A Game Theoretic Approach to Quantum Information

Xianhua Dai, MSci.

Thesis submitted to The University of Nottingham
for the degree of Doctor of Philosophy

March 2008

Abstract

In this project, bridging entropy econometrics, game theory and information theory, a game theoretic approach will be investigated to quantum information, during which new mathematical definitions for quantum relative entropy, quantum mutual information, and quantum channel capacity will be given and monotonicity of entangled quantum relative entropy and additivity of quantum channel capacity will be obtained rigorously in mathematics; also quantum state will be explored in Kelly criterion, during which game theoretic interpretations will be given to classical relative entropy, mutual information, and asymptotical information.

In specific, after briefly introducing probability inequalities, C^*-algebra, including von Neumann algebra, and quantum probability, we will overview quantum entanglement, quantum relative entropy, quantum mutual information, and entangled quantum channel capacity in the direction of R. L. Stratonovich and V. P. Belavkin, and upon the monotonicity property of quantum mutual information of Araki-Umegaki type and Belavkin-Staszewski type, we will prove the additivity property of entangled quantum channel capacities, extending the results of V. P. Belavkin to products of arbitrary quantum channel to quantum relative entropy of both Araki-Umegaki type and Belavkin-Staszewski type.

We will obtain a sufficient condition for minimax theorem in an introduction to strategic game, after which, in the exploration of classical/quantum estimate (here we still use the terminology of quantum estimate in the sense of game theory in accordance to classical estimate, but NOT in the sense of quantum physics or quantum probability), we will find the existence of the minimax value of this game and its minimax strategy, and applying biconjugation in convex analysis, we will arrive at one new approach to quantum relative entropy, quantum mutual entropy, and quantum channel capacity, in the sense, independent on Radon-Nikodym derivative, also the monotonicity of quantum relative entropy and the additivity of quantum communication channel capacity will be

obtained.

Applying Kelly's criterion, we will give a practical game theoretic interpretation, in the model to identify quantum state, to relative entropy, mutual information, and asymptotical information, during which we will find that the decrement in the doubling rate achieved with true knowledge of the distribution F over that achieved with incorrect knowledge G is bounded by relative entropy $\mathcal{R}(F;G)$ of F relative to G; the increment Δ in the doubling rate resulting from side information Y is less than or equal to the mutual information $\mathcal{I}(X,Y)$; a good sequence to identify the true quantum state leads to asymptotically optimal growth rate of utility; and applying the asymptotic behavior of classical relative entropy, the utility of the Bayes' strategy will be bounded below in terms of the optimal utility.

The first two main parts are to extend classical entropy econometrics, in the view of strategic game theory, to non-commutative data, for example, quantum data in physical implementation, while the third main part is to intrinsically and practically give a game theoretic interpretation of classical relative entropy, mutual information, and asymptotical information, in the model to identify quantum state, upon which a pregnant financial stock may be designed, which may be called "quantum" stock, for its physical implementation.

Acknowledgements

First I am most grateful to Prof. Slava Belavkin who cut out this beautiful project "a Game Theoretic Approach to Quantum Information". It was great fun to have many enthusiastic brain storms with Slava, on many subjects. I learned a lot from him setting free on its way to seek for ideas. Also it was with Slava that I first experienced the fascinating process of a new scientific idea coming to life. I also thank him for a careful reading of this thesis. Most of this thesis is owed to him, either directly or indirectly.

Secondly I want to thank two Annual Assessors, Dr. G. Tanner and Dr. J. M. Louko, for their kind advice on progress arrangement and thesis format.

I want to thank the University of Nottingham, for offering me the PhD position, Madam Chern Scholarship and University Research Scholarship during these years that directly resulted in this thesis, also Newton Institute for Mathematical Sciences, Cambridge, for funding me a junior membership.

Furthermore, it is a good chance to thank Max-Planck Institute for Mathematics in Bonn, German where I benefit from Prof. Yuri Manin and Prof. Matilde Marcolli before coming to Nottingham.

It is my pleasure to thank all colleagues in game theory, information theory and surrounding topics. It will be impossible to mention all, either for their original works, for sharing knowledge with me, or answering my questions. Just some are Prof. Andrew McLennan, Prof. Ariel Rubinstein and Prof. A. S. Holevo, Prof. Peter Shor.

Finally, I want to thank my wife for the love and sincere encouragement, and my parents for the love and freedom with me for life.

Contents

CHAPTER 1

Introduction

Some topics of game theory and information theory are fundamentally the same, for example, product distribution, in the information theoretic mathematical structure, provided a unified formulation of bounded rationality in game theory and a set of new types of mean field theory in statistical physics, as noted in [99], for example.

This project, bridging entropy econometrics, game theory, and information theory, is to investigate quantum information in a game theoretic approach, mainly on quantum relative entropy, quantum mutual information, quantum channel capacity and quantum state identification, critically speaking, extending entropy econometrics to noncommutative data, new in the following aspects:

1) At least as we know, this is the first, in working method, to formally apply classical game theory to quantum information, say, considering quantum relative entropy as one-spot game cost, which extended classical case (so as to obtain a new mathematical form of quantum information), and applying Kelly criterion to identify state instead of minimax criterion or others (to give a game theoretic interpretation of classical information theory).

2) Also this is the first to apply convex analysis to quantum information, approaching one new mathematical form of quantum relative entropy, though convex analysis was extensively applied in various classical subjects.

3) Our mathematical form of quantum relative entropy, corresponding quantum mutual information, and quantum channel capacity, is compact, much different from Araki-Umegaki type, Belavkin-Staszewski type, etc., independent on Radon-Nikodym derivative.

4) Simultaneously two important properties in quantum information theory, i.e., monotonicity of quantum relative entropy and additivity of quantum chan-

nel capacity, will be obtained rigorously, in mathematics, in our context, which extended the results of V. P. Belavkin to products of arbitrary quantum channel to quantum relative entropy of both Araki-Umegaki type and Belavkin-Staszewski type.

5) This project begins extending classical entropy econometrics to noncommutative data, for example, quantum data in physical implementation, though classical entropy econometrics was extensively explored and applied in various classical subjects.

6) This project intrinsically gives a practical game theoretic interpretation of classical relative entropy, mutual information, and asymptotical information, in the model to identify quantum state, upon which a pregnant financial stock may be designed potentially, or suitably "quantum" stock, for its physical implementation.

All in all, this project is closely motivated by the following subjects:

1) Entropy Econometrics

Information and entropy econometrics deals with statistical inference of problems, given incomplete knowledge or data, as well as the analysis, diagnostics and statistical properties of information measures.

Maximum entropy developed in two lines:

(1) To understand the general characteristics (distribution) of a system from partial and incomplete information, statistical inference was intensively investigated by Bernoulli, Bayes, Laplace, Jeffreys, Cox.

(2) To assign (initial) numerical values of probabilities, when only some (theoretical) limited global quantities of the investigated system are known, statistical modeling of problems was explored by Maxwell, Boltzmann, Gibbs, Shannon in mechanics, physics and information.

Recognizing the common basic objectives of these two lines, [48, 49] formulated the maximum entropy, dependent on the philosophy of the first line and the mathematics of the second line.

On the basic concepts and properties of information theory, Kullback and Leibler developed some fundamental statistics, for example, sufficiency and efficiency as well as a generalization of the Cramer-Rao inequality, which unify heterogeneous statistical procedures via the concepts of information theory [59–61].

[67], on the other hand, interpreted a statistical sample as a noisy communica-

tion channel, which conveys a message about a parameter according to a prior distribution. Also [67] applied Shannon's ideas to statistical theory by referring to the information in an experiment rather than in a message. [100] established Bayesian theorem as the optimal learning rule for information process based on logarithmic information measures.

On optimizing a certain informational-objective function, subject to certain moment representation of the data, or certain "conservation laws" representing the underlying system, some works pioneered [101–106], among which the same basic objective is analyzing limited and noisy data using minimal assumptions.

Once the underlying data generating process (or error margins) is uncertain or unknown, to avoid strong distributional assumptions or a pre-specified likelihood function, it seems inevitable, for a large class of models (linear and nonlinear, parametric and non-parametric), back to the foundations of information theory and maximum entropy and led to the above information-theoretic methods, which could be viewed as approaches to solving under-determined problems in the sense that without a pre-specified likelihood or distribution, there are always more unknowns than knowns regardless of the amount of data.

As information and entropy econometrics methods are used in conjunction, they are powerful in analyzing a wide variety of problems in most disciplines of science, for example, empirical likelihood and the generalized method of moments type methods in image reconstruction, spectral analysis in communication and information, operations research, and economics, statistical inference and estimation (Bayesian and non Bayesian methods) in information processing and information theory.

Despite these significant innovations, how about for non-commutative data? One simple but hard problem is to quantify the noisy non-commutative communication channel in quantum information, which highly depends on informational understanding of noncommutative entropy, for example, quantum relative entropy in physical implementation.

2) Game Theory

Extending the simpler optimization approach developed in variation methods and mathematical programming in mathematics, optimization theory and algorithms in information and computer science, and operation research in neoclassical economics, game theory studies situations where multiple players make decisions in an attempt to maximize their returns. The essential feature is to

provide a formal modeling approach to social or informational situations in which decision makers interact with other agents.

In history, game theory formally exists as a unique field after 1928 [77], which methodized finding mutually consistent solutions for two-person zero-sum games, during which period, game work was primarily on cooperative game theory, where they analyze optimal strategies for groups of individuals, assuming that they can enforce agreements between them about proper strategies.

The prisoner's dilemma was firstly discussed in 1950, and the RAND corporation undertake an experiment. Around them, John Nash developed Nash criterion for mutual consistency of players' strategies [73], which applied to a wider variety of games than [77]. This criterion is sufficiently general, in the sense to allow for the analysis of non-cooperative games except for cooperative ones.

Game theory developed flurry in the 1950s, where occurred the concepts of the core, the extensive form game, fictitious play, repeated games, and the Shapley value, and the first applications were appeared to philosophy and political science.

In 1960s, the solution concept of subgame perfect equilibria was introduced by Reinhard Selten [90], further refining the Nash equilibrium.

In 1970s, the concepts of complete information and Bayesian games was developed by John Harsanyi [40], and game theory was extensively applied, for example, evolutionary stable strategy [70] in biology, also correlated equilibrium, trembling hand perfection, and common knowledge were introduced and analyzed.

In addition, there are other contributions, for examples, Schelling on dynamic models (early examples of evolutionary game theory) and Aumann on the equilibrium (for example, equilibrium coarsening, correlated equilibrium, and an extensive formal analysis of the assumption of common knowledge and of its consequences), and Leonid Hurwicz, Roger Myerson, and Eric Maskin on mechanism design theory.

Game theory are playing a large role in many diverse fields, for example, social sciences, behavior sciences beginning in the 1970s, political science, ethics, information and computer sciences.

Some game theoretic analysis appear similar to decision theory, but game theory studies decisions made in an environment in which players interact. Alter-

natively, game theory studies choice of optimal behavior when costs and benefits of each option depend upon the choices of other individuals. In this sense, game theory, much more than decision theory, is of similar spirit and situation to information and computer science.

Among many interrelations between game theory and information theory, we here just list the following two:

(1) In "game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory" [38], to maximize entropy and to minimize worst-case expected loss were exposed dual to each other, in a formulation of the equilibrium theory of zero-sum games between Decision Maker and Nature.

Moreover, we can associate an appropriate generalized definition of entropy with arbitrary decision problems. Subject to certain regularity conditions, the above-mentioned duality still exists, which simultaneously provides a possible rationale for maximizing entropy and a tool for finding robust Bayesian acts. Thus the identity between the problem of maximizing entropy and that of minimizing a related divergence between distributions leads to an extension, to arbitrary divergences, of a well-known minimax theorem for the case of Kullback-Leibler divergence (the "redundancy-capacity theorem" of information theory).

Does it hold quantum correspondence for this identity? Undoubtedly, it difficultly depends on informational understanding of noncommutative entropy, for example, quantum relative entropy in physical implementation.

(2) In "a new interpretation of information rate" [52], applying information theory to gambling theory, the Kelly criterion was formulated to maximize the long-term growth rate of repeated plays of a given gamble (with positive expected value).

How about application to quantum aspects? Undoubtedly, it is worth since not only information theory and game theory can offer a conceptional understanding of quantum theory, but also quantum information and game theory are vital themselves and can be applied extensively in engineering, well, it is still a long way to grasp their common intrinsics.

In particular, how about Kelly criterion in quantum world? For example, let $\{X \in \mathbb{R}^{m\times m} : X_{ij} \geq 0, i, j = 1, 2, ..., m\}$ denote a set of matrix-valued awards and $B = \{b \in \mathbb{C}^{m\times m} : b \geq 0, \mathrm{Tr}b = 1\}$ the set of all quantum states, how to apply Kelly criterion to identify quantum state? This project will investigate identifying quantum state, during which a practical game theoretic interpreta-

tion of classical information will be given.

3) Quantum Information

To find fundamental limits on compressing and reliably communicating data, information theory is to quantify information.

In lossless data compression, lossy data compression, and channel coding, classical information theory fundamentally applied, which is also crucial to the Voyager missions to deep space, the CD, as well as the mobile phones, the Internet, and numerous other fields.

Information theory crossroads between mathematics, probability theory, computer science, and game theory, economics, etc., for examples, Limit Theorems and Large Deviations in Probability Theory, Kolmogorov Complexity in Computer Science, Kelly criterion in Game Theory, Portfolio Theory and Entropy Metrics in Statistics and Economics.

Since [42], there are several fundamental advances made in quantum information, for example, in quantum statistical mechanics, von Neumann entropy refers to the extension of classical entropy concepts to the field of quantum mechanics, but it is still one open problem to give an information theoretic formulation of Von Neumann entropy, like Shannon entropy, though in the view of statistical physics, quantum mutual information was extensively investigated by [7, 16–19, 87], etc.

Also to see whether it is the proper definition of entropy for quantum information in general, we need source coding, accessible information, quantum teleportation and its converse superdense coding, quantum channel capacity. Since [50, 85], there were much research on quantum coding. Since [45, 86], various of quantum channel capacity were extensively investigated.

Well, it is still a long way to find, in general, the information frame for quantum entropy, quantum channel capacity, let alone quantum error correction and quantum cryptography.

Many information measures exist for quantum signals, say, von Neumann entropy, quantum conditional entropy, quantum relative entropy, quantum mutual information, among which quantum relative entropy is central, in the sense, from which others may be derived.

At least presently, several mathematical forms exist for quantum relative entropy, for example, Araki-Umegaki type [68, 93], Belavkin-Staszewski type [14], and Hammersley-Belavkin type [39], unlike Shannon information entropy, which

was uniquely derived mathematically and originally from theoretic information frame in [91].

Quantum relative entropy of Araki-Umegaki type is commonly used in quantum information. However unlike the classical case, this is not only possible choice for informational divergence of the states ϖ and φ, and it does not relate explicitly the informational divergence to the Radon-Nikodym (RN) density

$$\omega_\varphi = \phi^{-1/2}\omega\phi^{-1/2} \tag{1.0.1}$$

of the state ϖ with respect to φ as in the classical case.

Quantum relative entropy of Hammersley-Belavkin type includes the other two relative quantum information as special cases, but can not exhaust all possibilities except for commutative algebra, for example, the trace distance

$$\mathcal{R}_{tr}(\varrho;\varsigma) = \lambda(|\varrho - \varsigma|), \tag{1.0.2}$$

and the fidelity distance

$$\mathcal{R}_{fid}(\varrho;\varsigma) = 1 - \lambda(|\sqrt{\varrho}\sqrt{\varsigma}|). \tag{1.0.3}$$

It is still a challenge to obtain a general information-theoretic formulation for quantum relative entropy which reduces to a general mathematical formula for quantum relative entropy, including those well known mathematical forms.

Motivated by them jointly, the rest of this thesis will be organized as follows.

In chapter 2, we will overview basic mathematics for our application, say, probability inequalities, then basic conceptions for C^*-algebra, including von Neumann algebra, and quantum probability, leaving some game theoretic knowledge in suit sections.

In chapter 3, we will overview the related development of quantum relative entropy, and quantum mutual information, further quantum channel capacity, mainly in the direction of Stratonovich and Belavkin, and will obtain additivity of entangled quantum channel capacity for arbitrary channels.

Chapter 4 is an application of strategic game to quantum information, and applying convex analysis, we reach one new approach to quantum relative entropy, and further in the view of Stratonovich and Belavkin, we reach one new approach to quantum mutual entropy, and quantum channel capacity, also monotonicity property of quantum relative entropy and additivity of entangled quantum channel capacity will be obtained in our context.

In chapter 5, after introducing repeated game and Kelly criterion, we apply them to identify quantum state, one spot or sequentially, with/without side information and will give a practical game theoretic interpretation to classical relative entropy, mutual information, and asymptotical information, in the sense of game theory.

Chapter 6 is for conclusion and some possible further problems for this project.

CHAPTER 2

Preliminaries

This chapter is to overview basic mathematics for our application, say, probability inequalities, then basic conceptions for C^*-algebra, including von Neumann algebra, and quantum probability, while leaving some game theoretic knowledge in suit sections.

Through this project, we adopt the convention that, except where noted otherwise, $\mathbb{P}(A)$ denotes the probability which the event A occurs, and $\mathbb{E}$ the expectation with respect to the true probability density.

2.1 Probability Inequalities

In this section, we will introduce several necessary probability inequalities, which will be applied in Chapter five, in specific, the Markov inequality, the Chebyshev inequality, then the weak and strong law of large numbers, mainly on their statements. See, for example, [31], for reference in detail. We only prove the strong law of large numbers.

2.1.1 Markov Inequality

The Markov inequality, in probability theory, bridges probabilities to expectations, and gives useful bounds for the cumulative distribution function of a random variable, that is, an upper bound for the probability that a non-negative function of a random variable is greater than or equal to some positive constant.

We state the Markov inequality in measure theory, then in probability theory respectively.

Theorem 2.1.1. *(**Markov inequality in measure theory**, [31]) On a measure space*

(X, S, μ), f is a measurable extended real-valued function. Then, for any $t > 0$,

$$\mu(\{x \in X :\mid f(x) \mid \geq t\}) \leq \frac{1}{t} \int_X \mid f \mid d\mu. \tag{2.1.1}$$

In particular, for a measure space (X, S, μ) of measure 1, the Markov inequality can be restated in probability as follows.

Theorem 2.1.2. *(**Markov inequality in probability theory**, [31]) For any random variable X and $a > 0$. Then*

$$\mathbb{P}(\mid X \mid \geq a) \leq \frac{1}{a}\mathbb{E}(\mid X \mid). \tag{2.1.2}$$

2.1.2 Chebyshev Inequality

In any data sample or probability distribution, nearly all the values are close to the mean value, which gives a quantitative description of "nearly all" and "close to".

Theorem 2.1.3. *(**Chebyshev Inequality**, [31]) For a discrete random variable X with expected value $\mu = \mathbb{E}(X)$, and finite variance $\mathrm{Var}(X)$. Then, for any positive real number $\varepsilon > 0$,*

$$\mathbb{P}(\mid X - \mu \mid \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}. \tag{2.1.3}$$

Theorem 2.1.4. *(**Chebyshev inequality in measure theory**, [31]) For a measure space $(X, \sum, \mu)$, let f be an extended real-valued measurable function defined on X. Then for any real number $t > 0$,*

$$\mu(\{x \in X :\mid f(x) \mid \geq t\}) \leq \frac{1}{t^2} \int_X \mid f \mid^2 d\mu. \tag{2.1.4}$$

In general, for a nonnegative extended real-valued measurable function g, nondecreasing on the range of function f. Then

$$\mu(\{x \in X : f(x) \geq t\}) \leq \frac{1}{g(t)} \int_X g \circ f d\mu. \tag{2.1.5}$$

Obviously, the previous can be obtained from the generous case by defining $g(t)$ as t^2 if $t \geq 0$, and $g(t)$ equals 0 otherwise, in the meanwhile taking $\mid f \mid$ instead of f.

Theorem 2.1.5. *(**Chebyshev inequality in probability theory**, [31]) For a random variable X with expected value μ and finite variance σ^2. Then, for any real number $k > 0$,*

$$\mathbb{P}(\mid X - \mu \mid \geq k\sigma) \leq \frac{1}{k^2}. \tag{2.1.6}$$

Obviously, only the cases $k > 1$ offers useful information, since in general there holds

$$\mathbb{P}(\mid X - \mu \mid \geq k\sigma) \leq 1. \tag{2.1.7}$$

Remark 2.1.6. *This Chebyshev inequality can be proved by the Markov inequality in probability theory (Theorem 2.1.2) with the random variable $(X - \mu)^2$ and the constant $a = (k\sigma)^2$. This Chebyshev inequality can also be proved by the Chebyshev inequality (Theorem 2.1.3) for the random variable X with*

$$\mathbb{E}(X) = \mu, \tag{2.1.8}$$

$$\mathrm{Var}(X) = \sigma^2, \tag{2.1.9}$$

and the constant $\varepsilon = k\sigma$.

2.1.3 the Law of Large Numbers

The "law of large numbers" was developed [47], and there exists two different forms, namely, the "weak" law and the "strong" law, describing the convergence of the observed or measured probability to the actual probability.

