Quantum Computation, Quantum Error Correcting Codes and Information Theory

K R Parthasarathy

TIFR
Mumbai

Narosa

Lectures on
Quantum Computation, Quantum Error Correcting Codes and Information Theory

K.R. Parthasarathy

Published for the
Tata Institute of Fundamental Research

International Distribution by
American Mathematical Society

Narosa Publishing House
New Delhi Chennai Mumbai Kolkata

K.R. Parthasarathy
Indian Statistical Institute
Delhi, India

NAROSA PUBLISHING HOUSE PVT. LTD.

22 Daryaganj, Delhi Medical Association Road, New Delhi 110 002
35-36 Greams Road, Thousand Lights, Chennai 600 006
306 Shiv Centre, D.B.C. Sector 17, K.U. Bazar P.O., Navi Mumbai 400 703
2F-2G Shivam Chambers, 53 Syed Amir Ali Avenue, Kolkata 700 019

www.narosa.com

Printed from the camera-ready copy provided by the Authors

ISBN 81-7319-688-5

Published by N.K. Mehra for Narosa Publishing House Pvt. Ltd.,
22 Daryaganj, Delhi Medical Association Road, New Delhi 110 002
and printed at Rajkamal Press, New Delhi 110 033, India.

Lectures on
Quantum Computation, Quantum Error Correcting Codes and Information Theory

Preface

These notes were prepared by Amitava Bhattacharya on a course of lectures I gave at the Tata Institute of Fundamental Research (Mumbai) in the months of April 2001 and February 2002. I am grateful to my colleagues at the TIFR, in general, and Professor Parimala Raman in particular, for providing me a receptive and enthusiastic audience and showering on me their warm hospitality. I thank Professor Jaikumar for his valuable criticism and insight and several fruitful conversations in enhancing my understanding of the subject. Finally, I express my warm appreciation of the tremendous effort put in by Amitava Bhattacharya for the preparation of these notes and their LaTeX files.

Financial support from the Indian National Science Academy (in the form of C. V. Raman Professorship), TIFR (Mumbai) and Indian Statistical Institute (Delhi Centre) is gratefully acknowledged.

K. R. Parthasarathy
Delhi

Contents

Lecture 1

Quantum Probability

In the Mathematical Congress held at Berlin, Peter Shor presented a new algorithm for factoring numbers on a *quantum computer*. In this series of lectures, we shall study the areas of quantum computation (including Shor's algorithm), quantum error correcting codes and quantum information theory.

1.1 Classical Versus Quantum Probability Theory

We begin by comparing classical probability and quantum probability. In classical probability theory (since Kolmogorov's 1933 monograph [11]), we have a sample space, a set of events, a set of random variables, and distributions. In quantum probability (as formulated in von Neumann's 1932 book [14]), we have a state space (which is a Hilbert space) instead of a sample space; events, random variables and distributions are then represented as operators on this space. We now recall the definitions of these notions in classical probability and formally define the analogous concepts in quantum probability. In our discussion we will be concerned only with *finite* classical probability spaces, and their quantum analogues—finite dimensional Hilbert spaces.

Spaces	
1.1 The sample space Ω: This is a finite set, say $\{1, 2, \ldots, n\}$.	**1.2 The state space H:** It is a complex Hilbert space of dimension n.

Events	
1.3 The set of events $\mathcal{F}_\Omega$: This is the set of all subsets of Ω. $\mathcal{F}_\Omega$ is a Boolean algebra with the *union* ($\cup$) operation for 'or' and the *intersection* ($\cap$) operation for 'and'. In particular, we have $E \cap (F_1 \cup F_2) = (E \cap F_1) \cup (E \cap F_2)$.	**1.4 The set of events $\mathcal{P}(H)$:** This is the set of all orthogonal projections in $\mathcal{H}$. An element $E \in \mathcal{P}(H)$ is called an *event*. Here, instead of '$\cup$' we have the max ($\vee$) operation, and instead of '$\cap$' the min ($\wedge$) operation. Note, however, that $E \wedge (F_1 \vee F_2)$ is not always equal to $(E \wedge F_1) \vee (E \wedge F_2)$. (They are equal if E, F_1, F_2 commute with each other).

