jakob Meyer. «Tomography of Parametrized Quantum States». In: *PRX Quantum* 6.2 (giu. 2025), p. 020346. DOI: 10.1103/PRXQuantum.6.020346. arXiv: 2407.12916 [quant-ph].

[16] Daniel F. V. James et al. «Measurement of Qubits». In: *Physical Review A* 64.5 (ott. 2001), p. 052312. DOI: 10.1103/PhysRevA.64.052312.

[17] A. J. Scott. «Tight Informationally Complete Quantum Measurements». In: *Journal of Physics A: Mathematical and General* 39.43 (ott. 2006), pp. 13507–13530. DOI: 10.1088/0305-4470/39/43/009.

[18] Paul Hausladen e William K. Wootters. «A 'Pretty Good' Measurement for Distinguishing Quantum States». In: *Journal of Modern Optics* (1994). DOI: 10.1080/09500349414552221.

[19] Ian Goodfellow, Yoshua Bengio e Aaron Courville. *Deep Learning*. MIT Press. URL: https://www.deeplearningbook.org/.

[20] Diederik P. Kingma e Jimmy Ba. *Adam: A Method for Stochastic Optimization*. Gen. 2017. DOI: 10.48550/arXiv.1412.6980. arXiv: 1412.6980 [cs.LG].

[21] Hsin-Yuan Huang, Richard Kueng e John Preskill. *Predicting Many Properties of a Quantum System from Very Few Measurements*. Feb. 2020. DOI: 10.1038/s41567-020-0932-7.

[22] Andreas Elben et al. «The Randomized Measurement Toolbox». In: *Nature Reviews Physics* 5.1 (gen. 2023), pp. 9–24. DOI: 10.1038/s42254-022-00535-2.

[23] Senrui Chen et al. «Robust Shadow Estimation». In: *PRX Quantum* 2.3 (set. 2021), p. 030348. DOI: 10.1103/PRXQuantum.2.030348.

[24] Hsin-Yuan Huang, Richard Kueng e John Preskill. «Efficient Estimation of Pauli Observables by Derandomization». In: *Physical Review Letters* 127.3 (lug. 2021), p. 030503. DOI: 10.1103/PhysRevLett.127.030503.

[25] Karol yczkowski et al. «Generating Random Density Matrices». In: *Journal of Mathematical Physics* 52.6 (giu. 2011), p. 062201. DOI: 10.1063/1.3595693.

[26] M. J. D. Powell. «An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives». In: *The Computer Journal* 7.2 (gen. 1964), pp. 155–162. DOI: 10.1093/comjnl/7.2.155.

[27] Jason Ansel et al. «PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation». In: *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*. ACM Conferences. Apr. 2024, pp. 929–947. DOI: 10.1145/3620665.3640366.