Given a sample of independent and identically distributed random variables

$$X_1, X_2, ..., X_n \tag{2.1.10}$$

with the finite expected value

$$\mathbb{E}(X_1) = \mathbb{E}(X_2) = ... = \mathbb{E}(X_n) = \mu < \infty, \tag{2.1.11}$$

and the finite variance

$$\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = ... = \mathrm{Var}(X_2) = \sigma^2 < \infty, \tag{2.1.12}$$

the average of these observations

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + ... + X_n) \tag{2.1.13}$$

will eventually approach and stay close to the finite expected value μ. Thus the long-term stability of a random variable was reached.

the Weak Law of Large Numbers

In probability, the sample average of many observations will eventually reach close to the mean within any specified small nonzero margin.

Theorem 2.1.7. *(**the Weak Law of Large Numbers**, [31]) For $n \to \infty$, the sample average*

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + ... + X_n) \tag{2.1.14}$$

converges in probability towards the expected value

$$\mathbb{E}(X_1) = \mathbb{E}(X_2) = ... = \mathbb{E}(X_n) = \mu, \tag{2.1.15}$$

that is,

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + ... + X_n) \to \mu, \tag{2.1.16}$$

in probability.

Simply, for any positive number $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\mid \bar{X}_n - \mu \mid < \varepsilon) = 1. \tag{2.1.17}$$

Obviously, the convergence of random variables is weak in probability, thus this version is called the weak law of large numbers.

the Strong Law of Large Numbers

Theorem 2.1.8. *(**the Strong Law of Large Numbers**, [31]) For $n \to \infty$, the sample average*

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + ... + X_n) \tag{2.1.18}$$

converges almost surely to the finite expected value μ, that is,

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + ... + X_n) \to \mu, \tag{2.1.19}$$

almost surely.

In other word,

$$\mathbb{P}(\lim_{n\to\infty} \bar{X}_n = \mu) = 1. \tag{2.1.20}$$

The sample average converges almost surely to the expected value, which is strong convergence of random variables, thus this version is called the strong law of large numbers,

To prove this theorem, we introduce the following crucial lemma [95].

Lemma 2.1.9. *([95]) Let $\{Y_n\}$ $(n \geq 1)$ be a sequence of nonnegative random variables, each with the finite expectation, and*

$$Y_n \leq Y_{n+1} \tag{2.1.21}$$

for each $n \geq 1$ *and*

$$\alpha = \lim_{n\to\infty} \mathbb{E}(Y_n) < \infty. \tag{2.1.22}$$

If

$$Y = \lim_{n\to\infty} Y_n. \tag{2.1.23}$$

Then

$$\mathbb{P}(Y < +\infty) = 1. \tag{2.1.24}$$

Proof. (adapted from [31]) For $a > 0$, take the events $A = \{Y_n > a\}$, $n \geq 1$, and $A = \{Y > a\}$. Thus for each $n \geq 1$, $A_n \subseteq A_{n+1}$, and $A = \bigcup_{n=1}^{n} A_n$, therefore

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n). \tag{2.1.25}$$

According to the Markov inequality,

$$\mathbb{P}(A_n) \leq \frac{\mathbb{E}(Y_n)}{a}. \tag{2.1.26}$$

Then, for every $a > 0$,

$$\mathbb{P}(Y = +\infty) \leq \mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n) \leq \frac{\mathbb{E}(Y_n)}{a} = \frac{\alpha}{a}. \tag{2.1.27}$$

Let $a \to +\infty$, since $\alpha < \infty$, we obtain

$$\mathbb{P}(Y = +\infty) = 0, \tag{2.1.28}$$

which indicates $\mathbb{P}(Y < +\infty) = 1$. □

Proof of the Strong Law of Large Numbers

Proof. (adapted from [95]) Let $S_n = \sum_{i=1}^{n} X_i$, for all real number t, we take the moment-generating function $M_n(t)$ (See [96], etc, for reference) of S_n as

$$M_n(t) \equiv \mathbb{E}(\exp(tS_n)) = [1 + \mu(e^t - 1)]^n, \tag{2.1.29}$$

for $n \geq 1$. And

$$M_n(t) \leq \exp\{n\mu(e^t - 1)\}, \tag{2.1.30}$$

for all $n \geq 1$, and all real t, since $1 + x \leq e^x$ for all real x.

Let $0 < \varepsilon < \frac{1}{2}$,

$$U_n = S_n - n(\mu + \varepsilon) \tag{2.1.31}$$

and for $n \geq 1$,

$$V_n = n(\mu - \varepsilon) - S_n. \tag{2.1.32}$$

Then

$$\mathbb{E}(\exp(\varepsilon U_n)) = e^{-n\varepsilon(\mu+\varepsilon)} M_n(\varepsilon) \tag{2.1.33}$$

$$\leq \exp\{-n\varepsilon\mu - n\varepsilon^2 + n\mu(e^{\varepsilon} - 1)\} \tag{2.1.34}$$

$$= \exp\{-n\varepsilon^2 + n\mu(e^{\varepsilon} - 1 - \varepsilon)\}; \tag{2.1.35}$$

$$\mathbb{E}(\exp(\varepsilon V_n)) = e^{n\varepsilon(\mu-\varepsilon)} M_n(-\varepsilon) \tag{2.1.36}$$

$$\leq \exp\{n\varepsilon\mu - n\varepsilon^2 - n\mu(e^{-\varepsilon} - 1)\} \tag{2.1.37}$$

$$= \exp\{-n\varepsilon^2 + n\mu(e^{-\varepsilon} - 1 + \varepsilon)\}. \tag{2.1.38}$$

Since $2(1-\varepsilon) > 1$, applying the Maclaurin expansion of e^{ε} and $e^{-\varepsilon}$ respectively, we obtain

$$e^{\varepsilon} - 1 - \varepsilon = \sum_{k=2}^{n} \frac{\varepsilon^k}{k!} \leq \sum_{k=2}^{n} \frac{\varepsilon^k}{2} = \frac{\varepsilon^2}{2(1-\varepsilon)} < \varepsilon^2; \tag{2.1.39}$$

$$e^{-\varepsilon} - 1 + \varepsilon = \sum_{k=2}^{n} \frac{(-\varepsilon)^k}{k!} \leq \sum_{k=2}^{n} \frac{(-\varepsilon)^k}{2} = \frac{\varepsilon^2}{2(1+\varepsilon)} < \varepsilon^2. \tag{2.1.40}$$

Thus, we obtain

$$\mathbb{E}(\exp(\varepsilon U_n)) < \exp\{-n(1-\mu)\varepsilon^2\} = r^n; \tag{2.1.41}$$

$$\mathbb{E}(\exp(\varepsilon V_n)) < \exp\{-n(1-\mu)\varepsilon^2\} = r^n, \tag{2.1.42}$$

where $r = \exp\{-(1-\mu)\varepsilon^2\} \in (0,1)$, so the geometric series $\sum_{n=1}^{\infty} r^n$ converges. Then

$$\sum_{n=1}^{\infty} \mathbb{E}(\exp(\varepsilon U_n)) < \infty; \tag{2.1.43}$$

$$\sum_{n=1}^{\infty} \mathbb{E}(\exp(\varepsilon V_n)) < \infty, \tag{2.1.44}$$

for $0 < \varepsilon < \frac{1}{2}$.

Let $Y_n = \sum_{k=1}^{n} \mathbb{E}(\exp(\varepsilon U_n))$, and $Y_n' = \sum_{k=1}^{n} \mathbb{E}(\exp(\varepsilon V_n))$, for $n \geq 1$, above lemma indicates

$$\mathbb{P}(\sum_{n=1}^{\infty} \mathbb{E}(\exp(\varepsilon U_n)) < \infty) = 1; \tag{2.1.45}$$

$$\mathbb{P}(\sum_{n=1}^{\infty} \mathbb{E}(\exp(\varepsilon V_n)) < \infty) = 1. \tag{2.1.46}$$

Since the summands of a convergent series necessarily converge to zero, we find that

$$\mathbb{P}(\lim_{n\to\infty} U_n = -\infty) = \mathbb{P}(\lim_{n\to\infty} \exp(\varepsilon U_n) = 0) = 1; \tag{2.1.47}$$

$$\mathbb{P}(\lim_{n\to\infty} V_n = -\infty) = \mathbb{P}(\lim_{n\to\infty} \exp(\varepsilon V_n) = 0) = 1. \tag{2.1.48}$$

Hence, for all sufficiently large n,

$$\mathbb{P}(\frac{S_n}{n} - \mu < \varepsilon) = 1; \tag{2.1.49}$$

$$\mathbb{P}(\frac{S_n}{n} - \mu > -\varepsilon) = 1. \tag{2.1.50}$$

Both statements gives the strong law of large numbers. □

2.2 C^*-algebra

In the theory of unitary representations of locally compact groups, and algebraic formulations of quantum mechanics, C^*-algebras are intensively investigated in detail, for example, [26, 33, 35, 84, 88]. Critically, C^*-algebras was abstractly characterized by Israel Gelfand, Mark Naimark and Irving Segal independent on operators.

This section introduces abstract definition with some examples, including von Neumann algebra, and their properties respectively, which will be employed all around this project.

Since matrix mechanics [28], there exists many examples in C^*-algebras, a subfield in functional analysis. For example, $\mathcal{A}$, a complex algebra of linear operators on a complex Hilbert space is topologically closed in the norm topology of operators, and closed under the operation of taking adjoints of operators, thus is a C^*-algebras.

2.2.1 Abstract Definition

Definition 2.2.1. *([26]) A C^*-algebra, $\mathcal{A}$, is a Banach algebra, over the field $\mathbb{C}$ of complex numbers, with an anti-automorphic involution*

$$* : \mathcal{A} \to \mathcal{A} \tag{2.2.1}$$

which satisfies

$$(x^*)^* = x, \tag{2.2.2}$$

$$x^* y^* = (yx)^*, \tag{2.2.3}$$

$$x^* + y^* = (x + y)^*, \tag{2.2.4}$$

$$(cx)^* = \bar{c}x^*, \tag{2.2.5}$$

for $x, y \in \mathcal{A}$, $c \in \mathbb{C}$, where $\bar{c}$ is the complex conjugate of c, and whose norm satisfies

$$\|xx^*\| = \|x\|\|x^*\|. \tag{2.2.6}$$

Remark 2.2.2. *([26]) Because of the norm condition, for all $x \in \mathcal{A}$,*

$$\|x\| = \|x^*\|, \tag{2.2.7}$$

any C^*-algebra, $\mathcal{A}$, is automatically a B^*-algebra (that is, a B-algebra with an involution $*$). On the contrary, not every B^*-algebra is a C^*-algebra.

Definition 2.2.3. *([26]) We call a bounded linear map*

$$\pi : \mathcal{A} \to \mathcal{B} \tag{2.2.8}$$

between B^*-algebras $\mathcal{A}$ and $\mathcal{B}$ a $*$-homomorphism, if for any $x, y \in \mathcal{A}$,

$$\pi(x)\pi(y) = \pi(xy), \tag{2.2.9}$$

$$\pi(x^*) = \pi(x)^*. \tag{2.2.10}$$

Remark 2.2.4. *([26]) Any $*$-homomorphism π between C^*-algebras is non-expansive, that is, bounded with norm ≤ 1. Therefore, a $*$-homomorphism between C^*-algebras is isometry due to the norm condition.*

Definition 2.2.5. *([26]) We call a bijective $*$-homomorphism π a C^*-isomorphism, in which case, $\mathcal{A}$ and $\mathcal{B}$ are called isomorphic.*

2.2.2 Some Examples

This section introduces some examples of C^*-algebras, namely, finite-dimensional C^*-algebras, C^*-algebras of compact operators, and von Neumann Algebra.

Finite-dimensional C^*-algebras

Consider matrices as operators on the Euclidean space $\mathbb{C}^n$, take the operator norm $\|.\|$ on matrices, and give the involution by the conjugate transpose, thus algebra $M_n(\mathbb{C})$ of n by n matrices over $\mathbb{C}$ is definitely a C^*-algebra.

The self-adjoint condition induces following theorem of Artin-Wedderburn type.

Theorem 2.2.6. *([33]) A finite-dimensional C^*-algebra, $\mathcal{A}$, is canonically isomorphic to a finite direct sum*

$$\mathcal{A} = \oplus_{e \in \min \mathcal{A}} \mathcal{A}e, \tag{2.2.11}$$

where $\min \mathcal{A}$ *is the set of minimal nonzero self-adjoint central projections of $\mathcal{A}$.*

Every C^*-algebra $\mathcal{A}e$ is non-canonically isomorphic to the full matrix algebra $M_{\dim(e)}(\mathbb{C})$. We can define the finite family (indexed on $\min \mathcal{A}$ given by $\dim(e)e$) the dimension vector of $\mathcal{A}$, uniquely determining the isomorphism class of a finite-dimensional C^*-algebra.

C^*-algebras of compact operators

On a separable infinite-dimensional Hilbert space $\mathcal{H}$, the algebra $\mathcal{K}(\mathcal{H})$ of compact operators is norm closed, and closed under involution. Therefore, $\mathcal{K}(\mathcal{H})$ is a C^*-algebra.

Similar to Wedderburn's theorem for finite dimensional C^*-algebras, concrete C^*-algebras of compact operators have following characterization theorem.

Theorem 2.2.7. *([33]) For $\mathcal{A}$, a C^*-subalgebra of $\mathcal{K}(\mathcal{H})$, there are Hilbert spaces $\{\mathcal{H}_i\}_{i\in I}$ such that*

$$\mathcal{A} \cong \oplus_{i\in I}\mathcal{K}(\mathcal{H}_i), \tag{2.2.12}$$

where the C^*-direct sum consists of elements (T_i) of the form

$$\prod \mathcal{K}(\mathcal{H}_i) \tag{2.2.13}$$

with

$$\|T_i\| \to 0. \tag{2.2.14}$$

von Neumann Algebra

Von Neumann algebras are C^*-algebras, but it is not useful to consider von Neumann algebras only as C^*-algebras. In this section, we simply introduce three different definitions of von Neumann algebras [35] as follows.

Definition 2.2.8. *([35]) The von Neumann algebra, as a Banach space, is the dual of some other Banach space called the predual, where the predual of a von Neumann algebra is unique up to an isomorphism.*

Definition 2.2.9. *([35]) The von Neumann algebras are defined as weakly closed $*$-algebras of bounded operators (on a Hilbert space) containing the identity.*

In this way, any von Neumann algebra is a C^*-algebra, since the $*$-algebras of bounded operators (closed in the norm topology) are C^*-algebras.

Definition 2.2.10. *([35]) A von Neumann algebra is taken as a subset of the bounded operators closed under $*$ and equal to its double commutant, or equivalently the commutant of some subset closed under $*$ [76].*

Obviously, the first defines von Neumann algebras abstractly as C^*-algebras with a predual, thus some use W^*-algebra for this abstract concept. Other two definitions take a von Neumann algebra concretely as a set of operators acting

on some given Hilbert space, thus some use von Neumann algebra for those two definitions. Therefore, a von Neumann algebra is a W^*-algebra together with a Hilbert space and a suitable faithful unital action on the Hilbert space.

Remark 2.2.11. *Similar to the abstract and concrete definitions of a C^*-algebra, the abstract and concrete definitions of a von Neumann algebra may be defined either as Banach $*$-algebras such that*

$$\|aa^*\| = \|a\|\|a^*\|, \tag{2.2.15}$$

or as norm-closed $*$-algebras of operators on a Hilbert space.

2.2.3 Some Properties

Due to the Gelfand isomorphism [26], we can reduce C^*-algebras to commutative C^*-algebras, thus a number of properties [26] follows for C^*-algebras, among which will be applied as follows. See [26], for example, for proof in detail.

Theorem 2.2.12. *([26]) The elements of a C^*-algebra $\mathcal{A}$ with the form x^*x forms a closed convex cone.*

Theorem 2.2.13. *([26]) The self-adjoint elements of a C^*-algebra $\mathcal{A}$ are naturally partially ordered.*

Theorem 2.2.14. *([26]) There exists a directed family*

$$\{e_\lambda\}_{\lambda \in I} \tag{2.2.16}$$

of self-adjoint elements of C^*-algebra $\mathcal{A}$ such that

$$xe_\lambda \to x, \tag{2.2.17}$$

and if $\lambda \leq \mu$,

$$0 \leq e_\lambda \leq e_\mu. \tag{2.2.18}$$

Theorem 2.2.15. *([26]) With the natural norm, the algebraic quotient of a C^*-algebra by a closed proper two-sided ideal is a C^*-algebra.*

Theorem 2.2.16. *([26]) A closed two-sided ideal of a C^*-algebra is a C^*-algebra.*

2.3 Quantum Probability

For quantum probability, we can see [1, 62, 71] in detail, for example for reference. Here we just introduce its definition, and will return back in continuing chapter.

We can take $\mathcal{A}$ as the σ-algebra $\sigma(X)$ generated by a random variable X, containing X (the information on the values). Classically, we often summarize information by sigma-algebra

$$\mathcal{A} \in \mathcal{F} \tag{2.3.1}$$

of events in a classical probability space

$$(\Omega, \mathcal{F}, P). \tag{2.3.2}$$

Similarly, to describe the non-commutative features and the information available in an quantum experiment, we take the appropriate algebraic structure for observables as the $*$-algebra of operators on a Hilbert space to represent quantum information in analogous algebraic pictures. It is often assumed $*$-algebra of operators consisting of bounded operators and closed in the operator norm; or closed in the strong operator topology, thus becomes a von Neumann algebra. Traditional quantum mechanics uses the algebra $\mathcal{B}(\mathcal{H})$ of all bounded operators on some Hilbert space $\mathcal{H}$.

Finite dimensional C^*-algebras are isomorphic to a direct sum of copies of full matrix algebras. A unital $*$-algebra is a complex vector subspace $\mathcal{A}$ of operators on a Hilbert space $\mathcal{H}$ containing the identity I and closed under composition (a multiplication) and adjoints (an involution).

A (mixed) state ρ on C^*-algebra $\mathcal{A}$ can be abstractly taken as a linear functional

$$\rho : \mathcal{A} \to \mathbb{C}, \tag{2.3.3}$$

such that (positivity) $\rho(A^*A) \geq 0$ for all $A \in \mathcal{A}$ and (normalization) $\rho(I) = 1$. A projection is taken as an operator $P \in \mathcal{A}$, such that

$$P^2 = P = P^\dagger. \tag{2.3.4}$$

Therefore, we obtain the general definition of quantum probability for finite dimensional space.

Definition 2.3.1. *(**Quantum probability space of finite dimension**, [71]) A pair $(\mathcal{N}, \rho)$ is called a finite dimensional quantum probability space, where $\mathcal{N}$ is a $*$-algebra, ρ is a state, the projections $P \in \mathcal{N}$ are the events in $\mathcal{N}$, and $\rho(P)$ gives the probability of the event P.*

CHAPTER 3

Entangled Quantum Channel Capacity

3.1 Introduction

Unlike classical channels, quantum channels have several different capacities (e.g. for sending classical information or quantum information, one-way or two-way communication, prior or via entanglement, etc.).

Well, the problem of characterizing in general the capacity of a noisy quantum channel is unsolved although several attempts have been made to define a quantum analog of Shannon mutual information (see the conceptions of coherent information [86] or von Neumann mutual entropy [24, 25, 30].

Unfortunately most of these attempts do not give a satisfactory solution in the sense that the defined quantities fail to preserve such important property of classical informational capacity as additivity and some do not have even monotonicity property. This chapter is following the approach to quantum channel capacity in [16–19], which is free of the above difficulties due to the enlargement of the class of input encodings, including the encodings via entanglement for one-way communication.

Quantum entanglement is a uniquely quantum mechanical resource that plays a key role, along with the celebrating paper [36] of Einstein-Podolsky-Rosen, in many of the most interesting applications of quantum information and quantum computation, for example, quantum entanglement is extensively used in teleporting an unknown quantum state via dual classic and Einstein-Podolsky-Rosen channels in the subject of quantum teleportation [23], quantum cryptography was investigated based on Bell's theorem [37], quantum noiseless coding

theorem appeared [50] as a quantum analogous of Shannon's classical noiseless coding theorem.

This chapter will concentrate on application of quantum entanglement to quantum source entropy and quantum channel capacity in the subject of quantum information.

Recently tremendous effort has been made to better understand the properties of quantum entanglement as a fundamental resource of nature. Although there is as yet no complete understanding and proof of physical realizability of quantum entanglement for quantum technologies, a theoretical progress has been made in understanding this strange property of quantum mechanics, for example, mathematical aspects of quantum entanglement are extensively studied as follows.

V. P. Belavkin [20] described the dynamical procedure of quantum entanglement in terms of transpose-completely positive maps in the subject of quantum decoherence and stochastic filtering theory.

V. P. Belavkin and M. Ohya [18, 19] initiated mathematical study of quantum entanglement as truly quantum couplings in the terminology of algebraic approach.

Peter Levay [65, 66] investigated geometry of quantum entanglement for two qubits (quantum entanglement of two qubits corresponds to the twisting of the bundle).

R. Penrose [82] treated quantum entanglement via spinor representation in the subject of mathematical physics.

Peter Levay [66] investigated twistor geometry of quantum entanglement for three qubits still in mathematical physics.

This chapter will follow [16–19] to treat with quantum entanglement in algebraic approach.