Random variables and observables	
1.5 The set of random variables $\mathcal{B}_\Omega$: This is the set of all complex valued functions on Ω. The elements of $\mathcal{B}_\Omega$ are called *random variables*. $\mathcal{B}_\Omega$ is an Abelian C^*-algebra under the operations $(\alpha f)(\omega) = \alpha f(\omega)$; $(f+g)(\omega) = f(\omega) + g(\omega)$; $(f \cdot g)(\omega) = f(\omega) g(\omega)$; $f^*(\omega) \triangleq f^\dagger(\omega) = f(\bar{\omega})$. Here, $\alpha \in \mathbb{C}$, $f, g \in \mathcal{B}_\Omega$, and the 'bar' stands for complex conjugation. The random variable $\mathbf{1}$ (defined by $\mathbf{1}(\omega) \triangleq 1$), is the unit in this algebra.	**1.6 The set of observables $\mathcal{B}(\mathcal{H})$:** This is the (non-Abelian) C^*-algebra of all operators on $\mathcal{H}$, with '$+$' and '$\cdot$' defined as usual, and X^* defined to be the adjoint of X. We will use $X^\dagger$ instead of X^*. The identity projection I is the unit in this algebra. We say that an observable is real-valued if $X^\dagger = X$, that is, if X is Hermitian. For such an observable, we define $\mathrm{Sp}(X)$ to be the set of eigen values of X. Since X is Hermitian, $\mathrm{Sp}(X) \subseteq \mathbb{R}$, and by the spectral theorem, we can write X as $X = \sum_{\lambda \in \mathrm{Sp}(X)} \lambda E_\lambda$,

With each event $E \in \mathcal{F}_\Omega$ we associate the indicator random variable $\mathbf{1}_E$ defined by

$$\mathbf{1}_E(\omega) = \begin{cases} 1 & \text{if } \omega \in E; \\ 0 & \text{otherwise.} \end{cases}$$

For a random variable f, let $\mathrm{Sp}(f) \triangleq f(\Omega)$. Then, f can be written as the following linear combination of indicator random variables:

$$f = \sum_{\lambda \in \mathrm{Sp}(f)} \lambda \mathbf{1}_{f^{-1}(\{\lambda\})},$$

so that

$$\mathbf{1}_{f^{-1}(\{\lambda\})} \cdot \mathbf{1}_{f^{-1}(\{\lambda'\})} = \mathbf{0} \text{ for } \lambda \neq \lambda';$$
$$\sum_{\lambda \in \mathrm{Sp}(f)} \mathbf{1}_{f^{-1}(\{\lambda\})} = \mathbf{1}.$$

Similarly, we have

$$f^r = \sum_{\lambda \in \mathrm{Sp}(f)} \lambda^r \mathbf{1}_{f^{-1}(\{\lambda\})},$$

and, in general, for a function $\varphi : \mathbb{C} \to \mathbb{C}$, we have the random variable

$$\varphi(f) = \sum_{\lambda \in \mathrm{Sp}(f)} \varphi(\lambda) \mathbf{1}_{f^{-1}(\{\lambda\})}.$$

Later, we will be mainly interested in real-valued random variables, that is random variables f with $\mathrm{Sp}(f) \subseteq \mathbb{R}$ (or $f^\dagger = f$).

where E_λ is the projection on the subspace $\{u : Xu = \lambda u\}$ and

$$E_\lambda E_{\lambda'} = \mathbf{0}, \lambda, \lambda' \in \mathrm{Sp}(X), \lambda \neq \lambda';$$
$$\sum_{\lambda \in \mathrm{Sp}(X)} E_\lambda = I.$$