Taking entanglement as "true quantum" encoding, V. P. Belavkin and M. Ohya [16–19] introduced quantum conditional entropy of the entangled compound state related to the product of marginal states which is positive and obeys all natural properties of the classical conditional entropy as the relative conditional/unconditional entropy of a compound state. They studied its relation to the mutual information as the informational divergence (relative informational entropy) of the compound state with respect to the product of its marginal states in the sense of Lindblad, Araki and Umegaki [2, 68, 97]. This quantum mutual

information leads to an entropy bound of quantum mutual information and quantum channel capacity via entanglement (entanglement-assisted quantum capacity introduced in [24, 25], which considered the mutual information of input-output state of quantum channel).

Also V. P. Belavkin and P. Staszewski [14] investigated C*-algebraic generalization of relative and conditional entropy including two types of quantum relative entropy, such as Araki-Umegaki type and Belavkin-Staszewski type, and even more general informational divergencies which meet natural axiomatic properties of relative information were studied in quantum information in [39].

Based on the combination of these two original ideas, after introducing compound quantum state and two types of quantum relative entropy, namely Araki-Umegaki type and Belavkin-Staszewski type, this chapter treats two types of quantum mutual information via entanglement and corresponding quantum channel capacities via entanglement in algebraic approach.

It proves additivity property of quantum channel capacities via entanglement, which extends the results of V. P. Belavkin [16], [17] to products of arbitrary quantum channel and to quantum relative entropy of not only Araki-Umegaki type but also Belavkin-Staszewski type.

Extending [21], the rest of this chapter is organized as follows.

Section two and three introduce related notion of quantum mechanics, such as quantum state and quantum entanglement respectively.

Section four introduces two types of quantum relative entropy via entanglement.

Section five introduces quantum channel capacity via entanglement and shows additivity of quantum channel capacity via entanglement.

3.2 Quantum States in Algebraic Approach

This section is a brief mathematical review of quantum state in quantum mechanics in a discrete algebraic approach.

Anyone can turn to [63] for general physical review, or [78] for mathematical foundations of quantum mechanics, [79, 92] for a brief review of quantum mechanics principles in quantum information and computation.

In order to keep a closer link with classical information theory, we will allow

for a possibility of having classical-quantum combined systems described in what follows by discrete non-commutative W*-algebras $\mathcal{A} = (\mathcal{A}_i)$ represented by block-diagonal matrices $A = [A(i)\delta^i_j]$ with arbitrary uniformly bounded operators $A(i) \in \mathcal{A}_i$ on some separable Hilbert spaces $\mathcal{G}_i$.

Let $\mathcal{H}$ denote the separable Hilbert space of a quantum system, and denote the algebra of all linear bounded operators on $\mathcal{H}$, with a decomposable subalgebra $\mathcal{B} \subseteq \mathcal{L}(\mathcal{H})$ of elements $B \in \mathcal{B}$ of the block-diagonal form $B = [B(j)\delta^i_j]$, where $B(j) \in \mathcal{L}(\mathcal{H}_j)$, corresponding to an orthogonal decomposition $\mathcal{H} = \bigoplus_j \mathcal{H}_j$.

Note that any such algebra is weakly closed in $\mathcal{L}(\mathcal{H})$, i.e. is a W^*-algebra having a predual space $\mathcal{B}_*$, which can be identified with the trace class subspace of $\mathcal{B}$ with respect to the pairing

$$\langle \varsigma | B \rangle = \sum_j \mathrm{Tr}_{\mathcal{H}_j}[\varsigma(j)^\dagger B(j)] = \mathrm{Tr}_{\mathcal{H}}[B\varsigma^\dagger],$$

where $\varsigma(j) \in \mathcal{B}_j$ are such operators in $\mathcal{H}_j$ that

$$\mathrm{Tr}_{\mathcal{H}} \sqrt{\varsigma^\dagger \varsigma} < \infty \tag{3.2.1}$$

and $\mathrm{Tr}_{\mathcal{H}}$ is the standard trace on $\mathcal{B}$ normalized on one dimensional projectors

$$P_\psi = \psi\psi^\dagger \tag{3.2.2}$$

for $\psi \in \mathcal{H}_j$.

We now remind the definition of quantum normal state.

Definition 3.2.1. *A bounded linear functional*

$$\sigma : \mathcal{B} \to \mathbb{C} \tag{3.2.3}$$

of the form

$$\sigma(B) = \mathrm{Tr}_{\mathcal{H}}[B\varsigma] \tag{3.2.4}$$

for a $\varsigma = \varsigma^\dagger \in \mathcal{B}_*$ is called the state on $\mathcal{B}$ if it is positive for any positive operator $B \in \mathcal{B}$ and normalized $\sigma(I) = 1$ for the identity operator I in $\mathcal{B}$. The operator ς, uniquely defined as a positive trace one operator on $\mathcal{H}$, is called density operator of the state σ.

Let $\mathcal{G}$ be another separable Hilbert space and χ be a Hilbert-Schmidt operator from $\mathcal{G}$ to $\mathcal{H}$ defining a decomposition

$$\varsigma = \chi\chi^\dagger \tag{3.2.5}$$

of the state density with the adjoint operator $\chi^\dagger$ from $\mathcal{H}$ to $\mathcal{G}$. We now equip $\mathcal{G}$ with an isometric involution

$$J = J^\dagger, J^2 = I, \tag{3.2.6}$$

the complex conjugation on $\mathcal{G}$,

$$J\Sigma_k \lambda_k \zeta_k = \Sigma_k \bar{\lambda_k} J\zeta_k, \forall \lambda_k \in \mathbb{C}, \zeta_k \in \mathcal{G}, \tag{3.2.7}$$

defining an isometric transposition

$$\widetilde{A} = JA^\dagger J = \overline{A}^\dagger \tag{3.2.8}$$

on the algebra $\mathcal{L}(\mathcal{G})$, where $\overline{A} = JAJ$.

A normal state $\rho : \mathcal{A} \to \mathbb{C}$ on the algebra $\mathcal{A} \subseteq \mathcal{L}(\mathcal{G})$ is called real (or equivalently symmetric) if its density is real, $\overline{\varrho} = \varrho$ (or equivalently symmetric, $\widetilde{\varrho} = \varrho$).

Given a state, J can be always chosen in such a way that $\varrho = \overline{\varrho}$ as it was done in [16–18], but here we fix J but not ϱ, and in general we will not assume that $\varrho = \widetilde{\varrho}$. Instead, we may assume that the transposition leaves invariant the decomposable subalgebra $\mathcal{A} \subseteq \mathcal{L}(\mathcal{G})$ such that

$$\overline{\mathcal{A}} := J\mathcal{A}J = \mathcal{A}, \tag{3.2.9}$$

however, from the notational and operational point of view, it is preferable to distinguish the algebra $\mathcal{A}$ from the transposed algebras

$$\widetilde{\mathcal{A}} = \{\widetilde{A} : A \in \mathcal{A}\} = \overline{\mathcal{A}}. \tag{3.2.10}$$

Lemma 3.2.2. *[78] Any normal state ρ on $\mathcal{A} \subseteq \mathcal{L}(\mathcal{G})$ can be expressed as*

$$\rho(A) = \mathrm{Tr}_{\mathcal{H}}[\chi \widetilde{A} \chi^\dagger] = \mathrm{Tr}_{\mathcal{H}}[A\varrho], \tag{3.2.11}$$

where the density operator $\varrho \in \mathcal{A}_*$ is uniquely defined by

$$\widetilde{\varrho} = \chi^\dagger \chi = \overline{\varrho} \tag{3.2.12}$$

iff

$$\chi^\dagger \chi \in \widetilde{\mathcal{A}}. \tag{3.2.13}$$

Thus we have an operational expression

$$\rho(A) = \left\langle \chi \overline{A} \chi^\dagger | I \right\rangle \tag{3.2.14}$$

of quantum normal state, which is called standard in the case $\mathcal{G} = \mathcal{H}$ and $\chi = \sqrt{\varsigma}$, in which case $\varrho = \overline{\varsigma}$.

Generally χ is named as the amplitude operator, or simply amplitude given by a vector $\chi = \psi \in \mathcal{H}$ with

$$\psi^\dagger \psi = \|\psi\|^2 = 1 \tag{3.2.15}$$

in the case of one dimensional $\mathcal{G} = \mathbb{C}$, corresponding to the pure state

$$\sigma(B) = \psi^\dagger B \psi, \tag{3.2.16}$$

where $\chi^\dagger$ is the functional $\psi^\dagger$ from $\mathcal{H}$ to complex field $\mathbb{C}$.

Remark 3.2.3. *The amplitude operator χ is unique up to a unitary transform in $\mathcal{H}$ as a probability amplitude satisfying the conditions $\chi^\dagger \chi \in \widetilde{\mathcal{A}}$ such that $\varrho = \overline{\chi^\dagger \chi}$ is positive decomposable trace one operator $\varrho = \oplus_i \varrho(i)$ with the components $\varrho(i) \in \mathcal{L}(\mathcal{G}_i)$ normalized as*

$$\mathrm{Tr}_{\mathcal{G}_i} \varrho(i) = k(i) \geq 0, \sum_i k(i) = 1. \tag{3.2.17}$$

Therefore we can identify the predual space $\mathcal{A}_*$ with the direct sum

$$\bigoplus \mathcal{T}(\mathcal{G}_i) \subseteq \mathcal{A} \tag{3.2.18}$$

of the Banach spaces $\mathcal{T}(\mathcal{G}_i)$ of trace class operators in $\mathcal{G}_i$.

Note that we denote the probability operators

$$P_{\mathcal{A}} = \varrho \in \mathcal{A}_*, P_{\mathcal{B}} = \varsigma \in \mathcal{B}_* \tag{3.2.19}$$

as trace densities of the states ρ, σ defined as the expectations on the algebras $\mathcal{A}, \mathcal{B}$ respectively by the variations of Greek letters ρ, σ which are also used in [19] for the transposed (contravariant) density operators

$$\widetilde{\varrho} \equiv \rho = \overline{\varrho}, \widetilde{\varsigma} \equiv \sigma = \overline{\varsigma} \tag{3.2.20}$$

with respect to the bilinear pairings

$$\rho(A) = \langle A, \rho \rangle \equiv \langle \overline{\rho} | A \rangle, \sigma(B) = \langle B, \sigma \rangle \equiv \langle \overline{\sigma} | B \rangle. \tag{3.2.21}$$

We now define an entangled state ω on the W^*-tensor product algebra $\mathcal{A} \otimes \mathcal{B}$ of bounded operators on the Hilbert product space $\mathcal{G} \otimes \mathcal{H}$ by

$$\mathrm{Tr}_{\mathcal{G}}[\widetilde{A} \chi^\dagger B \chi] = \omega(A \otimes B) = \mathrm{Tr}_{\mathcal{H}}[\chi \widetilde{A} \chi^\dagger B]. \tag{3.2.22}$$

Obviously ω can be uniquely extended by linearity to a normal state on the algebra $\mathcal{A} \otimes \mathcal{B}$ generated by all the linear combinations $C = \Sigma_k \lambda_k A_k \otimes B_k$ such that

$$\omega(C^\dagger C) = \mathrm{Tr}_{\mathcal{G}}[X^\dagger X] \geq 0, \tag{3.2.23}$$

where $X = \Sigma_k \lambda_k B_k \chi \tilde{A}_k$, and

$$\omega(I \otimes I) = \mathrm{Tr}[\chi^\dagger \chi] = 1. \tag{3.2.24}$$

Remark 3.2.4. *The state (3.2.11) is pure on $\mathcal{L}(\mathcal{G} \otimes \mathcal{H})$, since it is given by an amplitude $\psi \in \mathcal{G} \otimes \mathcal{H}$ defined as*

$$(\zeta \otimes \eta)^\dagger \psi = \eta^\dagger \chi J \zeta, \forall \zeta \in \mathcal{G}, \eta \in \mathcal{H}, \tag{3.2.25}$$

with the states ρ on $\mathcal{A}$ and σ on $\mathcal{B}$ as the marginals of ω:

$$\sigma(B) = \omega(I \otimes B) = \mathrm{Tr}_{\mathcal{H}}[B\varsigma],\ \rho(A) = \omega(A \otimes I) = \mathrm{Tr}_{\mathcal{G}}[\widetilde{A}\varrho]. \tag{3.2.26}$$

Therefore, we call the state ω defined above as a pure entanglement state for $\mathcal{A} = \mathcal{L}(\mathcal{G})$, $\mathcal{B} = \mathcal{L}(\mathcal{H})$.

More general, mixed entangled states for $\mathcal{A} = \mathcal{L}(\mathcal{G})$, $\mathcal{B} = \mathcal{L}(\mathcal{H})$ can be obtained by using a stochastic amplitude operator $\chi : \mathcal{G} \to \mathcal{F} \otimes \mathcal{H}$.

Given an amplitude operator $v : \mathcal{F} \longrightarrow \mathcal{G} \otimes \mathcal{H}$ on a Hilbert space $\mathcal{F}$ into the tensor product Hilbert space $\mathcal{G} \otimes \mathcal{H}$ such that

$$\varpi := vv^\dagger \in \mathcal{A} \otimes \mathcal{B} \tag{3.2.27}$$

and

$$\mathrm{Tr}_{\mathcal{F}}[v^\dagger v] = 1, \tag{3.2.28}$$

we define a compound state $\omega : \mathcal{A} \otimes \mathcal{B} \longrightarrow \mathbb{C}$ as

$$\omega(A \otimes B) = \mathrm{Tr}_{\mathcal{F}}[v^\dagger (A \otimes B) v] = \mathrm{Tr}[(A \otimes B)\varpi]. \tag{3.2.29}$$

Lemma 3.2.5. *([16–19]) Any compound state (3.2.29) can be achieved via an entanglement χ as*

$$\mathrm{Tr}_{\mathcal{G}}[\widetilde{A}\chi^\dagger (I \otimes B)\chi] = \omega(A \otimes B) = \mathrm{Tr}_{\mathcal{F}\otimes\mathcal{G}}[\chi \widetilde{A} \chi^\dagger (I \otimes B)], \tag{3.2.30}$$

with

$$\omega(A \otimes I) = \mathrm{Tr}_{\mathcal{G}}[A\varrho], \omega(I \otimes B) = \mathrm{Tr}_{\mathcal{H}}[B\varsigma], \widetilde{\varrho} = \chi^\dagger \chi \tag{3.2.31}$$

and

$$\varsigma = \mathrm{Tr}_{\mathcal{F}}[\chi \chi^\dagger], \tag{3.2.32}$$

where χ is an operator $\mathcal{G} \longrightarrow \mathcal{F} \otimes \mathcal{H}$ with

$$\mathrm{Tr}_{\mathcal{F}}[\chi \mathcal{A} \chi^\dagger] \subset \mathcal{B}, \chi^\dagger (I \otimes B)\chi \subset \mathcal{A}. \tag{3.2.33}$$

Moreover, the operator χ is uniquely defined by $\widetilde{\chi}U = v$, where

$$(\zeta \otimes \eta)^{\dagger}\widetilde{\chi}\xi = (J\xi \otimes \eta)^{\dagger}\chi J\zeta, \ \ \forall \xi \in \mathcal{F}, \zeta \in \mathcal{G}, \eta \in \mathcal{H}, \tag{3.2.34}$$

up to a unitary transformation U of the minimal space $\mathcal{F} = \mathrm{rank}v^{\dagger}$ equipped with an isometric involution J.

Note that we have used the invariance of trace under the transposition such that

$$\mathrm{Tr}_{\mathcal{G}}[\widetilde{\varrho}] = \mathrm{Tr}_{\mathcal{G}}[\varrho]. \tag{3.2.35}$$

Proof. In an ortho-normal basis $\{\xi_k\} \subset \mathcal{F}$, we define the involution $J : \mathcal{F} \to \mathcal{F}$ by $J = U^{\dagger}CU$, thus $J\xi_k = \xi_k$, and

$$v = \sum |n\rangle \otimes \psi_k(n)\xi_k^{\dagger} = \tilde{\chi}U, \tag{3.2.36}$$

where $U = \sum |k\rangle\xi_k^{\dagger}$ and the canonical basis

$$\{|k\rangle\} \subset \mathcal{F} \in \ell^2(\mathbf{N}). \tag{3.2.37}$$

Due to the real isometric transformation U of $\{\xi_k\} \subset \mathcal{F}$ into $\{|k\rangle\} \subset \mathcal{F}$,

$$\bar{U} \equiv CUJ = U, \tilde{U} \equiv CU^{\dagger}J = U^{\dagger}, \tag{3.2.38}$$

the amplitude operator $\chi : \mathcal{G} \longrightarrow \mathcal{F} \otimes \mathcal{H}$ (defined by the transposition of $vU^{\dagger} = v\tilde{U} \equiv \tilde{\chi}$) is equivalent to $\tilde{v} : \chi = (U \otimes I)\tilde{v}$.

Therefore

$$\chi^{\dagger}\chi = \mathrm{Tr}_{\mathcal{H}}[vv^{\dagger}] = \rho, \mathrm{Tr}_{\mathcal{F}}[\chi\chi^{\dagger}] = \mathrm{Tr}_{\mathcal{G}}[vv^{\dagger}] = \sigma. \tag{3.2.39}$$

Since $\omega = vv^{\dagger} \in \mathcal{A} \otimes \mathcal{B}$, we can know

$$\chi^{\dagger}(I \otimes B)\chi = \mathrm{Tr}_{\mathcal{H}}[(I \otimes B)\omega] \in \mathcal{A}, \mathrm{Tr}_{\mathcal{F}}[\chi\tilde{A}\chi^{\dagger}] = \mathrm{Tr}_{\mathcal{G}}[(A \otimes I)\omega] \in \mathcal{B}. \tag{3.2.40}$$

Moreover, recalling the equality

$$\mathrm{Tr}_{\mathcal{G}}[|n\rangle \langle m|\chi^{\dagger}(I \otimes \eta\eta^{\dagger})\chi] = \mathrm{Tr}_{\mathcal{F}}[v^{\dagger}(|m\rangle \langle n| \otimes \eta\eta^{\dagger})v], \tag{3.2.41}$$

the families of the vectors

$$(I \otimes \eta^{\dagger})\chi|n\rangle \ \subseteq \mathcal{F} \in \ell^2(\mathbf{N}) \tag{3.2.42}$$

and of the vectors

$$(\langle n| \otimes \eta^{\dagger})v \subseteq \mathcal{F} \tag{3.2.43}$$

are isometric, therefore the entangling operator is uniquely defined in the minimal space $\mathcal{F}$ up to the unitary operator U intertwining the involutions C and J. □

3.3 Entanglement as Quantum Operation

Quantum entanglement is iron to the classical world's bronze age. Quantum entanglement are recently researched extensively, such as Peter Levay [65, 66] via geometric method, Penrose [82] and Peter Levay [65, 66] via spinor and twistor representation, Belavkin [16–19] via algebraic approach.

We now follow [16–19] for entangled state.

Let us write the entangled state as

$$\omega(A \otimes B) = \mathrm{Tr}_{\mathcal{H}}[B\pi^*(A)] = \mathrm{Tr}_{\mathcal{G}}[A\pi(B)], \tag{3.3.1}$$

where the operator

$$\pi^*(A) = \mathrm{Tr}_{\mathcal{F}}[\chi\widetilde{A}\chi^\dagger] \in \mathcal{B}, \tag{3.3.2}$$

bounded by $\|A\|\varsigma \in \mathcal{B}_*$, is in the predual space $\mathcal{B}_* = \mathcal{T}(\mathcal{H})$ of $\mathcal{B}$ for any $A \in \mathcal{G}$, and

$$\pi(B) = J\chi^\dagger(I \otimes B^\dagger)\chi J = \tilde{\chi}(I \otimes \widetilde{B})\bar{\chi}, \tag{3.3.3}$$

with $\widetilde{B}$ defined by isometric involution in $\mathcal{H}$ as $\widetilde{B} = JB^+J$, is in $\mathcal{A}_*$ as a trace-class operator in $\mathcal{G}$, bounded by $\|B\|\varsigma \in \mathcal{A}_*$.

The dual linear maps π and π^* in (3.3.1), $\pi^{**} = \pi$, with respect to the standard pairing $\langle A|A\rangle = \mathrm{Tr}[A^*A]$, are both positive, but in general not completely positive but transpose-completely positive maps, with

$$\pi^*(I) = \varsigma, \pi(I) = \varrho. \tag{3.3.4}$$

Remark 3.3.1. *For the entangled state $\omega(A \otimes B) = \mathrm{Tr}[(A \otimes B)\varpi]$, in terms of the compound density operator $\varpi = vv^\dagger$, the entanglements π and π^* can be written as*

$$\pi(B) = \mathrm{Tr}_{\mathcal{H}}[(I \otimes \widetilde{B})\varpi], \pi^*(A) = \mathrm{Tr}_{\mathcal{G}}[(\widetilde{A} \otimes I)\varpi]. \tag{3.3.5}$$

Definition 3.3.2. *([16–19]) The transpose-completely positive map $\pi : \mathcal{B} \to \mathcal{A}_*$, (or its dual map $\pi^* : \mathcal{A} \to \mathcal{B}_*$), normalized as $\mathrm{Tr}_{\mathcal{G}}[\pi(I)] = 1$ (or, equivalently, $\mathrm{Tr}_{\mathcal{H}}[\pi^*(I)] = 1$) is called the quantum entanglement of the state $\sigma(B) = \mathrm{Tr}_{\mathcal{H}}[\pi(B)]$ to a state on $\mathcal{A}$ described by the density operator $\varrho = \pi(I)$ (or of $\rho(A) = \mathrm{Tr}_{\mathcal{G}}[\pi^*(A)]$ to $\varsigma = \pi^*(I)$).*

We call the standard entanglement $\pi = \pi_q$ for $(\mathcal{B}, \sigma)$ the entanglement to $\varrho = \widetilde{\varsigma}$ on $\mathcal{A} = \tilde{\mathcal{B}}$ by

$$\pi_q(B) = \varrho^{1/2}\widetilde{B}\varrho^{1/2}, B \in \tilde{\mathcal{B}}. \tag{3.3.6}$$

Obviously, $\pi_q^*(A) = \varsigma^{1/2}\widetilde{A}\varsigma^{1/2}$, where $\varsigma = \widetilde{\varrho}$, and $\pi_q^* = \pi_q$ iff $\mathcal{B} = \tilde{\mathcal{B}}$ and $\varsigma = \widetilde{\varsigma}$

The standard entanglement defines the standard compound state

$$\omega_q(A \otimes B) = \mathrm{Tr}_{\mathcal{H}}[B\varsigma^{1/2}\widetilde{A}\varsigma^{1/2}] = \mathrm{Tr}_{\mathcal{H}}[A\varrho^{1/2}\widetilde{B}\varrho^{1/2}]. \tag{3.3.7}$$

Theorem 3.3.3. *Every entanglement π on $\mathcal{B}$ to the state $\varrho \in \mathcal{A}_*$ has a decomposition*

$$\pi(B) = \sqrt{\varrho}\widetilde{\Pi(B)}\sqrt{\varrho} \equiv \pi_q(\Pi(B)), \tag{3.3.8}$$

where Π is a normal completely positive map $\mathcal{B} \to \tilde{\mathcal{A}}$ normalized to the identity operator at least on the minimal Hilbert subspace supporting density operator $\widetilde{\varrho}$.

Moreover, this decomposition is unique by the condition $\Pi(I) = E_{\widetilde{\varrho}}$, where $E_{\widetilde{\varrho}} \in \widetilde{\mathcal{A}}$ is the orthoprojector on this minimal Hilbert subspace $\tilde{\mathcal{G}}_\varrho \subseteq \mathcal{G}$.

Proof. Π can be found as a solution to the linear equation

$$\widetilde{\varrho}^{1/2}\Pi(B)\widetilde{\varrho}^{1/2} \equiv \widetilde{\pi(B)} \quad \forall B \in \mathcal{B} \tag{3.3.9}$$

which is unique if ϱ and therefore $\widetilde{\varrho}$ is not degenerate:

$$\Pi(B) = \widetilde{\varrho}^{-1/2}\widetilde{\pi(B)}\widetilde{\varrho}^{-1/2}. \tag{3.3.10}$$

If ϱ is degenerate, we should consider the Hilbert subspace $\mathcal{G}_{\widetilde{\varrho}} = E_{\widetilde{\varrho}}\mathcal{G}$ given by the minimal orthoprojector $E_{\widetilde{\varrho}} \in \tilde{\mathcal{A}}$ supporting the state $\widetilde{\rho}(A) = \rho(\widetilde{A})$ on the transposed algebra $\tilde{\mathcal{A}}$ such that $\widetilde{\rho}\left(E_{\widetilde{\varrho}}\right) = 1$. □

3.4 Quantum Mutual Information via Entanglement

Quantum mutual information is extensively researched in [7, 16–18, 87].

Belavkin and Ohya [18, 19] introduced quantum mutual information as the von Neumann negaentropy $\mathcal{R}(\varpi) = -\mathcal{S}(\varpi)$ of the entangled compound state related to negaentropy $\mathcal{R}(\varrho \otimes \varsigma) = -\mathcal{S}(\varrho \otimes \varsigma)$ of the product of marginal states, i.e. as the relative negaentropy $\mathcal{R}^{(\mathrm{a})}(\varpi; \varphi) = -\mathcal{S}^{(\mathrm{a})}(\varpi; \varphi)$, in the sense of Lindblad, Araki and Umegaki relative entropy [2, 68, 97] with respect to $\varphi = \varrho \otimes \varsigma$;

Cerf and Adami [30] discussed mutual quantum information entropy and its subadditivity property via entropy diagram.

Note that we prefer to use in what is following the term "information" for negaentropy, leaving the term "entropy" for the opposite quantities like relative negainformation $\mathcal{S}^{(\mathrm{a})}(\varpi; \varphi) = -\mathcal{R}^{(\mathrm{a})}(\varpi; \varphi)$, which coincides with usual von Newmann entropy $\mathcal{S}(\varpi)$ if it is taken with respect to the trace $\phi = \mathrm{Tr}$.

We now follow [7, 18, 19] to define quantum mutual information via quantum entanglement.

Definition 3.4.1. *Relative quantum information of Araki-Umegaki type to compound state ω on the algebra $\mathcal{A} \otimes \mathcal{B}$, (or information divergence of the state ω with respect to a reference state ϕ) is defined by the density operator ϖ, φ of these states ω and ϕ as*

$$\mathcal{R}^{(a)}(\varpi; \varphi) = \mathrm{Tr}[\varpi(\ln \varpi - \ln \varphi)]. \tag{3.4.1}$$

This quantity is used in most definitions of quantum relative information.

However unlike the classical case, this is not only possible choice for informational divergence of the states ω and ϕ, and it does not relate explicitly the informational divergence to the Radon-Nikodym type (RN) density $\varpi_\phi = \varphi^{-1/2}\varpi\varphi^{-1/2}$ of the state ω with respect to ϕ as in the classical case.

Another quantum relative information (of Belavkin-Staszewski type [80]) was introduced in [14] as

$$\mathcal{R}^{(b)}(\varpi; \varphi) = \mathrm{Tr}[\varpi \ln(\varphi^{-1}\varpi)], \tag{3.4.2}$$

where $\varpi \ln(\varphi^{-1}\varpi) = \ln(\varpi\varphi^{-1})\varpi$ is understood as the Hermitian operator

$$\varpi^{1/2} \ln(\varpi^{1/2}\varphi^{-1}\varpi^{1/2})\varpi^{1/2} = v \ln(v^\dagger \varphi^{-1} v)v^\dagger. \tag{3.4.3}$$

This relative information can be explicitly written in terms of the RN density ϖ_ϕ as $\mathcal{R}^{(b)}(\varpi; \varphi) = \phi(r(\varpi_\phi))$, where $r(\varpi_\phi) = \varpi_\phi \ln \varpi_\phi$.

In finite dimensions and faithful states, the Belavkin-Staszewski information divergence based on quantum relative information of Belavkin and Staszewski type gives better distinction of ϖ and φ in the sense that it is greater than relative quantum information of Araki-Umegaki type [80], and that it satisfies the following important property.

Lemma 3.4.2. *Given a normal completely positive unital map* $\mathrm{K} : \mathcal{M} \to \mathcal{M}^0$*, if*

$$\omega = \omega_0 \mathrm{K}, \phi = \phi_0 \mathrm{K}, \tag{3.4.4}$$

then for both relative informations,

$$\mathcal{R}(\varpi; \varphi) \leq \mathcal{R}(\varpi_0; \varphi_0). \tag{3.4.5}$$

Generally this is called monotonicity property of relative information, which is well known since [68], [93] for Araki-Umegaki type, while Belavkin-Staszewski type also satisfies this inequality [14]. Of course it is worth mathematically proving this inequality of Belavkin-Staszewski type in the most general case.

Definition 3.4.3. *We define the mutual quantum information $\mathcal{I}_{\mathcal{A},\mathcal{B}}(\pi) = \mathcal{I}_{\mathcal{B},\mathcal{A}}(\pi^*)$ of both types in a compound state ω achieved by a quantum entanglement $\pi : \mathcal{B} \to \mathcal{A}_*$, or by $\pi^* : \mathcal{A} \to \mathcal{B}_*$ with*

$$\rho(A) = \omega(A \otimes I) = \mathrm{Tr}_{\mathcal{G}}[A\varrho], \sigma(B) = \omega(I \otimes B) = Tr_{\mathcal{H}}[B\varsigma] \tag{3.4.6}$$

as the relative information of each type of the state ω on $\mathcal{M} = \mathcal{A} \otimes \mathcal{B}$ with the respect to the product state $\phi = \rho \otimes \sigma$:

$$\mathcal{I}^{(\mathrm{a})}_{\mathcal{A},\mathcal{B}}(\pi) = \mathrm{Tr}[\varpi(\ln \varpi - \ln(\varrho \otimes I) - \ln(I \otimes \varsigma))]. \tag{3.4.7}$$

$$\mathcal{I}^{(\mathrm{b})}_{\mathcal{A},\mathcal{B}}(\pi) = \mathrm{Tr}[\varpi \ln((\varrho \otimes \varsigma)^{-1}\varpi)]. \tag{3.4.8}$$

The definition of mutual quantum entropy for Araki-Umegaki type can be found in [16–19]. Note that $\mathcal{I}^{(\mathrm{a})}_{\mathcal{A},\mathcal{B}}(\pi) \leq \mathcal{I}^{(\mathrm{b})}_{\mathcal{A},\mathcal{B}}(\pi)$ as it follows from Ohya and Petz [80].

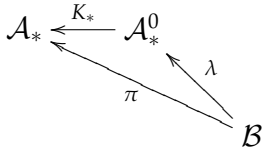
The following inequality for Araki-Umegaki type can also be found in [16–19]. Similarly this inequality for Belavkin-Staszewski type holds.

Theorem 3.4.4. *Let $\lambda : \mathcal{B} \to \mathcal{A}^0_*$ be an entanglement of the state $\sigma(B) = \mathrm{Tr}[\lambda(B)]$ to $(\mathcal{A}^0, \rho^0)$ with $\mathcal{A}^0 \subseteq \mathcal{L}(\mathcal{G}_0)$, $\varrho^0 = \lambda(I)$ on $\mathcal{B}$, and $\pi = \mathrm{K}_*\lambda$ be entanglement to the state $\rho = \rho^0\mathrm{K}$ on $\mathcal{A} \subseteq \mathcal{G}$ defined as the composition of λ with the predual operator $\mathrm{K}_* : \mathcal{A}^0_* \to \mathcal{A}_*$ normal completely positive unital map $\mathrm{K} : \mathcal{A} \to \mathcal{A}^0$.*

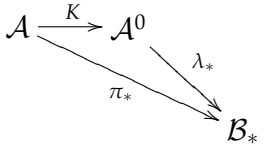
Then for both mutual quantum informations, the following monotonicity holds

$$\mathcal{I}_{\mathcal{A},\mathcal{B}}(\pi) \leq \mathcal{I}_{\mathcal{A}^0,\mathcal{B}}(\lambda). \tag{3.4.9}$$

Proof. This follows from the commutativity of the following diagrams:



Commutative diagram for entanglement π



Dual commutative diagram for entanglement π_*

Applying the monotonicity property of the relative information on $\mathcal{M} = \mathcal{A} \otimes \mathcal{B}$ with respect to the predual map $\varpi_0 \mapsto (\mathrm{K}_* \otimes \mathrm{Id})(\varpi_0)$ corresponding to $\omega_0 \mapsto \omega_0(\mathrm{K} \otimes \mathrm{Id})$ as the ampliation $\mathrm{K} \otimes \mathrm{Id}$ of a normal completely positive unital map $\mathrm{K} : \mathcal{A} \to \mathcal{A}^0$. □

Definition 3.4.5. *The maximal quantum mutual information $\mathcal{J}_{\tilde{\mathcal{B}},\mathcal{B}}(\pi_q)$ for both types as the supremum*

$$H_{\mathcal{B}}(\varsigma) = \sup_{\pi^*(I)=\varsigma} \mathcal{I}_{\mathcal{B},\mathcal{A}}(\pi^*) = \mathcal{J}_{\mathcal{B},\tilde{\mathcal{B}}}(\pi^*_q) \tag{3.4.10}$$

over all entanglements π^* of any $(\mathcal{A},\rho)$ to $(\mathcal{B},\sigma)$ is achieved on $\mathcal{A}^0 = \tilde{\mathcal{B}}$, $\varrho^0 = \tilde{\varsigma}$ by the standard quantum entanglement $\pi^*_q(A) = \varsigma^{1/2}\widetilde{A}\varsigma^{1/2}$ for a fixed $\sigma(B) = \mathrm{Tr}_{\mathcal{H}}[B\varsigma]$, which is named as entangled, or true quantum entropy of each type of the state σ.

This definition for Araki-Umegaki type can be found in [16–19].

Definition 3.4.6. *We call the positive difference*

$$H_{\mathcal{B}|\mathcal{A}}(\pi) = H_{\mathcal{B}}(\varsigma) - \mathcal{I}_{\mathcal{A},\mathcal{B}}(\pi) \tag{3.4.11}$$

entangled (or true quantum) conditional entropy respectively of each type on $\mathcal{B}$ with respect to $\mathcal{A}$.

This definition for Araki-Umegaki type can be found in [16–19].

Obviously, the conditional mutual quantum entropies of both types are positive, unlike the "conditional entropies" considered for example in [24].

3.5 Entangled Channel Capacity and its Additivity

Entanglement-assisted quantum capacity, or entangled quantum capacity is extensively researched recently, for example, entangled quantum capacity [16–19], and entanglement-assisted quantum capacity [24, 25].

Generally C. H. Bennett, P. W. Shor, J. A. Smolin and A. V. Thapliyal [24, 25] defined entanglement-assisted capacity of quantum channel via a common framework, we now discuss quantum channel capacity via entangled mutual quantum information entropy.

Let $\mathcal{B} \subseteq \mathcal{L}(\mathcal{H})$ be the W^*-algebra of operators in a (not necessarily finite dimensional unitary) Hilbert space $\mathcal{H}$. Generally we denote the set of states, i.e. positive unit trace operators in $\mathcal{B}(\mathcal{H})$ by $\mathcal{S}(\mathcal{H})$, the set of all m-dimensional projections by $\mathcal{P}_m(\mathcal{H})$ and the set of all projections by $\mathcal{P}(\mathcal{H})$.

Definition 3.5.1. *A quantum channel Λ is a normal unital completely positive linear map (UCP) of $\mathcal{B}$ into the same or another algebra $\mathcal{B}^0 \subseteq \mathcal{B}(\mathcal{H}^0)$.*

These maps admit the Kraus decomposition, which is usually written in terms of the dual map $\Lambda^* : \mathcal{B}^0_* \to \mathcal{B}_*$ as

$$\Lambda^*(\varsigma^0) = \sum_k A_k \varsigma^0 A_k^* \equiv \Lambda_*(\varsigma^0) \tag{3.5.1}$$

(W. F. Stinespring [93], G. Lindblad [69], A. S. Holevo [44]), $\Lambda(B) = \sum_k A_k^* B A_k$, for A_k are operators $\mathcal{H}^0 \to \mathcal{H}$ satisfying $\sum_k A_k^* A_k = I^0$.

For example, quantum noiseless channel in the case $\mathcal{B} = \mathcal{L}(\mathcal{H})$, $\mathcal{B}^0 = \mathcal{L}(\mathcal{H}^0)$ is described by a single isometric operator $Y : \mathcal{H}^0 \to \mathcal{H}$ as $\Lambda(B) = Y^* B Y$. See, for example, [29,30] for the simple cases $\mathcal{B} = \mathcal{L}(\mathcal{H})$, $\dim(\mathcal{H}) < \infty$.

A noisy quantum channel sends input pure states $\sigma_0 = \rho_0$ on the algebra $\mathcal{B}^0 = \mathcal{L}(\mathcal{H}^0)$ into mixed states described by the output densities $\varsigma = \Lambda^*(\varsigma^0)$ on $\mathcal{B} \subseteq \mathcal{L}(\mathcal{H})$ given by the predual $\Lambda_* = \Lambda^* \mid \mathcal{B}_*^0$ to the normal completely positive unital map $\Lambda : \mathcal{B} \to \mathcal{B}^0$ which can always be written as

$$\Lambda(B) = \mathrm{Tr}_{\mathcal{F}_+}[Y^\dagger B Y], \tag{3.5.2}$$

here Y is a linear operator from $\mathcal{H}^0 \otimes \mathcal{F}_+$ to $\mathcal{H}$ with $\mathrm{Tr}_{\mathcal{F}_+}[Y^\dagger Y] = I$, and $\mathcal{F}_+$ is a separable Hilbert space of quantum noise in the channel.

Each input mixed state σ^0 is transmitted into an output state $\sigma = \sigma^0 \Lambda$ given by the density operator

$$\Lambda^*(\varsigma^0) = Y(\varsigma^0 \otimes I_+)Y^\dagger \in \mathcal{B}_* \tag{3.5.3}$$

for each density operator $\varsigma^0 \in \mathcal{B}_*^0$, the identity operator $I_+ \in \mathcal{F}_+$.

We follow [16–19] to denote K_q the set of all normal transpose-completely positive maps $\kappa : \mathcal{A} \to \mathcal{B}^0$ with any probe algebra $\mathcal{A}$, normalized as $\mathrm{Tr}[\kappa(I)] = 1$, and $\mathcal{K}_q(\varsigma^0)$ be the subset of $\kappa \in \mathcal{K}_q$ with $\kappa(I) = \varsigma^0$.

We take the standard entanglement π_q^0 on $(\mathcal{B}^0, \sigma^0) = (\mathcal{A}_0, \rho^0)$, where $\rho_0(A_0) = \mathrm{Tr}[A_0 \varrho_0]$ given by the density operator $\varrho_0 = \varsigma^0$, and denote by K a normal unital completely positive map $\mathcal{A} \to \mathcal{A}^0 = \widetilde{\mathcal{A}}_0$ that decomposes κ as $\kappa(A) = \varrho_0^{1/2} \widetilde{\mathrm{K}(A)} \varrho_0^{1/2}$. It defines an input entanglement $\kappa^* = \mathrm{K}_* \pi_q^0$ on the input of quantum channel as transpose-completely positive map on $\mathcal{A}_0 = \mathcal{B}^0$ into $\mathcal{A}_*$ normalized to $\varrho = \mathrm{K}_* \varrho^0$, $\varrho^0 = \widetilde{\varrho}_0$.

The channel Λ transmits this input entanglement as a true-quantum encoding into the output entanglement

$$\pi = \mathrm{K}_* \pi_q^0 \Lambda \equiv \mathrm{K}_* \lambda \tag{3.5.4}$$

mapping $\mathcal{B}$ via the channel Λ into $\mathcal{A}_*$ with $\pi(I) = \varrho$. The mutual entangled information, transmitted via the channel for quantum encoding κ, is therefore

$$\mathcal{J}_{\mathcal{A},\mathcal{B}}(\kappa^* \Lambda) = \mathcal{J}_{\mathcal{A},\mathcal{B}}(\mathrm{K}_* \pi_q^0 \Lambda) = \mathcal{J}_{\mathcal{A},\mathcal{B}}(\mathrm{K}_* \lambda), \tag{3.5.5}$$

where $\lambda = \pi_q^0 \Lambda$ is the standard input entanglement $\pi_q^0(B) = \varsigma_0^{1/2} \widetilde{B} \varsigma_0^{1/2}$ with $\varsigma_0 = \widetilde{\varsigma}^0$, transmitted via the channel Λ.

Lemma 3.5.2. *Given a quantum channel $\Lambda : \mathcal{B} \to \mathcal{B}^0$, and an input state σ^0 on $\mathcal{B}^0$, the entangled input-output quantum information capacity via a channel $\Lambda : \mathcal{B} \to \mathcal{B}^0$ as the supremum over the set $\mathcal{K}_q(\varsigma^0)$ including true-quantum encodings κ achieves the maximal value*

$$\mathcal{J}(\varsigma^0, \Lambda) = \sup_{\kappa \in \mathcal{K}_q(\varsigma^0)} \mathcal{J}(\kappa^* \Lambda) = \mathcal{I}_{\mathcal{A}^0, \mathcal{B}}(\lambda), \tag{3.5.6}$$

where $\lambda = \pi_q^0 \Lambda$ is given by the corresponding extremal input entanglement π_q^0 mapping $\mathcal{B}^0 = \tilde{\mathcal{A}}^0$ into $\mathcal{A}^0 = \tilde{\mathcal{B}}^0$ with $\mathrm{Tr}[\pi_q(B)] = \sigma^0(B)$ for all $B \in \mathcal{B}^0$.

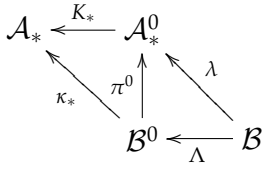
Note that this Lemma for Araki-Umegaki type can be found in [16–19].

Proof. Given to the monotonicity

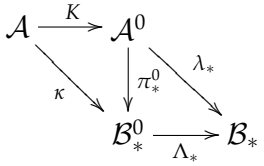
$$\mathcal{R}(\varpi_{01}(\mathrm{K} \otimes \Lambda); \varrho_0 \mathrm{K} \otimes \mathrm{Tr}) \leq \mathcal{R}(\varpi_{01}(I \otimes \Lambda); \varrho_0 \otimes \mathrm{Tr}), \tag{3.5.7}$$

the supremum of $\mathcal{J}(\kappa^* \Lambda)$ over all $\kappa \in \mathcal{K}_q(\varsigma^0)$ is achieved on the standard entanglement $\mathcal{B} \to \mathcal{A}^0$ given by $\kappa^* = \pi_q^0 \equiv \kappa^0$. □

The following definition uses commutativity of diagrams:



Commutative diagram for quantum channel Λ with standard entanglement $\pi^0 = \pi_q^0$ for $\mathcal{A} = \widetilde{\mathcal{B}}^0$



Dual commutative diagram for quantum channel Λ with standard entanglement π_*^0 for $\mathcal{A}_*^0 = \widetilde{\mathcal{B}}_*^0$

Definition 3.5.3. *Given a quantum channel $\Lambda : \mathcal{B} \to \mathcal{B}^0$, and a input state σ^0 on $\mathcal{B}^0$, we can define the input-output entangled information capacity as the maximal mutual quantum information*

$$\mathcal{J}(\varsigma^0, \Lambda) = \mathcal{I}_{\mathcal{B}^0, \mathcal{B}}(\pi_q^0 \Lambda) \tag{3.5.8}$$

for input standard entanglement of the state ς^0 to the state $\varrho^0 = \widetilde{\varsigma}^0$.