Similarly, we have

$$X^r = \sum_{\lambda \in \mathrm{Sp}(X)} \lambda^r E_\lambda,$$

and in general, for a function $\varphi : \mathbb{R} \to \mathbb{R}$, we have

$$\varphi(X) = \sum_{\lambda \in \mathrm{Sp}(X)} \varphi(\lambda) E_\lambda.$$

Distributions and states

1.7 A distribution p: This is a function from $\mathcal{F}_\Omega$ to $\mathbb{R}$, determined by n real numbers $p_1, p_2, \ldots, p_n$, satisfying:

$$\begin{aligned} p_i &\geq 0; \\ \sum_{i=1}^{n} p_i &= 1. \end{aligned}$$

The probability of the event $E \in \mathcal{F}_\Omega$ (under the distribution $\boldsymbol{p}$) is

$$\Pr(E; \boldsymbol{p}) \triangleq \sum_{i \in E} p_i.$$

When there is no confusion we write $\Pr(E)$ instead of $\Pr(E; \boldsymbol{p})$. We will identify $\boldsymbol{p}$ with the sequence $(p_1, p_2, \ldots, p_n)$. The probability that a random variable f takes the value $\lambda \in R$ is

$$\Pr(f = \lambda) \triangleq \Pr(f^{-1}(\{\lambda\}));$$

thus, a real-valued random variable f has a distribution on the real line with mass $\Pr(f^{-1}(\{\lambda\}))$ at $\lambda \in \mathbb{R}$.

1.8 A state ρ: In quantum probability, we have a state ρ instead of the distribution $\boldsymbol{p}$. A state is a non-negative definite operator on $\mathcal{H}$ with $\operatorname{Tr}\rho = 1$. The *probability* of the event $E \in \mathcal{P}(H)$ in the state ρ is defined to be $\operatorname{Tr}\rho E$, and the probability that the real-valued observable X takes the value λ is

$$\Pr(X = \lambda) = \begin{cases} \operatorname{Tr}\rho E_\lambda \text{ if } \lambda \in \operatorname{Sp}(X); \\ 0 \qquad \text{otherwise.} \end{cases}$$

Thus, a real-valued observable X has a distribution on the real line with mass $\operatorname{Tr}\rho E_\lambda$ at $\lambda \in \mathbb{R}$.

Expectation, moments, variance

The expectation of a random variable f is

$$\mathop{\mathbb{E}}_{\boldsymbol{p}} f \triangleq \sum_{\omega \in \Omega} f(\omega) p_\omega.$$

The r-*th* moment of f is the expectation of f^r, that is

The expectation of an observable X in the state ρ is

$$\mathop{\mathbb{E}}_{\rho} X \triangleq \operatorname{Tr}\rho X.$$

The map $X \mapsto \mathbb{E}_\rho X$ has the following properties:

$$\begin{aligned}\mathop{\mathbb{E}}_{p} f^r &= \sum_{\omega\in\Omega}(f(\omega))^r p_\omega \\ &= \sum_{\lambda\in \mathrm{Sp}(f)} \lambda^r \Pr(f^{-1}(\lambda)),\end{aligned}$$

and the *characteristic function* of f is the expectation of the complex-valued random variable e^{itf}, that is,

$$\mathop{\mathbb{E}}_{p} e^{itf} = \sum_{\lambda\in \mathrm{Sp}(f)} e^{it\lambda} \Pr(f^{-1}(\lambda)).$$

The variance of a real-valued random variable f is

$$\mathbf{var}(f) \stackrel{\Delta}{=} \mathop{\mathbb{E}}_{p}(f - \mathop{\mathbb{E}}_{p} f)^2 \geq 0.$$

Note that

$$\mathbf{var}(f) = \mathop{\mathbb{E}}_{p} f^2 - (\mathop{\mathbb{E}}_{p} f)^2;$$

also, $\mathbf{var}(f) = 0$ if and only if all the mass in the distribution of f is concentrated at $\mathbb{E}_p f$.

(1) It is linear;

(2) $\mathbb{E}_\rho X^\dagger X \geq 0$, for all $X \in \mathcal{B}(\mathcal{H})$.

(3) $\mathbb{E}_\rho I = 1$.