Note that this definition for Araki-Umegaki type can be found in [16–19]. Thus we have at least two types of such mutual quantum entropy, and obviously, $\mathcal{J}^{(\mathrm{a})}(\varsigma^0, \Lambda) \leq \mathcal{J}^{(\mathrm{b})}(\varsigma^0, \Lambda)$ with input product state $\rho_0^{\otimes} = \otimes_{i=1}^n \rho_0^i$ corresponding to the states $\rho_0^i = \sigma_i^0$ on $\mathcal{B}_i^0$.

Here and below for notational simplicity we implement the agreements $\mathcal{A}_0^i = \mathcal{B}_i^0$, $\rho_0^i = \sigma_i^0$, $\mathcal{A}_0^{\otimes} = \otimes_{i=1}^n \mathcal{B}_i^0$, $\rho_0^{\otimes} = \otimes_{i=1}^n \sigma_i^0$ such that $\varsigma_0^{\otimes} = \otimes_{i=1}^n \varrho_i^0$ is transposed

input state $\widetilde{\varrho}_0^{\otimes} = \otimes_{i=1}^n \widetilde{\varsigma}_i^0$ on $\mathcal{B}_0^{\otimes} = \otimes_{i=1}^n \mathcal{A}_i^0$ with $\widetilde{\mathcal{B}}_i^0 = \mathcal{A}_i^0 \equiv \mathcal{B}_0^i = \widetilde{\mathcal{A}}_0^i$, $\widetilde{\varsigma}_i^0 = \varrho_i^0 \equiv \varsigma_0^i = \widetilde{\varrho}_0^i$.

Let Λ_i be channels respectively from the algebra $\mathcal{B}_i$ on $\mathcal{H}_i$ to $\mathcal{B}_i^0$ on $\mathcal{H}_i^0$ for $i = 1, 2, ..., n$, and let $\Lambda^{\otimes} = \otimes_{i=1}^n \Lambda_i$ be their tensor product.

We now show the additivity property of this entangled input-output quantum information capacity under a given input state, using monotonicity property (as indicated in [16–19] for Araki-Umegaki type).

Theorem 3.5.4. *Let $\Lambda^{\otimes}$ be product channel from the algebra $\mathcal{B}^{\otimes} = \otimes_{i=1}^n \mathcal{B}_i$ to $\mathcal{A}_0^{\otimes} = \otimes_{i=1}^n \mathcal{A}_0^i$, and let $\rho_0^{\otimes} = \otimes_{i=1}^n \rho_0^i$ be the tensor product of input states σ_0^i on $\mathcal{B}_0^i$. Then*

$$\mathcal{J}(\varrho_0^{\otimes}, \Lambda^{\otimes}) = \sum_{i=1}^n \mathcal{J}(\varrho_0^i, \Lambda_i). \tag{3.5.9}$$

Proof. Take $\Lambda_{i*} : \mathcal{B}_{i*}^0 \to \mathcal{B}_{i*}$, and $\varrho_0^i \in \mathcal{B}_{i*}^0$, $\varsigma_i = \Lambda_{i*}(\varrho_0^i) \in \mathcal{B}_{i*}$, and $\mathrm{K}_*^{(n)}; \mathcal{A}_*^{\otimes} \to \mathcal{A}_*^{(n)}$, where $\mathcal{A}_{0*}^{\otimes} = \otimes_{i=1}^n \mathcal{B}_{i*}^0$, but $\mathcal{A}_*^{(n)}$ is predual to a general, not necessarily product algebra $\mathcal{A}^{(n)} \subseteq \mathcal{L}(\mathcal{G}^{(n)})$.

For $\pi^{(n)} = \mathrm{K}_*^{(n)} \pi_q^{0\otimes} \Lambda^{\otimes}$, we consider quantum mutual information $\mathcal{I}_{\mathcal{A}^{(n)}, \mathcal{B}^{\otimes}}(\pi^{(n)})$ as quantum relative entropy

$$\mathcal{R}((\mathrm{K}_*^{(n)} \otimes \Lambda_*^{\otimes})\widetilde{\omega}_0^{\otimes}; \mathrm{K}_*^{(n)}(\varsigma_0^{\otimes}) \otimes \Lambda_*^{\otimes}(\varrho_0^{\otimes})), \tag{3.5.10}$$

where $\widetilde{\omega}_0^{\otimes} = \otimes_{i=1}^n \widetilde{\omega}_0^i$ is the density operator of the standard compound state $\otimes_{i=1}^n \omega_0^i$ with

$$\omega_0^i(A_i \otimes B_i) = \omega_i^0(A_i \otimes B_i) = \mathrm{Tr}[B_i \sqrt{\varrho_i^0} \widetilde{A}_i \sqrt{\varrho_i^0}] \tag{3.5.11}$$

for $A_i \in \widetilde{\mathcal{B}}_i^0, B_i \in \mathcal{B}_i^0$, corresponding to $\varsigma_i^0 = \varrho_0^i$.

Applying monotonicity property (3.4.5) of quantum relative entropy to the probe system $(\mathcal{G}^{(n)}, \mathcal{A}^{(n)})$ for this given ϱ_0^i and Λ_i, we obtain

$$\mathcal{R}((\mathrm{K}_*^{(n)} \otimes \Lambda_*^{\otimes})\widetilde{\omega}_0^{\otimes}; \mathrm{K}_*^{(n)}(\varsigma_0^{\otimes}) \otimes \Lambda_*^{\otimes}(\varrho_0^{\otimes})) \tag{3.5.12}$$

$$\leq \mathcal{R}((\mathrm{Id}^{\otimes} \otimes \Lambda^{\otimes})\widetilde{\omega}_0^{\otimes}; \mathrm{Id}^{\otimes}(\varsigma_0^{\otimes}) \otimes \Lambda_*^{\otimes}(\varrho_0^{\otimes})) \tag{3.5.13}$$

$$= \sum_{i=1}^n \mathcal{R}((\mathrm{Id} \otimes \Lambda_{i*})(\widetilde{\omega}_0); \mathrm{Id}(\varsigma_0^i) \otimes \Lambda_{i*}(\varrho_0^i)), \tag{3.5.14}$$

where $\varsigma_0^i = \varrho_i^0 = \widetilde{\varrho}_0^i$, $\varrho_0^i = \varsigma_i^0 = \widetilde{\varsigma}_0^i$.

The suprema over $\mathrm{K}^{(n)}$ is achieved on $\mathrm{K}^{(n)} = \mathrm{Id}^{\otimes}$ identically mapping $\mathcal{A}^{(n)} = \otimes_{i=1}^n \mathcal{A}_0^i$ to $\mathcal{B}_{0*}^{\otimes} = \otimes_{i=1}^n \mathcal{B}_0^i$, where $\mathcal{B}_0^i = \widetilde{\mathcal{B}}_i^0$, coinciding with such $\mathcal{A}^{(n)}$ due to $\mathcal{A}_0^i = \widetilde{\mathcal{B}}_i^0$.

Thus $\mathcal{J}(\varrho_0^{\otimes}, \Lambda^{\otimes}) = \sum_{i=1}^n \mathcal{J}(\varrho_0^i, \Lambda_i)$. □

Definition 3.5.5. *Given a normal unital completely positive map $\Lambda : \mathcal{B} \to \mathcal{A}$, the suprema*

$$C_q(\Lambda) = \sup_{\kappa \in \mathcal{K}_q} \mathcal{I}_{\mathcal{A},\mathcal{B}}(\kappa^* \Lambda) = \sup_{\varsigma^0} \mathcal{J}(\varsigma^0, \Lambda) \tag{3.5.15}$$

is called the quantum channel capacity via entanglement, or q-capacity.

Note that this definition for Araki-Umegaki type can be found in [16–19], there we have two types of entangled channel capacities, and obviously,

$$C_q^{(\mathrm{a})}(\Lambda) \leq C_q^{(\mathrm{b})}(\Lambda). \tag{3.5.16}$$

Lemma 3.5.6. *Let $\Lambda(B) = Y^\dagger B Y$ be a unital completely positive map $\Lambda : \mathcal{B} \to \mathcal{B}^0$ describing a quantum deterministic channel by an isometry $Y : \mathcal{H}^0 \to \mathcal{H}$. Then*

$$\mathcal{J}(\varsigma^0, \Lambda) = H_{\mathcal{B}_0}(\varsigma^0), \tag{3.5.17}$$

$$C_q(\Lambda) = \ln \dim \mathcal{B}^0. \tag{3.5.18}$$

Note that this Lemma for Araki-Umegaki type can be found in [16–19].

Proof. Taking $\nu = (X \otimes Y)(I_- \otimes \nu_{01})$, $\nu^\dagger \nu = (I_- \otimes \nu_{01})^\dagger (R \otimes I)(I_- \otimes \nu_{01}) = \nu_1^\dagger \nu_1$, where $R = X^\dagger X$, $\nu_1 = (X \otimes I)(I_- \otimes \nu_{01})$. Then

$$\nu^\dagger (\rho \otimes I)^{-1} \nu = (I_- \otimes \nu_{01})^\dagger (X^\dagger \rho^{-1} X \otimes I)(I_- \otimes \nu_{01}) = \nu_1^\dagger (\rho \otimes I)^{-1} \nu_1, \tag{3.5.19}$$

where $\rho = X(I_- \otimes \rho_0)X^\dagger$.

Therefore,

$$\mathcal{I}(k^* \Lambda) = \mathrm{Tr}_{\mathcal{F}}[\nu_1^\dagger \nu_1 \ln \nu_1^\dagger (\rho \otimes I)^{-1} \nu_1] = \mathcal{I}(k^*). \tag{3.5.20}$$

Then

$$\mathcal{J}(\varsigma^0, \Lambda) = H_{\mathcal{B}_0}(\varsigma^0), \tag{3.5.21}$$

$$C_q(\Lambda) = \sup_{\kappa \in \mathcal{K}_q} \mathcal{I}_{\mathcal{A},\mathcal{B}}(\kappa^* \Lambda) = \sup_{\varsigma^0} \mathcal{J}(\varsigma^0, \Lambda) = \ln \dim \mathcal{B}^0. \tag{3.5.22}$$

□

Let $\Lambda^\otimes$ be product channel from the algebra $\mathcal{B}^\otimes = \otimes_{i=1}^n \mathcal{B}_i$ to $\mathcal{A}_0^\otimes = \otimes_{i=1}^n \mathcal{B}_i^0$. The additivity problem for entangled quantum channel capacity is if it is true that

$$\mathcal{C}_q(\Lambda^\otimes) = \sum_{i=1}^n \mathcal{C}_q(\Lambda_i). \tag{3.5.23}$$

We now still follow the idea of [16–19] (for Araki-Umegaki type) to give a proof of this additivity property upon monotonicity property.

Theorem 3.5.7. *Let $\Lambda^{\otimes}$ be product channel from the algebra $\mathcal{B}^{\otimes} = \otimes_{i=1}^{n}\mathcal{B}_i$ to $\mathcal{A}_0^{\otimes} = \otimes_{i=1}^{n}\mathcal{B}_i^0$. Then*

$$\mathcal{C}_q(\Lambda^{\otimes}) = \sum_{i=1}^{n}\mathcal{C}_q(\Lambda_i). \tag{3.5.24}$$

Proof. It simply follows from the additivity (3.5.9). Indeed,

$$\mathcal{C}_q(\Lambda^{\otimes}) = \sup_{\kappa\in\mathcal{K}_q^{(n)}}\mathcal{I}_{\mathcal{A}^{(n)},\mathcal{B}}(\kappa^*\Lambda^{\otimes}) = \sup_{\varrho_0^{\otimes}}\mathcal{J}(\varrho_0^{\otimes},\Lambda^{\otimes}) = \sup_{\varrho_0^{\otimes}}\sum_{i=1}^{n}\mathcal{J}(\varrho_0^{i},\Lambda_i) \tag{3.5.25}$$

Therefore by further taking suprema over $\varrho_0^{\otimes}$ as over independently for each $i = 1, 2, ..., n$, thus we have

$$\mathcal{C}_q(\Lambda^{\otimes}) = \sum_{i=1}^{n}\sup_{\varrho_0^{\otimes}}\mathcal{J}(\varrho_0^{i},\Lambda_i) = \sum_{i=1}^{n}\mathcal{C}_q(\Lambda_i), \tag{3.5.26}$$

which is the additivity property of entangled quantum channel capacity due to encodings via entanglement obviously. □

Remark 3.5.8. *Note that there is no such additivity for the Holevo capacity for a arbitrary channel $\Lambda : \mathcal{B} \to \mathcal{B}^0$. Indeed, this smaller semiclassical capacity is defined as the supremum*

$$\mathcal{C}_d(\Lambda) = \sup_{\kappa\in\mathcal{K}_d}\mathcal{I}_{\mathcal{A},\mathcal{B}}(\kappa^*\Lambda) \tag{3.5.27}$$

over the smaller class $\mathcal{K}_d \subseteq \mathcal{K}_q$ of the diagonal [16–19] (semiclassical) encodings $\kappa : \mathcal{A} \to \mathcal{B}_*^0$ corresponding to only diagonal (Abelian) algebras $\mathcal{A}$.

This supremum cannot in general be achieved on the standard entanglement of $\mathcal{A}^0 = \widetilde{\mathcal{B}}^0 \equiv \mathcal{B}_0$ if $\mathcal{A}^0$ is non Abelian corresponding to the non Abelian input algebra $\mathcal{B}^0$.

Therefore the supremum $\mathcal{C}_d(\Lambda^{\otimes}) \leq \sum_{i=1}^{n}\mathcal{C}_d(\Lambda_i)$ can be achieved not on a product Abelian algebra $\mathcal{A}^{(n)}$ as is was in the true quantum case where we could take $\mathcal{A}^{(n)} = \otimes_{i=1}^{n}\mathcal{B}_0^i$ with non Abelian $\mathcal{B}_0^i = \widetilde{\mathcal{B}}_i^0$.

CHAPTER 4

Quantum Relative Entropy

4.1 Introduction

We are given observations distributed according to an unknown distribution P_θ (associated with award Q), which Nature chooses at random from the set $\{P_\theta : \theta \in \Theta\}$ according to a known prior distribution μ on Θ, we produce an estimate M for the unknown distribution P_θ. In the end, we will suffer a cost, measuring the quality of this estimate, therefore the whole utility is in terms of award and cost.

One such cost function is relative entropy function $\mathcal{R}(P; M)$, important in several fields, such as information theory, data compression, computational learning theory, game theory, statistics, statistical mechanics, and econometrics.

In the source coding interpretation of the estimate, the minimax value of this game can be interpreted as the capacity of the channel from Θ to X.

In computational learning theory, the minimax value of this game is the utility shared by an adaptive algorithm, predicting each observation before it arrives on the previous observation, compared to an algorithm predicting after knowing the real distribution.

In gambling theory and finance, the relative entropy measures the expected reduction in the logarithm of compounded wealth due to lack of knowledge of the true distribution, thus the minimax value of this game is the practical compounded wealth.

In this chapter, strategic game will be briefly introduced, during which a sufficient condition for minimax theorem is obtained. An estimate is explored in the frame of game theory, and in the view of convex conjugate, we reach one new

approach to quantum relative entropy, quantum mutual entropy, and quantum channel capacity, which are more general, in the sense, without Radon-Nikodym (RN) derivatives, therefore extending classical (econo)metrics to non-commutative data, or suitably quantum data in physical implementation. Also the monotonicity of quantum relative entropy and the additivity of quantum channel capacity will be obtained.

Extending [34], the structure of the chapter is organized as follows.

In the second section, we will give a brief introduction to strategic game, during which a sufficient condition for minimax theorem is obtained.

In the third section, we will introduce convex conjugate along with some examples, mainly on its important mathematical properties for our application.

In the fourth section, we introduce an estimate in the frame of game theory with the cost of classical relative entropy and reach one new approach to classical relative entropy in the view of convex conjugate.

Similarly, we introduce this approach to quantum relative entropy, in the section five, especially monotonicity of quantum relative entropy is obtained, and further one approach to quantum mutual entropy will be given in the sixth section.

The section seven is for quantum channel capacity and its additivity.

4.2 Strategic Game

In strategic games, agents choose strategies to maximize their return, given the strategies the other agents choose. Essentially it provides a formal modeling approach to social situations, for example, in which decision makers interact with other agents, extending the simpler optimization approach developed.

Since Von Neumann and and O. Morgenstern's classic Theory of Games and Economic Behavior [77] in 1944, there are many introductions to game theory, such as [81].

In this section, we briefly introduce strategic game, mainly general definitions, existence theorems, and competitive game.

4.2.1 General Definitions

In game theoretic models, the basic entity is a player. A player may be interpreted as an individual or as a group of individuals making a decision. Once we define the set of players, we may distinguish between two types of models: those in which the sets of possible strategies of individual players are primitive; those in which the sets of possible joint strategies of groups of players are primitive. Models of the first type are referred to as "noncooperative".

A strategic game is a model of interactive decision-making in which each decision-maker chooses his plan of strategy once for all, and that these choices are made simultaneously.

The model consists of a finite set N of players and, for each player i, a set of A_i of strategies and a utility function on the set of strategy profiles $A_1 \times \ldots \times A_N$.

Definition 4.2.1. *([27], [75]) A non-cooperative finite strategic game consists of*

- *a finite set N (the set of players)*

  and for each player $i \in N$,

- *a set A_i (the set of strategies available to player i on strategy profile)*

- *a payoff function $u_i : A \to R$, where $A \equiv \times_{j=1}^{N} A_j$ (the payoff of player i).*

Denote by Σ_i the set of probability measures over A_i, which are player i's mixed strategies. And denote by the suffix $-i$ all players except i.

In the play of a strategic game, each player holds the correct expectation about the other players' behavior and acts rationally, thus a steady state is reached. If not attempting to examine the process by which a steady state is reached, we call it Nash equilibrium.

Definition 4.2.2. *([73]) A mixed Nash equilibrium of a finite strategic game $\langle N, (A_i), (u_i)\rangle$ is a vector $(\pi_1, \pi_2, \ldots, \pi_n)$, with $\pi_i \in \Sigma_i$ for all $i \in N$, such that*

$$\Sigma_{a\in A}\pi_i(a_i)\pi_{-i}(a_{-i})u_i(a_i, a_{-i}) \geq \Sigma_{a\in A}\rho_i(a_i)\pi_{-i}(a_{-i})u_i(a_i, a_{-i}) \tag{4.2.1}$$

for all $\rho_i \in \Sigma_i$ and for all $i \in N$.

Therefore, pure-strategy Nash equilibria are those which only involve degenerate mixed strategies.

The following restatement of the definition is useful elsewhere.

Definition 4.2.3. *For any $a_{-i} \in A_{-i}$, we define $B_i(a_{-i})$ best actions to be the set of player i's given a_{-i}:*

$$B_i(a_{-i}) = \{a_i \in A_i : u_i(a_i, a_{-i}) \geq u_i(a_i', a_{-i})\}. \tag{4.2.2}$$

The set-valued function B_i is called the best-response function of player i.

Therefore, a Nash equlibrium is a profile a^* of actions for which

$$a_i \in B_i(a^*_{-i}), \tag{4.2.3}$$

for all $i \in N$.

This alternative definition formulation points us to a (not necessarily efficient) method of finding Nash equilibria: At first to calculate the best response function of each player, then to find a profile a^* of actions for which $a^* \in B_i(a^*_{-i})$ for all $i \in N$.

Obviously, if the function B_i are singleton-valued, the second step deduces to solve $|N|$ equations in the $|N|$ unknowns $(a^*_i)_{i \in N}$.

4.2.2 Existence Theorems

An existence result has two purposes as follows.

At first, if we have a game satisfying the hypothesis of the result, it is hopeful to find an equilibrium.

Secondly, the existence of an equilibrium ensures the game consistent with a steady state solution.

Furthermore, the existence of an equilibria for a family of games allows us to study properties of these equilibria without finding them explicitly and without the risk to study the empty set.

It is extensively investigated under which conditions the set of Nash equilibria of a game is nonempty. We here just introduce one of the simplest existence result, whose mathematical level is much more advanced.

To prove that a Nash equilibrium exists for a game, it suffices to show that there is a profile a^* of actions such that $a^* \in B_i(a^*_{-i})$ for all $i \in N$, which is $a^* \in B(a^*)$, if we define the set-valued function $B : A \longrightarrow A$ by $B(a) = \times_{i \in N} B_i(a_{-i})$. Luckily fixed point theorems give conditions on B under which there exists a value of a^* for which $a^* \in B(a^*)$. Generally we apply the following fixed point theorem.

Theorem 4.2.4. *([51]) Let X be a compact convex subset of $\mathbb{R}^n$ and let $f : X \longrightarrow X$ be a set-valued function for which*

- *for all $x \in X$, the set $f(x)$ is nonempty and convex; and*
- *the graph of f is closed (i.e. for all sequences $\{x_n\}$ and $\{y_n\}$ such that $\{y_n\} \in f(\{x_n\})$ for all n, $x_n \longrightarrow x$ and $y_n \longrightarrow y$, we have $y \in f(\{x\})$).*

Then there exists a $x^* \in f(x^*)$.

Theorem 4.2.5. *([73, 74]) A Nash equilibrium of strategic game $\langle N, (A_i), (u_i)\rangle$ exists if for all $i \in N$,*

- *the set A_i of actions of player i is a nonempty compact convex subset of a Euclidian space; and*
- *the utility function u_i is continuous and quasi-concave on A_i.*

Proof. Let set-valued function B_i the best-response function of player i, we define $B : A \longrightarrow A$ by $B(a) = \times_{i \in N} B_i(a_{-i})$.

For every $i \in N$, the set $B_i(a_{-i})$ is nonempty since the utility function u_i is continuous and the set A_i is compact, and also is convex since the utility function u_i is quasi-concave on A_i; B has a closed graph since each utility function u_i is continuous.

Following the Kakutani's fixed point theorem, B has a fixed point; any fixed point is a Nash equilibrium of the game as noted. □

This result states that a strategic game satisfying certain conditions has at least one Nash equilibrium; but a game may have more than one equilibrium.

Note that this theorem does not apply to any game in which some player has finitely many actions, since the set of actions of every player is not convex, but a mixed strategy Nash equilibrium of every finite strategic game $\langle N, (A_i), (u_i)\rangle$ always exists.

Theorem 4.2.6. *([73]) A mixed strategy Nash equilibrium of every finite strategic game $\langle N, (A_i), (u_i)\rangle$ always exists.*

Proof. Let $G = \langle N, (A_i), (u_i)\rangle$ be a strategic game, and for each player i, let m_i be the number of members of the set A_i, then we identify the set $\Delta(A_i)$ of player i's mixed strategies with the set of vectors $(p_1, p_2, ..., p_{m_i})$ for which $p_k \geq 0$ for

all k and $\sum_{k=1}^{m_i} p_k = 1$ (p_k being the probability with which player i uses his ith pure strategy). This set is nonempty, convex, and compact.

Since expected payoff is linear in the probabilities, each player's payoff function in the mixed extension of G is quasi-concave in his strategy and continuous.

Thus a mixed strategy Nash equilibrium exists due to above theorem. □

As further application, we need the definition of the Kuhn-Tucker conditions as follows, which is, in fact, one special case of Lagrange Multiplier Method in theoretic mechanics.

Definition 4.2.7. *([55]) The Kuhn-Tucker conditions for the problem*

$$\max_x f(x) \tag{4.2.4}$$

subject to

$$g_j(x) \leq c_j \tag{4.2.5}$$

for $j = 1, 2, ..., m$, are

$$L_j'(x) = 0 \tag{4.2.6}$$

for $i = 1, 2, ..., n$, $\lambda_j \geq 0, g_j(x) \leq c_j$ and $\lambda_j[g_j(x) - c_j] = 0$ for $j = 1, 2, ..., m$, where

$$L(x) = f(x) - \sum_{j=1}^{m} \lambda_j [g_j(x) - c_j]. \tag{4.2.7}$$

4.2.3 Competitive Game

Little obtained on the set of Nash equilibria of an arbitrary strategic game, we discuss strictly competitive games and its qualitative character of the equilibria.

Definition 4.2.8. *A strategic game $\langle \{1, 2\}, (A_i), (u_i) \rangle$ is strictly competitive if for any $a \in A$ and $b \in A$, we have $u_1(a) \geq u_1(b)$ if and only if $u_2(a) \leq u_1(b)$.*

Player i maxminimizes if he chooses an action best for him on the assumption that whatever he does, player j will choose her action to hurt him as much as possible.

We will find that for a strictly completive game possessing a Nash equilibrium, a pair of actions is a Nash equilibrium if and only of the action of each player is a maxminimizer, a striking result since providing a link between individual decision-making and the reasoning behind the notion of Nash equilibrium, during which we also find that for a strictly completive game possessing Nash equilibria yield the same payoffs.

Definition 4.2.9. *Let $G = \langle\{1,2\},(A_i),(u_i)\rangle$ be a strictly completive game, the action $x^* \in A_1$ is a maxminimizer for player 1 if for all $x \in A_1$,*

$$\min_{y\in A_2} u_1(x^*,y) \geq \min_{y\in A_2} u_1(x,y). \tag{4.2.8}$$

Similarly, the action $y^* \in A_2$ is a maxminimizer for player 2 if for all $y \in A_2$,

$$\min_{x\in A_1} u_2(x,y^*) \geq \min_{x\in A_1} u_2(x,y). \tag{4.2.9}$$

Theorem 4.2.10. *([75]) Let $G = \langle\{1,2\},(A_i),(u_i)\rangle$ be a strictly completive game.*

a) If (x^*,y^*) is a Nash equilibrium of G, then x^* is a maxminimizer for player 1 and y^* is a maxminimizer for player 2.

b) If (x^*,y^*) is a Nash equilibrium of G, then

$$\max_x \min_y u_1(x,y) = \min_y \max_x u_1(x,y) = u_1(x^*,y^*), \tag{4.2.10}$$

and all Nash equilibria of G yield the same payoffs.

c) If $\max_x \min_y u_1(x,y) = \min_y \max_x u_1(x,y)$, x^* is a maxminimizer for player 1, and y^* is a maxminimizer for player 2, then (x^*,y^*) is a Nash equilibrium of G.

Proof. First to prove (a) and (b). Let (x^*,y^*) is a Nash equilibrium of G, then $u_2(x^*,y^*) \geq u_2(x^*,y)$ for all $y \in A_2$, or $u_1(x^*,y^*) \leq u_1(x^*,y)$ for all $y \in A_2$.

Hence

$$u_1(x^*,y^*) = \min_y u_1(x^*,y) \leq \max_x \min_y u_1(x,y). \tag{4.2.11}$$

Similarly,

$$u_1(x^*,y^*) \geq \max_x \min_y u_1(x,y). \tag{4.2.12}$$

Thus $u_1(x^*,y^*) = \max_x \min_y u_1(x,y)$ and x^* is a maxminimizer for player 1.

Similar argument for player 2, y^* is a maxminimizer for player 2 and $u_2(x^*,y^*) = \max_y \min_x u_2(x,y)$.

Now to prove (c). Let $v^* = \max_x \min_y u_1(x,y) = \min_y \max_x u_1(x,y)$, for a strictly completive game, we have $-v^* = \max_y \min_x u_2(x,y)$. Since x^* is a maxminimizer for player 1, we have $u_1(x^*,y) \geq v^*$ for all $y \in A_2$; similarly, $u_2(x,y^*) \geq -v^*$ for all $x \in A_1$.

Taking $y = y^*$ and $x = x^*$ in those two inequalities, we have $v^* = u_1(x^*,y^*)$, again considering this strictly completive game, we reach that (x^*,y^*) is a Nash equilibrium of G.

Following part (c), a Nash equilibrium can be found by solving the problem

$$\max_x \min_y u_1(x, y); \tag{4.2.13}$$

following part (a) and (c), Nash equilibria of strictly completive game are interchangeable: if (x, y) and (x', y') are equilibria, so are (x, y') and (x', y);

Following (b), for any strictly competitive game with a Nash equilibrium,

$$\max_x \min_y u_1(x, y) = \min_y \max_x u_1(x, y) = u_1(x^*, y^*). \tag{4.2.14}$$

If $\max_x \min_y u_1(x, y) = \min_y \max_x u_1(x, y) = u_1(x^*, y^*)$, we say this equilibrium payoff of player 1 is the value of the game. □

Theorem 4.2.11. *Let A_1, A_2 be non-empty, convex and compact subsets of $\mathbb{R}^n$ for some n.*

Let payoff $u : A_1 \times A_2 \longrightarrow \mathbb{R}$ be a continuous function, such that

- *$\forall a_2 \in A_2$, the set $\{a_1 \in A_1 : u(a_1, a_2) \geq u(a_1', a_2), \forall a_1' \in A_1\}$ is convex; and*
- *$\forall a_1 \in A_1$, the set $\{a_2 \in A_2 : u(a_1, a_2) \leq u(a_1, a_2'), \forall a_2' \in A_2\}$ is convex.*

Then, there exists an $a^* \in A_1 \times A_2$, such that

$$\max_{a_1 \in A_1} \min_{a_2 \in A_2} u(a_1, a_2) = u(a^*) = \min_{a_2 \in A_2} \max_{a_1 \in A_1} u(a_1, a_2). \tag{4.2.15}$$

Proof. At first, continuous payoff function u is quasi-concave with respect to two arguments, since $\forall a_2 \in A_2$, the set $\{a_1 \in A_1 : u(a_1, a_2) \geq u(a_1', a_2), \forall a_1' \in A_1\}$ is convex, and $\forall a_1 \in A_1$, the set $\{a_2 \in A_2 : u(a_1, a_2) \leq u(a_1, a_2'), \forall a_2' \in A_2\}$ is convex.

Following [73, 74], a Nash equilibrium $a^* \in A_1 \times A_2$ of this strategic game exists.

Further according to [75], for this competitive game,

$$\max_x \min_y u(x, y) = \min_y \max_x u(x, y) = u(a^*). \tag{4.2.16}$$

□

4.3 Convex Conjugate

In mathematics, convex conjugation, as a generalization of the Legendre transformation (in which sense, is also taken as Legendre-Fenchel transformation

or Fenchel transformation elsewhere), addressed much attention in the study of extremum problems, among which are system inequalities, the minimum or maximum of a convex function over a convex set, Lagrange multipliers, and minimax theorems.

There are excellent books on the introduction to convex analysis, such as [83] for pure mathematics, [3] for application in theoretic mechanics.

In this section, we simply overview convex conjugation, first on its definition including some examples, then on some properties for our application.

4.3.1 General Definition

Definition 4.3.1. *Let X be a linear normed space, and X^* the dual space to X, we denote the dual pairing by*

$$\langle ., . \rangle : X^* \times X \longrightarrow \mathbb{R}. \tag{4.3.1}$$

Given a function $f : X \longrightarrow \mathbb{R} \cup \{+\infty\}$ taking values on the extended real number line, we define the convex conjugate $f^* : X^* \longrightarrow \mathbb{R} \cup \{+\infty\}$ by

$$f^*(x^*) \equiv \sup\{\langle x^*, x\rangle - f(x) | x \in X\}, \tag{4.3.2}$$

or, equivalently, by

$$f^*(x^*) \equiv -\inf\{f(x) - \langle x^*, x\rangle | x \in X\}. \tag{4.3.3}$$

We consider convex conjugates for some examples, via simple computations, following the above definition.

Example 4.3.2. *An affine function is generally written by*

$$f(x) \equiv \langle a, x\rangle - b, a \in \mathbb{R}^n, b \in \mathbb{R}. \tag{4.3.4}$$

Then its convex conjugate $f^*(x^*)$, denoted by $O_a(x^*)$, is

$$f^*(x^*) = O_a(x^*) = \begin{cases} a & x^* = a; \\ \infty & x^* \neq a. \end{cases}$$

Example 4.3.3. *The norm function is generally written by*

$$f(x) \equiv \|x\|. \tag{4.3.5}$$

Then its convex conjugate $f^*(x^*)$, denoted by $O_1(x^*)$, is

$$f^*(x^*) \equiv O_1(x^*) = \begin{cases} a & \|x^*\| \leq 1; \\ \infty & \|x^*\| = 1. \end{cases}$$

Example 4.3.4. *Let $K \subseteq X$ be a convex subset and $e(x)$ be the calibration function*

$$e(x) \equiv \sup\{\langle x^*, x\rangle | x^* \in K\}. \tag{4.3.6}$$

Then its convex conjugate $e^*(x^*)$ is $O_K(x^*)$, where $O_K(x^*)$ is defined as follows.

$$e^*(x^*) \equiv O_K(x^*) = \begin{cases} 0 & x^* \in K; \\ \infty & x^* \in K^c, \end{cases}$$

where K^c is the complement of K.

Example 4.3.5. *The convex conjugate $f^*(x^*)$ of exponential function $f(x) = e^x$ is*

$$f^*(x^*) = \begin{cases} x^* \ln x^* - x^* & x^* > 0; \\ 0 & x^* = 0; \\ \infty & x^* < 0. \end{cases}$$

Let a cone X_+ be $\{x \geq 0\} \subseteq X$, and $X_+^* \equiv \{x^* \in X^* : \langle x^*, x\rangle \geq 0\}$ its dual cone. Then the convex conjugate $f^*(x^*)$ of exponential function $f(x) = e^x$ on X_+ is

$$f^*(x^*) = \begin{cases} x^* \ln x^* - x^* & x^* > 0; \\ 0 & x^* = 0. \end{cases}$$

4.3.2 Some Properties

Theorem 4.3.6. *The conjugate function of a closed convex function is a closed convex function.*

Proof. For every $t \in \mathbb{R} \bigcap [0, 1]$, and every $x^*, y^* \in X^*$, according to the definition of convex conjugate,

$$f^*(tx^* + (1-t)y^*) \equiv \sup\{\langle tx^* + (1-t)y^*, x\rangle - f(x) | x \in X\} \tag{4.3.7}$$

$$= \sup\{(\langle tx^*, x\rangle - tf(x)) + (\langle (1-t)y^*, x\rangle - (1-t)f(x)) | x \in X\} \tag{4.3.8}$$

$$\leq t \sup\{\langle x^*, x\rangle - f(x) | x \in X\} + (1-t) \sup\{\langle y^*, x\rangle - f(x) | x \in X\} \tag{4.3.9}$$

$$\equiv tf^*(x^*) + (1-t)f^*(y^*). \tag{4.3.10}$$

□

Theorem 4.3.7. ***(Order-reversing)*** *Convex-conjugation is order-reversing, i.e., if $f \leq g$, then*

$$f^* \geq g^*, \tag{4.3.11}$$

where $f \leq g$ means for every $x \in X$, $f(x) \leq g(x)$.

Proof. Since $f \leq g$, then for every $x \in X$,

$$f(x) \leq g(x). \tag{4.3.12}$$

According to the definition of convex conjugate, for every $x^* \in X^*$,

$$f^*(x^*) \equiv \sup\{\langle x^*, x\rangle - f(x) | x \in X\} \tag{4.3.13}$$

$$\geq \sup\{\langle x^*, x\rangle - g(x) | x \in X\} \tag{4.3.14}$$

$$= g^*(x^*). \tag{4.3.15}$$

Thus $f^* \geq g^*$, since every $x^* \in X^*$. □

Theorem 4.3.8. *(**Biconjugate**) The convex conjugate of a function is lower semi-continuous.*

The biconjugate f^{**} (the convex conjugate of the convex conjugate) is the closed convex hull, that is, the largest lower semi-continuous convex function smaller than f.

Furthermore, for proper functions f, $f = f^{**}$ if and only if f is convex and lower semi-continuous.

Proof. For every $x^* \leq x_0^* \in X^*$,

$$f^*(x^*) \equiv \sup\{\langle x^*, x\rangle - f(x) | x \in X\} \tag{4.3.16}$$

$$\leq \sup\{\langle x_0^*, x\rangle - f(x) | x \in X\} \tag{4.3.17}$$

$$\equiv f^*(x_0^*), \tag{4.3.18}$$

which implies that the convex conjugate of a function is lower semi-continuous. □

Theorem 4.3.9. *(**Fenchel's inequality or Fenchel-Young inequality**) For any proper convex function f and its convex conjugate f^*, Fenchel's inequality holds:*

$$\langle p, x\rangle \leq f(x) + f^*(p), \tag{4.3.19}$$

for every $x \in X, p \in X^*$.

Proof. According to the definition of convex conjugate, for every $x \in X, p \in X^*$,

$$f(x) + f^*(p) = f(x) + \sup\{\langle p, x\rangle - f(x) | x \in X\} \tag{4.3.20}$$

$$\geq f(x) + (\langle p, x\rangle - f(x)) \tag{4.3.21}$$

$$= \langle p, x\rangle. \tag{4.3.22}$$

□

Theorem 4.3.10. ***(Infimal convolution)*** *Let $f_1, ..., f_m$ be proper convex functions on X. Then*

$$(f_1 \star_{\inf} ... \star_{\inf} f_m)^* = f_1^* + ... + f_m^*, \tag{4.3.23}$$

where the infimal convolution of two functions f and g on X is defined as

$$(f \star_{\inf} g)(x) \equiv \inf\{f(x-y) + g(y) | y \in X\}. \tag{4.3.24}$$

Proof. Here we just consider the case for $m = 2$, for $x^* \in X^*$,

$$(f \star_{\inf} g)^*(x^*) = (\inf\{f(x-y) + g(y) | y \in X\})^*(x^*) \tag{4.3.25}$$

$$= \sup\{\langle x^*, x\rangle - \inf\{f(x-y) + g(y) | y \in X\} | x \in X\} \tag{4.3.26}$$

$$= \sup\{\langle x^*, (x-y) + y\rangle - f(x-y) - g(y) | x, y \in X\} \tag{4.3.27}$$

$$= \sup\{\langle x^*, x-y\rangle - f(x-y) | x, y \in X\} + \sup\{\langle x^*, y\rangle - g(y) | y \in X\} \tag{4.3.28}$$

$$= f^*(x^*) + g^*(x^*); \tag{4.3.29}$$

Since the infimal convolution is associative, i.e.,

$$[(f_1 \star_{\inf} ... \star_{\inf} f_{m-1}) \star_{\inf} f_m]^* = (f_1 \star_{\inf} ... \star_{\inf} f_{m-1})^* + f_m^*, \tag{4.3.30}$$

the theorem follows from mathematical induction for general m. □

4.4 Classical information

In classical information theory, we need find fundamental limits on compressing and reliably communicating classical data. A key measure of information is well known as information entropy, usually expressed by the average number of bits needed for storage or communication.

This section introduces basic classical information quantities, say, Shannon entropy, relative entropy, see, for example, [91], for reference.

We first overview basic mathematical forms of Shannon entropy and relative entropy, then explore classical estimation in the frame of game theory, and reach a new approach to relative entropy in the view of convex conjugate.

4.4.1 Classical Relative Entropy

Suppose there is a random variable with true distribution F (for the density f). Then we could represent that random variable with a code of average length $H(F)$, where Shannon entropy $H(F)$ is expressed in mathematics as follows.

Definition 4.4.1. *(**Shannon Entropy** $H(F)$ **of the distribution** F)*

$$H(F) \equiv -\int f(x) \log f(x) dx. \tag{4.4.1}$$

However, due to incomplete information (we do not know F really), we assume G the distribution of the random variables instead. Then the code would need more bits to represent the random variable.

The difference, in the number of bits, denoted by $\mathcal{R}(F;G)$, between a "true" probability distribution F and an arbitrary probability distribution G is known as the relative entropy [91], or the Kullback-Leibler divergence, information divergence, information gain in probability theory and information theory, which is expressed in mathematics as follows.

Definition 4.4.2. *(**Relative Entropy** $\mathcal{R}(F;G)$ **of probability distributions** F **and** G)*

$$\mathcal{R}(F;G) \equiv \int \log(f/g) dF, \tag{4.4.2}$$

where f and g are the respective densities with respect to any dominating measure.

Though the relative entropy $\mathcal{R}(F;G)$ is often taken as a distance metric, but it is not a true metric, since it is not symmetric between distribution F and G.

There exists some interpretations for relative entropy, for example, the relative entropy $\mathcal{R}(F;G)$ may be interpreted as the error exponent for the hypothesis test F versus G.

4.4.2 Classical Estimate

In a classical estimate, we are given classical observations distributed according to an unknown distribution $P_\theta \in X \in \ell$ associated with award Q, which Nature chooses randomly from the set $\{P_\theta : \theta \in \Theta\}$ according to a known prior distribution μ on Θ, we produce an estimate M for the unknown distribution P_θ, denoted by P later without notation confusion. In the end, we will suffer a relative entropy cost $\mathcal{R}(P;M)$, measuring the quality of this estimate, thus the whole utility is $P \cdot Q - \mathcal{R}(P;M)$.

In this section, we will investigate the existence of minimax value of this utility, correspondingly its minimax strategy.

We consider the utility $P \cdot Q - \mathcal{R}(P; M)$, then the estimate problem is in fact the following optimization problem

$$\min_{M \geq 0, M \cdot I = 1} \max_{P \geq 0, P \cdot I = 1} [P \cdot Q - \mathcal{R}(P; M)]. \tag{4.4.3}$$

Considering the convex conjugation $\mathcal{R}^*_M(Q)$ of $\mathcal{R}(P; M)$ with respect to P, that is,

$$\max_{P \geq 0, P \cdot I = 1} [P \cdot Q - \mathcal{R}(P; M)] = \mathcal{R}^*_M(Q), \tag{4.4.4}$$

we can rewrite the above estimate problem as follows.

$$\min_{M \geq 0, M \cdot I = 1} \max_{P \geq 0, P \cdot I = 1} [P \cdot Q - \mathcal{R}(P; M)] = \min_{M \geq 0, M \cdot I = 1} \mathcal{R}^*_M(Q). \tag{4.4.5}$$

Remark 4.4.3. *If we take function $\mathcal{R}_1(P)$ as follows.*

$$\mathcal{R}_1(P) = \begin{cases} \mathcal{R}(P; M) & P \cdot I = 1; \\ \infty & P \cdot I \neq 1. \end{cases}$$

Then we can write in the following form.

$$\max_{P \geq 0} [P \cdot Q - \mathcal{R}_1(P)] = \mathcal{R}^*_M(Q), \tag{4.4.6}$$

where $\mathcal{R}_1(P) = \mathcal{R}(P; M) + O_{A_1}(P)$, and the hyperplane $A_1 = \{P : P \cdot I = 1\}$.