The *r-th moment* of X is the expectation of X^r; if X is real-valued, then using the spectral decomposition, we can write

$$\mathop{\mathbb{E}}_{\rho} X^r = \sum_{\lambda\in \mathrm{Sp}(X)} \lambda^r \operatorname{Tr} \rho E_\lambda.$$

The *characteristic function* of the real-valued observable X is the expectation of the observable e^{itX}. The variance of a (real-valued) observable X is

$$\begin{aligned}\mathbf{var}(X) &\stackrel{\Delta}{=} \operatorname{Tr} \rho(X - \operatorname{Tr} \rho X)^2 \\ &= \operatorname{Tr} \rho X^2 - (\operatorname{Tr} \rho X)^2 \\ &\geq 0.\end{aligned}$$

The variance of X vanishes if and only if the distribution of X is concentrated at the point $\operatorname{Tr} \rho X$. This is equivalent to the property that the operator range of ρ is contained in the eigensubspace of X with eigenvalue $\operatorname{Tr} \rho X$.

Extreme points

1.9 The set of distributions: The set of all probability distributions on Ω is a compact convex set (Choquet simplex) with exactly n extreme points, δ_j $(j = 1, 2, \ldots, n)$, where δ_j is determined by

1.10 The set of states: The set of all states in $\mathcal{H}$ is a convex set. Let ρ be a state. Since ρ is non-negative definite, its eigen values are non-negative reals, and we can write

$$\delta_j(\{\omega\}) \triangleq \begin{cases} 1 & \text{if } \omega = j; \\ 0 & \text{otherwise.} \end{cases}$$

If $P = \delta_j$, then every random variable has a degenerate distribution under P: the distribution of the random variable f is concentrated on the point $f(j)$.

$$\rho = \sum_{\lambda \in \mathrm{Sp}(\rho)} \lambda E_\lambda;$$

since $\mathrm{Tr}\,\rho = 1$, we have

$$\sum_{\lambda \in \mathrm{Sp}(\rho)} \lambda \times \dim(E_\lambda) = 1.$$

The projection E_λ can, in turn, be written as a sum of *one-dimensional* projections:

$$E_\lambda = \sum_{i=1}^{\dim(E_\lambda)} E_{\lambda,i}.$$

Then,

$$\rho = \sum_{\lambda \in \mathrm{Sp}(\rho)} \sum_{i=1}^{\dim(E_\lambda)} \lambda E_{\lambda,i}.$$

Proposition 1.1.1 *A one-dimensional projection cannot be written as a non-trivial convex combination of states.*

Thus, the extreme points of the convex set of states are precisely the one-dimensional projections. Let ρ be the extreme state corresponding to the one-dimensional projection on the ray $\mathbb{C}u$ (where $\|u\| = 1$). Then, the expectation m of the observable X is

$$m = \mathrm{Tr}\, uu^\dagger X = \mathrm{Tr}\, u^\dagger X u = \langle u, Xu \rangle,$$

and

$$\begin{aligned} \mathbf{var}(X) &= \mathrm{Tr}\, uu^\dagger (X-m)^2 \\ &= \mathrm{Tr}\, \|(X-m)u\|^2. \end{aligned}$$

	Thus, $\mathbf{var}(X) = 0$ if and only if u is an eigenvector of X. So, even for this extreme state, not all observables have degenerate distributions: *degeneracy of the state does not kill the uncertainty of the observables!*

The product	
1.11 Product spaces: If there are two statistical systems described by classical probability spaces $(\Omega_1, \boldsymbol{p}_1)$ and $(\Omega_2, \boldsymbol{p}_2)$ respectively, then the probability space $(\Omega_1 \times \Omega_2, \boldsymbol{p}_1 \times \boldsymbol{p}_2)$ determined by $$\Pr(\{(i,j)\}; \boldsymbol{p}_1 \times \boldsymbol{p}_2) \triangleq \Pr(\{i\}; \boldsymbol{p}_1) \Pr(\{j\}; \boldsymbol{p}_2),$$ describes the two independent systems as a single system.	**1.12 Product spaces:** If $(\mathcal{H}_1, \rho_1)$ and $(\mathcal{H}_2, \rho_2)$ are two quantum systems, then the quantum system with state space $\mathcal{H}_1 \otimes \mathcal{H}_2$ and state $\rho_1 \otimes \rho_2$ (which is a non-negative definite operator of unit trace on $\mathcal{H}_1 \otimes \mathcal{H}_2$) describes the two independent quantum systems as a single system.
Dynamics	
1.13 Reversible dynamics in Ω: This is determined by a bijective transformation $T : \Omega \to \Omega$. Then, $f \rightsquigarrow f \circ T$ (for random variables) $P \rightsquigarrow P \circ T^{-1}$ (for distributions)	**1.14 Reversible dynamics in $\mathcal{H}$:** This is determined by a unitary operator $U : \mathcal{H} \to \mathcal{H}$. Then, we have the dynamics of Heisenberg: $X \rightsquigarrow U^\dagger X U$ for $X \in \mathcal{B}(\mathcal{H})$; Schrödinger $\rho \rightsquigarrow U \rho U^\dagger$ for the state ρ.