Applying the Lagrange Theorem, we obtain the following result.

$$\mathcal{R}^*_M(Q) = \max_{P \geq 0, P \cdot I = 1} [P \cdot Q - \mathcal{R}_M(P)] = \min_{\lambda} \max_{P \geq 0} [P \cdot Q - \mathcal{R}_1(P) + \lambda(P \cdot I - 1)]. \tag{4.4.7}$$

Since the function $P \cdot Q - \mathcal{R}_1(P) + \lambda(P \cdot I - 1)$ is linear with respect to λ and convex with respect to P, the min and max can be exchanged, i.e.,

$$\min_{\lambda} \max_{P \geq 0} [P \cdot Q - \mathcal{R}_1(P) + \lambda(P \cdot I - 1)] = \max_{P \geq 0} \min_{\lambda} [P \cdot Q - \mathcal{R}_1(P) + \lambda(P \cdot I - 1)], \tag{4.4.8}$$

therefore we obtain the following result.

$$\mathcal{R}^*_M(Q) = \max_{P \geq 0, P \cdot I = 1} [P \cdot Q - \mathcal{R}_1(P)] = \max_{P \geq 0} \min_{\lambda} [P \cdot Q - \mathcal{R}_1(P) + \lambda(P \cdot I - 1)]. \tag{4.4.9}$$

Remark 4.4.4. *If considering this optimization problem, at first, with respect to λ, that is,*

$$\max_{P \geq 0} [P \cdot Q - \mathcal{R}_1(P) + O_{A_1}(P)] = \max_{P \geq 0} [P \cdot Q - \mathcal{R}_1(P) + \min_{\lambda} \lambda(P \cdot I - 1)], \tag{4.4.10}$$

we obtain the following result.

$$\mathcal{R}^*_M(Q) = \max_{P\geq 0}[P\cdot Q - \mathcal{R}_1(P) + \min_{\lambda}\lambda(P\cdot I - 1)]. \tag{4.4.11}$$

To consider the optimization problem

$$\min_{\lambda}\max_{P\geq 0}[P\cdot Q - \mathcal{R}_1(P) + \lambda(P\cdot I - 1)], \tag{4.4.12}$$

applying the variational method, it suffices to consider the function

$$U = P\cdot Q - \mathcal{R}_1(P) + \lambda(P\cdot I - 1), \tag{4.4.13}$$

or, equivalently, the function

$$\sum_x P_x(Q_x + \ln\frac{P_x}{M_x} + \lambda) - \lambda, \tag{4.4.14}$$

and we obtain the result.

$$\delta U = \delta P_x(Q_x + \ln\frac{M_x}{P_x} + \lambda - 1) + \delta\lambda(\sum_x P_x - 1), \tag{4.4.15}$$

from which we obtain the following result.

$$Q_x + \ln\frac{M_x}{P_x} + \lambda - 1 = 0. \tag{4.4.16}$$

Therefore, we reached following two results.

$$P^*_x = M_x\cdot\exp(Q_x + \lambda - 1), \tag{4.4.17}$$

$$\mathcal{R}^*_M(Q) = P^*\cdot Q - \mathcal{R}_M(P^*) = -\sum_x M_x\exp(Q_x + \lambda - 1)(\lambda - 1). \tag{4.4.18}$$

Considering further $P\cdot I = 1$, we obtain the following results.

$$\exp(\lambda - 1) = \frac{1}{\sum_x M_x\exp Q_x}, \tag{4.4.19}$$

$$P^*_x = \frac{M_x\cdot\exp Q_x}{\sum_x M_x\exp Q_x}, \tag{4.4.20}$$

$$\mathcal{R}^*_M(Q) = P^*\cdot Q - \mathcal{R}_M(P^*) = \ln(\sum_x M_x\exp Q_x). \tag{4.4.21}$$

Remark 4.4.5. *It is easy to see that $\mathcal{R}_M(P^*) = P^*\cdot Q - \mathcal{R}^*_M(Q)$ is the classical relative entropy under the maximal utility, or classical relative capacity under given utility.*

Since the set $\{M : M \geq 0, M\cdot I = 1\}$ is convex, the minimum of function $\mathcal{R}_M(P^*)$ always exists with respect to M.

Therefore, we reach the following main result.

Theorem 4.4.6. *The minimax value, associated with the above estimate game, defined by*

$$\overline{\mathrm{V}} = \inf_{M\geq 0, M\cdot I=1} \sup_{P\geq 0, P\cdot I=1} [P\cdot Q - \mathcal{R}(P;M)], \tag{4.4.22}$$

makes sense, and so does its minimax strategy.

Remark 4.4.7. *We can similarly define the maxmin value, associated with the above estimate game, by*

$$\underline{\mathrm{V}} = \sup_{P\geq 0, P\cdot I=1} \inf_{M\geq 0, M\cdot I=1} [P\cdot Q - \mathcal{R}(P;M)], \tag{4.4.23}$$

but it is needy to consider if this maxmin value and its maxmin strategy always exist, and further if this maxmin value is the same as the minimax value when this maxmin value always exists.

4.4.3 Convex Conjugate View

In the above, we applied the convex conjugate $\mathcal{R}^*_M(Q)$ of classical relative entropy $\mathcal{R}_M(P)$ with respect to P, i.e.,

$$\mathcal{R}^*_M(Q) = \max_{P\geq 0, P\cdot I=1} \{P\cdot Q - \mathcal{R}_M(P)\}, \tag{4.4.24}$$

and obtained the following formula

$$\mathcal{R}^*_M(Q) = \ln(\sum_x M_x \exp Q_x), \tag{4.4.25}$$

but starting from the result $\mathcal{R}^*_M(Q) = \ln(\sum_x M_x \exp Q_x)$ and applying the above biconjugate property, we can define the classical relative entropy as follows.

Definition 4.4.8. *Classical relative entropy $\mathcal{R}(\rho;M)$ of ρ relative to M is defined as*

$$\mathcal{R}(\rho;M) \equiv \mathcal{R}_M(\rho) = \max_Q \{<\rho, Q> - \mathcal{R}^*_M(Q)\}, \tag{4.4.26}$$

where $\mathcal{R}^*_M(Q) = \ln(\sum_x M_x \exp Q_x)$.

Obviously, the simple computation will give us the mathematical form of classical relative entropy.

Proposition 4.4.9. *Classical relative entropy of $\{\rho_x\}$ relative to $\{M_x\}$ is equal to $\sum_x [\rho_x \log(\rho_x / M_x)]$, that is,*

$$\mathcal{R}(\rho;M) = \sum_x [\rho_x \log(\rho_x / M_x)], \tag{4.4.27}$$

which confirms the unique mathematical form of classical relative entropy, though still open to explain $\mathcal{R}^*_M(Q) = \ln(\sum_x M_x \exp Q_x)$ completely in information theory.

4.5 Quantum Relative Entropy

Many information measures for quantum signals, for example, von Neumann entropy [80, 98], quantum conditional entropy [46], quantum relative entropy [97], quantum mutual entropy [79], etc, upon information theoretic explanation and mathematical formula of von Neumann entropy.

In fact, at present stage, no original information-theoretic definition, similar to shannon information entropy, is possible even for von Newmann entropy except for the mathematical formula.

This section is for one new and general quantum relative entropy, in the sense without Radon-Nikodym (RN) derivatives.

First we simply overview three types of quantum relative entropy, then investigate quantum estimate, in the frame of game theory, (here following the terminology of classical estimate, the same terminology of estimate is still used, but in a different sense away from quantum physics), and reach one new mathematical form of quantum relative entropy, also give some important properties including monotonicity of quantum relative entropy.

Throughout this chapter we prefer to use in what is following the term "information" for negaentropy, leaving the term "entropy" for the opposite quantities like relative negainformation $\mathcal{S}(\varpi;\varphi) = -\mathcal{R}(\varpi;\varphi)$, which coincides with usual von Newmann entropy $\mathcal{S}(\varpi)$ if it is taken with respect to the trace $\varphi = \mathrm{Tr}$.

4.5.1 Historic Review

There are several mathematical quantum relative entropies so far, mainly Araki-Umegaki type [2, 68, 97], Belavkin-Staszewski type [14], Hammersley-Belavkin type [39].

This part overviews quantum relative entropy of Hammersley-Belavkin type and reduces to Araki-Umegaki type, Belavkin-Staszewski type.

We define the Radon-Nikodym (RN) derivatives with respect to an arbitrary, not necessarily normalized, density $\gamma \in \mathcal{A}_\top$ (i.e. a positive linear functional on $\mathcal{A}$),

$$\varrho_\gamma = \tilde{\gamma}^{-1/2}\varrho\tilde{\gamma}^{-1/2}, \tag{4.5.1}$$

$$\varsigma_\gamma = \tilde{\gamma}^{-1/2}\varsigma\tilde{\gamma}^{-1/2}, \tag{4.5.2}$$

then we obtain quantum relative entropy of Hammersley-Belavkin type (known

as γ type elsewhere) as follows.

Definition 4.5.1. *Quantum relative entropy of Hammersley-Belavkin type (known as γ type elsewhere) to compound state ϖ on the algebra $\mathcal{A} \otimes \mathcal{B}$, (or information divergence of the state ϖ with respect to a reference state φ) is defined by the density operator ω, ϕ of these states ϖ and φ as*

$$\mathcal{R}^{g}_{\gamma}(\varpi;\varphi) = \lambda(\sqrt{\varsigma}g(\mathrm{L}^{-1}_{\varsigma_\gamma}\mathrm{L}'_{\varrho_\gamma})\sqrt{\varsigma}), \tag{4.5.3}$$

where $\mathrm{L}^{-1}_{\varsigma}\chi = \varsigma^{-1}\chi$ is the operator of left multiplication by ς^{-1} isomorphic to the operator $\varsigma^{-1} \otimes I_{\bar{A}}$, and $\mathrm{L}'_{\varrho}\chi = \chi\varrho$ is the right multiplication by ϱ isomorphic to the operator $\mathcal{I}_A \otimes \tilde{\varrho}$, $g(r)$ is any strictly positive function $g(r)$ at $r \geq 1$ with $g(1) = 0$ and operator-convex, λ is a faithful, semi-finite trace.

Note that this quantity is introduced in [39] in order to characterize the "distance" as an information divergence between states.

Quantum relative entropy of Hammersley-Belavkin type (known as γ type elsewhere) includes the other two relative quantum information as special cases.

Corollary 4.5.2. *If $\gamma = I$, it gives quantum relative entropy of Araki-Umegaki type (known as a type elsewhere) to compound state ϖ on the algebra $\mathcal{A} \otimes \mathcal{B}$, (or information divergence of the state ϖ with respect to a reference state φ) defined by the density operator ω, ϕ of these states ϖ and φ as*

$$\mathcal{R}^{(a)}(\varpi;\varphi) = \mathrm{Tr}[\omega(\ln\omega - \ln\phi)]. \tag{4.5.4}$$

This quantity is used in most definitions of quantum relative entropy.

However unlike the classical case, this is not only possible choice for informational divergence of the states ϖ and φ, and it does not relate explicitly the informational divergence to the Radon-Nikodym (RN) density $\omega_\varphi = \phi^{-1/2}\omega\phi^{-1/2}$ of the state ϖ with respect to φ as in the classical case.

Corollary 4.5.3. *If $\gamma = \sigma$, it gives quantum relative entropy of Belavkin-Staszewski type (known as b type elsewhere) introduced in [39] as*

$$\mathcal{R}^{(b)}(\varpi;\varphi) = \mathrm{Tr}[\omega\ln(\phi^{-1}\omega)], \tag{4.5.5}$$

where $\omega\ln(\phi^{-1}\omega) = \ln(\omega\phi^{-1})\omega$ is understood as the Hermitian operator

$$\omega^{1/2}\ln(\omega^{1/2}\phi^{-1}\omega^{1/2})\omega^{1/2}. \tag{4.5.6}$$

This relative entropy can be explicitly written in terms of the RN density ω_φ as

$$\mathcal{R}^{(b)}(\varpi;\varphi) = \varphi(r(\omega_\varphi)), \tag{4.5.7}$$

where $r(\omega_\varphi) = \omega_\varphi \ln \omega_\varphi$.

Remark 4.5.4. *In finite dimensions and faithful states, the Belavkin-Staszewski information divergence based on quantum relative entropy of Belavkin-Staszewski type gives better distinction of ω and ϕ [80] in the sense that it is greater than quantum relative entropy of Araki-Umegaki type, and that it satisfies the following important monotonicity property.*

For quantum relative entropy of the three types, it is easy to obtain the well-known monotonicity inequality.

Theorem 4.5.5. *Given a normal completely positive unital map $K : \mathcal{M} \to \mathcal{M}^0$, if $\varpi = \varpi_0 K$, $\varphi = \varphi_0 K$, then for both relative entropies,*

$$\mathcal{R}(\varpi;\varphi) \leq \mathcal{R}(\varpi_0;\varphi_0). \tag{4.5.8}$$

Remark 4.5.6. *In fact, this monotonicity property is proved in [44, 68] for Araki-Umegaki type (known as a type elsewhere). Belavkin-Staszewski type (known as b type elsewhere) and Hammersley-Belavkin type (known as γ type elsewhere) satisfies a system of properties [39] for quantum relative entropy including this monotonicity inequality. Therefore this monotonicity property holds for the three quantum relative entropies.*

Remark 4.5.7. *Quantum relative entropy of Hammersley-Belavkin type (known as γ type elsewhere) is more general quantum relative entropy, and includes Araki-Umegaki type (known as a type elsewhere) and Belavkin-Staszewski type (known as b type elsewhere), but can not exhaust all possibilities except for commutative algebra, for example, the trace distance $\mathcal{R}_{tr}(\varrho;\varsigma) = \lambda(|\varrho - \varsigma|)$ and the fidelity distance $\mathcal{R}_{fid}(\varrho;\varsigma) = 1 - \lambda(|\sqrt{\varrho}\sqrt{\varsigma}|)$ are not of Hammersley-Belavkin type (known as γ type elsewhere) [39].*

Undoubtedly, it is a challenge to obtain a general information-theoretic definition for quantum relative entropy which reduces to a general mathematical formula for quantum relative entropy.

4.5.2 Quantum Estimate

In a quantum estimate, we are given quantum observations distributed according to an unknown distribution $P_\theta \in X$ associated with award Q (where X

is a predual space of Hermitian W^*-continuous functionals on a W^*-algebra $X^* = \mathcal{A}$), which Nature chooses randomly from the set $\{P_\theta : \theta \in \Theta\}$ according to a known prior distribution μ on Θ, and we produce an estimate M for the unknown distribution P_θ. In the end, we will suffer a relative entropy cost $\mathcal{R}(P; M)$, measuring the quality of this estimate, thus the whole utility is taken as $P \cdot Q - \mathcal{R}(P; M)$.

There are several mathematical quantum relative entropies, mainly Araki-Umegaki type, Belavkin-Staszewski type, Hammersley-Belavkin type, and similarly three different mathematical forms of game theoretic utilities. Below we concentrate on quantum relative entropy of Araki-Umegaki type, since other two types of quantum relative entropy are just mathematical complex in Radon-Nikodym (RN) derivatives but with the same intrinsic method.

Applying the convex conjugate $\mathcal{R}^*_{M,\mu}(Q)$ of $\mathcal{R}(P; M)$ with respect to P, that is,

$$\mathcal{R}^*_{M,\mu}(Q) = \max_{P \geq 0, P \cdot I = \mu} [P \cdot Q - \mathcal{R}(P; M)], \tag{4.5.9}$$

the estimate problem is the following optimization problem

$$\min_{M \geq 0, M \cdot I = 1} \max_{P \geq 0, P \cdot I = \mu} [P \cdot Q - \mathcal{R}(P; M)] = \min_{M \geq 0, M \cdot I = 1} \mathcal{R}^*_{M,\mu}(Q), \tag{4.5.10}$$

where the cost function $\mathcal{R}(P; M)$ here is taken as quantum relative entropy of Araki-Umegaki type.

Remark 4.5.8. *If taking function $\mathcal{R}_\mu(P)$ as follows.*

$$\mathcal{R}_1(P) = \begin{cases} \mathcal{R}(P; M) & P \cdot I = \mu; \\ \infty & P \cdot I \neq \mu. \end{cases}$$

Then we can rewrite $\mathcal{R}^*_{M,\mu}(Q)$ as follows.

$$\max_{P \geq 0} [P \cdot Q - \mathcal{R}_\mu(P)] = \mathcal{R}^*_{M,\mu}(Q), \tag{4.5.11}$$

where $\mathcal{R}_1(P) = \mathcal{R}(P; M) + O_{A_1}(P)$, and the hyperplane $A_1 = \{P : P \cdot I = \mu\}$.

Applying the Lagrange Theorem, we will find.

$$\mathcal{R}^*_{M,\mu}(Q) = \max_{P \geq 0, P \cdot I = \mu} [P \cdot Q - \mathcal{R}_\mu(P)] = \min_\lambda \max_{P \geq 0} [P \cdot Q - \mathcal{R}_\mu(P) + \lambda(P \cdot I - \mu)]. \tag{4.5.12}$$

Since the function $P \cdot Q - \mathcal{R}_\mu(P) + \lambda(P \cdot I - \mu)$ is linear with respect to λ and convex with respect to P, therefore the min and max can be exchanged, that is,

$$\min_\lambda \max_{P \geq 0} [P \cdot Q - \mathcal{R}_\mu(P) + \lambda(P \cdot I - \mu)] = \max_{P \geq 0} \min_\lambda [P \cdot Q - \mathcal{R}_\mu(P) + \lambda(P \cdot I - \mu)], \tag{4.5.13}$$

we can rewrite $\mathcal{R}^*_{M,\mu}(Q)$ as follows.

$$\mathcal{R}^*_{M,\mu}(Q) = \max_{P\geq 0, P\cdot I=\mu}[P\cdot Q - \mathcal{R}(P;M)] = \max_{P\geq 0}\min_{\lambda}[P\cdot Q - \mathcal{R}_\mu(P) + \lambda(P\cdot I - \mu)]. \tag{4.5.14}$$

Remark 4.5.9. *Considering this optimization problem at first with respect to λ, then the following formula is reached.*

$$\mathcal{R}^*_{M,\mu}(Q) = \max_{P\geq 0}[P\cdot Q - \mathcal{R}_\mu(P) + \min_{\lambda}\lambda(P\cdot I - \mu)] = \max_{P\geq 0}[P\cdot Q - \mathcal{R}(P;M) + O_{A_1}(P)]. \tag{4.5.15}$$

Applying the variational method to the function $U = P\cdot Q - \mathcal{R}_1(P) + \lambda(P\cdot I - \mu)$, we obtain the following results.

$$Q + \ln M - \ln P + (\lambda - 1)I = 0, \tag{4.5.16}$$

$$\mathrm{Tr}P = \mu. \tag{4.5.17}$$

Simple calculations give us results as follows.

$$P^* = M\cdot\exp[Q + (\lambda - 1)I], \tag{4.5.18}$$

$$\mathcal{R}^*_{M,\mu}(Q) = P^*\cdot Q - \mathcal{R}_{M,\mu}(P^*) = -\mathrm{Tr}[M\exp(Q + \lambda - 1)(\lambda - 1)]. \tag{4.5.19}$$

Considering further about two conditions $P\cdot I = \mu$ and $M\cdot I = 1$, we will find following three results.

$$\exp(\lambda - 1) = \frac{\mu}{\mathrm{Tr}[M\exp Q]}, \tag{4.5.20}$$

$$P^* = \mu\frac{M\cdot\exp Q}{\mathrm{Tr}[M\exp Q]}, \tag{4.5.21}$$

$$\mathcal{R}^*_{M,\mu}(Q) = P^*\cdot Q - \mathcal{R}_{M,\mu}(P^*) = \mu\ln[\frac{1}{\mu}\mathrm{Tr}(M\exp Q)]. \tag{4.5.22}$$

Remark 4.5.10. *It is easy to see that $\mathcal{R}_{M,\mu}(P^*) = \mathcal{R}^*_{M,\mu}(Q) - P^*\cdot Q$ is the quantum relative entropy under the maximal utility over $P \geq 0$ and $P\cdot I = \mu$, or quantum relative capacity under given utility.*

It is one interesting problem to explain $\mathcal{R}^*_{M,\mu}(Q) = \mu\ln[\frac{1}{\mu}\mathrm{Tr}(M\exp Q)]$ completely in quantum information theory.

Easily we can find the following properties of $\mathcal{R}^*_{M,\mu}(Q)$.

Lemma 4.5.11. *The function $\mathcal{R}^*_{M,\mu}(Q) = \mu\ln[\frac{1}{\mu}\mathrm{Tr}(M\exp Q)]$ is monotonous with respect to M.*

Lemma 4.5.12. *The function* $\mathcal{R}^*_{M,\mu}(Q) = \mu \ln[\frac{1}{\mu}\mathrm{Tr}(M \exp Q)]$ *is concave with respect to* M.