1.2 Three Distinguishing Features

We now state the first distinguishing feature.

Proposition 1.2.1 *Let E and F be projections in $\mathcal{H}$ such that $EF \neq FE$. Then, $E \vee F \leq E + F$ is false.*

Proof Suppose $E \vee F \leq E + F$. Then, $E \vee F - E \leq F$. So,

$$F(E \vee F - E) = (E \vee F - E)F.$$

That is, $FE = EF$, a contradiction.

□

Corollary 1.2.2 *Suppose E and F are projections such that $EF \neq FE$. Then, for some state ρ, the inequality $\operatorname{Tr} \rho(E \vee F) \leq \operatorname{Tr} \rho E + \operatorname{Tr} \rho F$ is false.*

Proof By the above proposition, $E \vee F \leq E + F$ is false; that is, there exists a unit vector u such that

$$\langle u, (E \vee F)u \rangle \not\leq \langle u, Eu \rangle + \langle u, Fu \rangle .$$

Choose ρ to be the one dimensional projection on the ray $\mathbb{C}u$. Then,

$$\begin{aligned} \operatorname{Tr}(E \vee F)\rho &= \langle u, (E \vee F)u \rangle , \\ \operatorname{Tr} E\rho &= \langle u, Eu \rangle , \\ \operatorname{Tr} F\rho &= \langle u, Fu \rangle . \end{aligned}$$

□

The second distinguishing feature is:

Proposition 1.2.3 (Heisenberg's inequality) *Let X and Y be observables and let ρ be a state in $\mathcal{H}$. Assume $\operatorname{Tr} \rho X = \operatorname{Tr} \rho Y = 0$. Then,*

$$\begin{aligned} \operatorname*{\mathbf{var}}_{\rho}(X) \operatorname*{\mathbf{var}}_{\rho}(Y) &\geq \left(\operatorname{Tr} \rho \frac{1}{2}\{X, Y\} \right)^2 + \left(\operatorname{Tr} \rho \frac{1}{2} i\,[X, Y] \right)^2 \\ &\geq \frac{1}{4}(\operatorname{Tr} \rho\, i\,[X, Y])^2, \end{aligned}$$

where

$$\begin{aligned} \{X, Y\} &\triangleq XY + YX; \text{ and} \\ [X, Y] &\triangleq XY - YX. \end{aligned}$$

Proof For $z \in \mathbb{C}$, we have

$$\operatorname{Tr} \rho(X+zY)^\dagger(X+zY) \geq 0.$$

If $z = re^{i\theta}$,

$$r^2 \operatorname{Tr} \rho Y^2 + 2r\Re e^{-i\theta} \operatorname{Tr} \rho YX + \operatorname{Tr} \rho X^2 \geq 0.$$

The left hand side is a degree-two polynomial in the variable r. Since, it is always non-negative, it can have at most one root. Thus, for all θ,

$$\begin{aligned}(\operatorname{Tr} \rho X^2)(\operatorname{Tr} \rho Y^2) &\geq (\Re e^{-i\theta} \operatorname{Tr} \rho YX)^2 \\ &\geq \left(\cos\theta \operatorname{Tr} \rho \frac{XY+YX}{2} + \sin\theta \operatorname{Tr} \rho i \frac{XY-YX}{2}\right)^2 \\ &= (x\cos\theta + y\sin\theta)^2,\end{aligned}$$

where $x \stackrel{\Delta}{=} \operatorname{Tr} \rho \, \frac{1}{2}\{X,Y\}$ and $y \stackrel{\Delta}{=} \operatorname{Tr} \rho \, \frac{i}{2}[X,Y]$. Note that the right hand side is maximum when $\cos\theta = \frac{x}{\sqrt{x^2+y^2}}$ and $\sin\theta = \frac{y}{\sqrt{x^2+y^2}}$ and the proposition follows.

□

Now we state the third distinguishing feature:

Extremal states (one-dimensional projections) are called *pure states.* The set of all pure states in an n-dimensional complex Hilbert space is a manifold of dimension $2n-2$. (The set of all extremal probability distributions on a sample space of n points has cardinality n).

1.3 Measurements: von Neumann's Collapse Postulate

Suppose X is an observable (i.e. a Hermitian operator) with spectral decomposition

$$X = \sum_{\lambda \in \mathrm{Sp}(X)} \lambda E_\lambda.$$

Then, the measurement of X in the quantum state ρ yields the value λ with probability $\operatorname{Tr} \rho E_\lambda$. If the observed value is λ, then the state collapses to

$$\tilde{\rho}_\lambda = \frac{E_\lambda \rho E_\lambda}{\operatorname{Tr} \rho E_\lambda}.$$

The collapsed state $\tilde{\rho}_\lambda$ has its support in the subspace $E_\lambda(\mathcal{H})$.

1.4 Dirac Notation

Elements of the Hilbert space $\mathcal{H}$ are called *ket vectors* and denoted by $|u\rangle$. Elements of the dual space $\mathcal{H}^*$ are called *bra vectors* and denoted by $\langle u|$. The bra $\langle u|$ evaluated on the ket $|v\rangle$ is the bracket $\langle u \mid v\rangle$, the scalar product between u, v as elements of $\mathcal{H}$.

The operator $|u\rangle\langle v|$ is defined by

$$|u\rangle\langle v|(|w\rangle) \triangleq \langle v \mid w\rangle\, |u\rangle.$$

It is a rank one operator when u and v are non-zero.

$$\begin{aligned}
\operatorname{Tr} |u\rangle\langle v| &= \langle v \mid u\rangle \\
(|u\rangle\langle v|)^\dagger &= |v\rangle\langle u| \\
|u_1\rangle\langle v_1||u_2\rangle\langle v_2| \cdots |u_n\rangle\langle v_n| &= (\langle v_1 \mid u_2\rangle \langle v_2 \mid u_3\rangle \cdots \langle v_{n-1} \mid u_n\rangle)|u_1\rangle\langle v_n|.
\end{aligned}$$

The scalar product $\langle u \mid v\rangle$ is anti-linear (conjugate-linear) in the first variable and linear in the second variable.

1.4.1 Qubits

The Hilbert space $\mathbf{h} \triangleq \mathbb{C}^2$, with scalar product $\left\langle \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right], \left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right] \right\rangle = \bar{a}c + \bar{b}d$, is called a 1-*qubit Hilbert space*. Let

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Then,

$$\begin{bmatrix} a \\ b \end{bmatrix} = a|0\rangle + b|1\rangle,$$

and the ket vectors $|0\rangle$ and $|1\rangle$ form an orthonormal basis for $\mathbf{h}$.

The Hilbert space $\mathbf{h}^{\otimes n} = (\mathbb{C}^2)^{\otimes n}$ is called the n-*qubit* Hilbert space. If $x_1 x_2 \cdots x_n$ is an n-length word from the binary alphabet $\{0, 1\}$, we let

$$\begin{aligned}
|x_1 x_2 \cdots x_n\rangle &\triangleq |x_1\rangle|x_2\rangle \cdots |x_n\rangle \\
&\triangleq |x_1\rangle \otimes |x_2\rangle \otimes \cdots \otimes |x_n\rangle \\
&\triangleq |\mathbf{x}\rangle,
\end{aligned}$$

where $\mathbf{x} = x_1 \times 2^{n-1} + x_2 \times 2^{n-2} + \cdots + x_{n-1} \times 2 + x_n$ (that is, as $x_1 x_2 \ldots x_n$ varies over all n-length words, the integer $\mathbf{x}$ varies in the range $\{0, 1, \ldots, 2^n - 1\}$).