Since the set $\{M : M \geq 0, M \cdot I = 1\}$ is convex, there always exists the minimum of function $\mathcal{R}_{M,\mu}(P^*)$ with respect to M, and upon the above lemmas, we reach the following main result.

Theorem 4.5.13. *The minimax value, associated with the above problem, defined by*

$$\overline{\mathrm{V}} = \inf_{M \geq 0, M \cdot I = 1} \sup_{P \geq 0, P \cdot I = \mu} [P \cdot Q - \mathcal{R}(P; M)], \tag{4.5.23}$$

makes sense, and so does its minimax strategy.

Remark 4.5.14. *We can similarly define the maxmin value, associated with the above problem, by*

$$\underline{\mathrm{V}} = \sup_{P \geq 0, P \cdot I = \mu} \inf_{M \geq 0, M \cdot I = 1} [P \cdot Q - \mathcal{R}(P; M)], \tag{4.5.24}$$

but it is a problem if this maxmin value and its maxmin strategy exist and if this maxmin value is equal to the minimax value.

4.5.3 Convex Conjugate View

In the above analysis, we applied convex conjugate $\mathcal{R}^*_M(Q)$ of quantum relative entropy $\mathcal{R}_M(P)$, that is,

$$\mathcal{R}^*_{M,\mu}(Q) = \max_{P \geq 0, P \cdot I = \mu} \{P \cdot Q - \mathcal{R}_M(P)\}, \tag{4.5.25}$$

and obtained the following formula

$$\mathcal{R}^*_{M,\mu}(Q) = \mu \ln[\frac{1}{\mu}\mathrm{Tr}(M \exp Q)], \tag{4.5.26}$$

but starting from the result $\mathcal{R}^*_{M,\mu}(Q) = \mu \ln[\frac{1}{\mu}\mathrm{Tr}(M \exp Q)]$, and applies the biconjugate property of convex conjugate, we can define the quantum relative entropy $\mathcal{R}_\mu(\rho; M)$ as follows.

Definition 4.5.15. *Quantum relative entropy* $\mathcal{R}_\mu(\rho; M)$ *of* ρ *relative to* M *is defined as*

$$\mathcal{R}_\mu(\rho; M) \equiv \mathcal{R}_{M,\mu}(\rho) = \max_Q \{< \rho, Q > - \mathcal{R}^*_{M,\mu}(Q)\}, \tag{4.5.27}$$

where $\mathcal{R}^*_{M,\mu}(Q) = \mu \ln[\frac{1}{\mu}\mathrm{Tr}(M \exp Q)]$.

In particular, for the case $\mu = 1$,

$$\mathcal{R}_1(\rho; M) \equiv \mathcal{R}_{M,1}(\rho) = \max_Q \{< \rho, Q > -\mathcal{R}^*_M(Q)\}, \tag{4.5.28}$$

where $\mathcal{R}^*_M(Q) = \ln[\mathrm{Tr}(M \exp Q)]$.

Obviously, the general mathematical form of this quantum relative entropy may not include the mathematical form of the Araki-Umegaki type, Belavkin-Staszewski type, or Hammersley-Belavkin type.

Here we obtain one property for this quantum relative entropy.

Theorem 4.5.16.

$$\forall \mu > 0, \mathcal{R}_\mu(\rho; M) = \mu \mathcal{R}_1(\frac{1}{\mu}\rho; \frac{1}{\mu}M). \tag{4.5.29}$$

Proof. According to the definition of quantum relative entropy,

$$\mathcal{R}_\mu(\rho; M) \equiv \mathcal{R}_{M,\mu}(\rho) \tag{4.5.30}$$

$$= \max_Q \{< \rho, Q > -\mu \ln[\frac{1}{\mu}\mathrm{Tr}M \exp Q)]\} \tag{4.5.31}$$

$$= \mu \max_Q \{< \frac{1}{\mu}\rho, Q > - \ln[\mathrm{Tr}(\frac{1}{\mu}M \exp Q)]\} \tag{4.5.32}$$

$$= \mu \mathcal{R}_1(\frac{1}{\mu}\rho; \frac{1}{\mu}M). \tag{4.5.33}$$

□

4.5.4 Monotonicity of Quantum Relative Entropy

To obtain its monotonicity of $\mathcal{R}_1(\rho; M)$ (or $\mathcal{R}(\rho; M)$ without notation confusion), we need several lemmas in our notation system.

Let ρ, ζ be positive trace class operators in a separable Hilbert space $\mathcal{H}$, Γ a trace-preserving from $\mathcal{B}(\mathcal{H})$ to a von Neumann subalgebra $\mathcal{A}$.

Lemma 4.5.17. *([53]) Let $\mathcal{A}$ be a finite von Neumann algebra. For each $X \in \mathcal{A}$ and each inner $^*-$automorphism α of $\mathcal{A}$, there exists $\overline{X} \in \mathcal{A}$ such that*

$$\frac{1}{N}\sum_{n=0}^{N-1} \alpha^n(X) \longrightarrow \overline{X} \tag{4.5.34}$$

as $N \longrightarrow \infty$ in the strong operator topology.

Lemma 4.5.18. *([88]) Let $\mathcal{A}'$ denote the group of all unitary transformations in $\mathcal{A}$, $E(\mathcal{A}')$ the set of all non-negative real valued functions on $\mathcal{A}'$ which vanish expect at a finite number of points of $\mathcal{A}'$ and which satisfy*

$$\sum_{U\in\mathcal{A}'} f(U) = I. \tag{4.5.35}$$

Take $f(X) = \sum_{U\in\mathcal{A}'} f(U)UXU^*$ for each bounded operator X, then for each bounded operator X, there exists a sequence $f_n \in E(\mathcal{A}')$ such that $f_n(X)$ converges weakly to an element of $\mathcal{A}'$.

Lemma 4.5.19. *Let $X \in T(\mathcal{H})$ be a trace class operators in a separable Hilbert space $\mathcal{H}$, and $K(X)$ the weakly closed convex hull of the set*

$$\{UXU^{-1}, \text{unitary} U \in \mathcal{A}'\}. \tag{4.5.36}$$

Then

$$K(X) \cap \mathcal{A} = \{\Gamma(X)\}, \tag{4.5.37}$$

$\{\Gamma(Y)\} = \{\Gamma(X)\}$ *for all $Y \in K(X)$.*

Moreover, let $E(\mathcal{A}')$ be the set of nonnegative real functions on the set $U(\mathcal{A}')$ of unitary operators in $\mathcal{A}'$ which are nonzero only on a finite number of points and which satisfy $\sum f(U) = I$. Take $f(X) = \sum f(U)UXU^{-1}$, then there is a sequence $\{f_n\} \subset E(\mathcal{A}')$ such that $f_n(X) \longrightarrow \Gamma(X)$.

Proof. Γ is normal, since the trace is normal, hence ultra-weakly continuous.

If $X \in T(\mathcal{H})$ be a trace class operators in a separable Hilbert space $\mathcal{H}$, $\Gamma(X)$ is the unique element of $\mathcal{A}$ such that

$$\mathrm{Tr}(\Gamma(X)Y) = \mathrm{Tr}(XY), \tag{4.5.38}$$

for all $Y \in \mathcal{A}$, which implies that $\Gamma(UXU^{-1}) = \Gamma(X)$ for all unitary $U \in \mathcal{A}'$, hence Γ is ultra-weakly continuous $\Gamma(X) = \Gamma(Y)$ for all $Y \in K(X)$.

The first statement results from Lemma 4.5.17; the last statement follows from Lemma 4.5.18. □

Lemma 4.5.20. *$\mathcal{R}(\rho;\zeta)$ is jointly convex in ρ and ζ: If $\lambda_i > 0, \sum\lambda_i = 1$, then*

$$\mathcal{R}(\sum\lambda_i\rho_i;\sum\lambda_i\zeta_i) \leq \sum\lambda_i\mathcal{R}(\rho_i;\zeta_i). \tag{4.5.39}$$

Proof. Since

$$\mathcal{R}_\mu(\rho;\zeta) \equiv \mathcal{R}_{\zeta,\mu}(\rho) \tag{4.5.40}$$

$$= \max_Q \{< \rho, Q > -\mu \ln[\frac{1}{\mu}\mathrm{Tr}(\zeta \exp Q)]\}, \tag{4.5.41}$$

$$\mathcal{R}(\sum \lambda_i \rho_i; \sum \lambda_i \zeta_i) = \max_Q \{< \sum \lambda_i \rho_i, Q > - \ln[\mathrm{Tr}(\sum \lambda_i \zeta_i \exp Q)]\} \tag{4.5.42}$$

$$\leq \max_Q \sum \lambda_i \{< \rho_i, Q > - \ln[\mathrm{Tr}(\zeta_i \exp Q)]\} \tag{4.5.43}$$

by additivity of inner product, $< \sum \lambda_i \rho_i, Q >= \sum \lambda_i < \rho_i, Q >$, and convexity of logarithm function,

$$= \sum \lambda_i \max_Q \{< \rho_i, Q > - \ln[\mathrm{Tr}(\zeta_i \exp Q)]\} \tag{4.5.44}$$

by the positivity of λ_i for all i,

$$= \sum \lambda_i \mathcal{R}(\rho_i; \zeta_i) \tag{4.5.45}$$

by the definition of $\mathcal{R}(\rho_i; \zeta_i)$. □

Lemma 4.5.21. *Let P be a projection in $\mathcal{H}$, and take $\rho_P \equiv P\rho P$, etc..*

Then

$$\mathcal{R}(\rho_P; \zeta_P) + \mathcal{R}(\rho_{I-P}; \zeta_{I-P}) \leq \mathcal{R}(\rho; \zeta). \tag{4.5.46}$$

Proof. Note that $U = 2P - I$ is unitary and that

$$\rho' \equiv \rho_P + \rho_{I-P} = \frac{1}{2}(\rho + U^+ \rho U). \tag{4.5.47}$$

Following Lemma 4.5.20, we have

$$\mathcal{R}(\rho'; \zeta') \leq \frac{1}{2}\mathcal{R}(\rho; \zeta) + \frac{1}{2}\mathcal{R}(U^+ \rho U; U^+ \zeta U) = \mathcal{R}(\rho; \zeta). \tag{4.5.48}$$

Note that

$$\mathcal{R}(\rho'; \zeta') \geq \mathcal{R}(\rho_P; \zeta_P) + \mathcal{R}(\rho_{I-P}; \zeta_{I-P}), \tag{4.5.49}$$

we reach the result

$$\mathcal{R}(\rho_P; \zeta_P) + \mathcal{R}(\rho_{I-P}; \zeta_{I-P}) \leq \mathcal{R}(\rho; \zeta). \tag{4.5.50}$$

□

Lemma 4.5.22. *Let P_n be a sequence of projections such that $P_m \leq P_n$ for $m \leq n$, $\dim P_n$ is finite for all n, and $P_n \longrightarrow I$ strongly when $n \longrightarrow \infty$. Take $\rho_n = P_n \rho P_n$.*

Then the sequences $\{\mathcal{R}(\rho_n; \zeta_n)\}$ are monotonously increasing and

$$\mathcal{R}(\rho_n; \zeta_n) \longrightarrow \mathcal{R}(\rho; \zeta). \tag{4.5.51}$$

Proof. The monotonicity of $\{\mathcal{R}(\rho_n;\zeta_n)\}$ follows from Lemma 4.5.21.

Recalling the facts,

$$\mathrm{Tr}(P_n\rho^2) \longrightarrow \mathrm{Tr}\rho^2, \tag{4.5.52}$$

and

$$0 \le \mathrm{Tr}[P_n(\rho_n^2-\rho^2)] = \mathrm{Tr}[P_n\rho(I-P_n)\rho] \le \mathrm{Tr}[\rho^2(I-P_n)] \longrightarrow 0, \tag{4.5.53}$$

we have the result

$$\mathrm{Tr}(\rho-\rho_n)^2 = \mathrm{Tr}(\rho^2-\rho_n^2) = \mathrm{Tr}[\rho^2(I-P_n)] + \mathrm{Tr}[P_n(\rho_n^2-\rho^2)] \longrightarrow 0, \tag{4.5.54}$$

but $\|\rho-\rho_n\|^2 \le \mathrm{Tr}(\rho-\rho_n)^2$, consequently $\|\rho-\rho_n\| \longrightarrow 0$, that is, the convergence $\rho_n \longrightarrow \rho$ is uniform.

Note that $\mathcal{R}(\rho;\zeta)$ is lower semi-continuous under the convergence $(\rho_n;\zeta_n) \longrightarrow (\rho;\zeta)$,

$$\mathcal{R}(\rho;\zeta) \le \liminf \mathcal{R}(\rho_n;\zeta_n), \tag{4.5.55}$$

but from Lemma 4.5.21, we know that $\mathcal{R}(\rho_n;\zeta_n) \le \mathcal{R}(\rho;\zeta)$, hence we have

$$\liminf \mathcal{R}(\rho_n;\zeta_n) = \mathcal{R}(\rho;\zeta). \tag{4.5.56}$$

□

Lemma 4.5.23. *Assume that* $\{f_k\} \subset E(\mathcal{A}')$ *satisfies*

$$f_k(\rho) \longrightarrow \Gamma(\rho), \tag{4.5.57}$$

$$f_k(\zeta) \longrightarrow \Gamma(\zeta) \tag{4.5.58}$$

weakly.

Then

$$\mathcal{R}(\Gamma(\rho);\Gamma(\zeta)) \le \liminf \mathcal{R}(f_k(\rho);f_k(\zeta)) \le \mathcal{R}(\rho;\zeta). \tag{4.5.59}$$

Proof. Since Γ is trace-preserving, use the spectral measure of $\Gamma(\rho)$ (where positive operator ρ has the support projection I), there is a sequence of projections P_n in $\mathcal{A}$ satisfying the conditions of Lemma 4.5.22.

From the definition of Γ, we have

$$\Gamma(P_n\rho P_n) = P_n\Gamma(\rho)P_n. \tag{4.5.60}$$

Since f_k is built up of elements of $\mathcal{A}'$, we result that

$$f_k(P_n\rho P_n) = P_n\Gamma[f_k(\rho)]P_n. \tag{4.5.61}$$

In the finite dimensional space, $\mathcal{H}_n = P_n\mathcal{H}$, $f_k(\rho_n) \longrightarrow \Gamma(\rho_n)$ is convergent uniformly, hence when $k \longrightarrow \infty$,

$$\mathcal{R}(f_k(\rho_n); f_k(\zeta_n)) \longrightarrow \mathcal{R}(\Gamma(\rho_n); \Gamma(\zeta_n)). \tag{4.5.62}$$

According to Lemma 4.5.22, we obtain that

$$\mathcal{R}(\rho;\zeta) = \sup \mathcal{R}(\rho_n;\zeta_n), \tag{4.5.63}$$

hence $\mathcal{R}(\rho;\zeta)$ is lower semi-continuous, that is,

$$\mathcal{R}(\Gamma(\rho);\Gamma(\zeta)) \leq \liminf \mathcal{R}(f_k(\rho); f_k(\zeta)). \tag{4.5.64}$$

Following Lemma 4.5.20, we have

$$\mathcal{R}(f_k(\rho); f_k(\zeta)) \leq \sum f_k(U)\mathcal{R}(U\rho U^+; U\zeta U^+) \tag{4.5.65}$$

$$= \mathcal{R}(\rho;\zeta) \tag{4.5.66}$$

by the unitary invariance of $\mathcal{R}$.

Hence,

$$\mathcal{R}(\Gamma(\rho);\Gamma(\zeta)) \leq \mathcal{R}(\rho;\zeta). \tag{4.5.67}$$

□

Remark 4.5.24. *According to Lemma 4.5.23, it is hard to prove Monotonicity Theorem since we do not know if there is a sequence $\{f_k\} \subset E(\mathcal{A}')$ which implements Γ on both ρ and ζ.*

Theorem 4.5.25. *(**Monotonicity**) For trace-preserving expectations Γ from $\mathcal{B}(\mathcal{H})$ to a von Neumann subalgebra $\mathcal{A}$, if ρ and ζ are positive trace class operators in a separable Hilbert space $\mathcal{H}$.*

Then

$$\mathcal{R}(\Gamma(\rho);\Gamma(\zeta)) \leq \mathcal{R}(\rho;\zeta). \tag{4.5.68}$$

Proof. Choose a sequence of projections $P_n \in \mathcal{A}$, such that $P_m \leq P_n$ for $m \leq n$, $\dim P_n$ is finite for all n, and $P_n \longrightarrow I$ strongly when $n \longrightarrow \infty$, and $\{f_k\} \subset E(\mathcal{A}')$ such that $f_k(\rho) \longrightarrow \Gamma(\rho)$ weakly.

Then

$$f_k(\rho_n) \longrightarrow \Gamma(\rho_n) \tag{4.5.69}$$

in norm.

For a given k, there exists $g_j \subset E(\mathcal{A}')$ such that $g_j[f_k(\zeta)] \longrightarrow \Gamma(\zeta)$ weakly when $j \longrightarrow \infty$, then

$$g_j[f_k(\zeta_n)] \longrightarrow \Gamma(\zeta_n) \tag{4.5.70}$$

in norm.

If $\|(f_k - \Gamma)\rho_n\| \leq \varepsilon(k)$, choose $g_{j,k}$ such that

$$\|(g_{j,k}f_k - \Gamma)(\zeta_n)\| \leq \varepsilon(k). \tag{4.5.71}$$

Obviously,

$$\|(g_{j,k}f_k - \Gamma)(\rho_n)\| = \|g_{j,k}(f_k - \Gamma)(\rho_n)\| \leq \|(f_k - \Gamma)(\rho_n)\| \leq \varepsilon(k). \tag{4.5.72}$$

Then $h_k \equiv g_{j,k}f_k$ satisfies

$$h_k(\rho_n) \longrightarrow \Gamma(\rho_n), \tag{4.5.73}$$

$$h_k(\zeta_n) \longrightarrow \Gamma(\zeta_n) \tag{4.5.74}$$

in norm.

According to the proof of Lemma 4.5.23, we have

$$\mathcal{R}(\Gamma(\rho_n); \Gamma(\zeta_n)) \leq \mathcal{R}(\rho_n; \zeta_n), \tag{4.5.75}$$

and further following Lemma 4.5.22,

$$\mathcal{R}(\Gamma(\rho); \Gamma(\zeta)) \leq \mathcal{R}(\rho; \zeta). \tag{4.5.76}$$

□

Corollary 4.5.26. *Let P_k be a set of mutually orthogonal projections in $\mathcal{H}$ satisfying $\sum P_k = I$ and the map $\Gamma : \rho \longrightarrow \sum P_k \rho P_k$ is trace-preserving describing the interaction of a finite quantum system with a classical apparatus measuring an observable with eigen-spaces P_k.*

Then

$$\mathcal{R}(\Gamma(\rho); \Gamma(\zeta)) = \sum \mathcal{R}(P_k \rho P_k; P_k \zeta P_k) \leq \mathcal{R}(\rho; \zeta). \tag{4.5.77}$$

4.6 Quantum Mutual Information

Quantum mutual information was discussed in [30] via entropy diagram, which seems not starting originally from the information-theoretic sense.

In the view of quantum control, quantum mutual information is extensively researched, starting from [7] and more recently [16–19, 87]. Belavkin and Ohya

[18, 19] introduced quantum mutual information as the von Neumann negaentropy $\mathcal{R}(\varpi) = -\mathcal{S}(\varpi)$ of the entangled compound state related to negaentropy $\mathcal{R}(\varrho \otimes \varsigma) = -\mathcal{S}(\varrho \otimes \varsigma)$ of the product of marginal states, i.e. as the relative negaentropy $\mathcal{R}^{(a)}(\varpi;\varphi) = -\mathcal{S}^{(a)}(\varpi;\varphi)$, in the sense of Lindblad, Araki and Umegaki relative entropy [2, 68, 97] with respect to $\varphi = \varrho \otimes \varsigma$.

Naturally this approach treats quantum mutual information via entanglement based on quantum relative entropy, therefore we can further define quantum mutual information via quantum entanglement.

Definition 4.6.1. *We define the quantum mutual information $\mathcal{I}_{\mathcal{A},\mathcal{B}}(\pi) = \mathcal{I}_{\mathcal{B},\mathcal{A}}(\pi^*)$ of both types in a compound state ω achieved by a quantum entanglement $\pi : \mathcal{B} \to \mathcal{A}_*$, or by $\pi^* : \mathcal{A} \to \mathcal{B}_*$ with*

$$\varrho(A) = \varpi(A \otimes I) = \mathrm{Tr}_{\mathcal{G}}[A\rho], \varsigma(B) = \varpi(I \otimes B) = Tr_{\mathcal{H}}[B\sigma] \tag{4.6.1}$$

as the quantum relative entropy of the state ϖ on $\mathcal{M} = \mathcal{A} \otimes \mathcal{B}$ with the respect to the product state $\phi = \varrho \otimes \varsigma$:

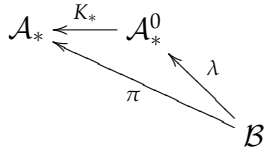
$$\mathcal{I}_{\mathcal{A},\mathcal{B}}(\pi) = \mathcal{R}(\omega; \rho \otimes \sigma). \tag{4.6.2}$$

Theorem 4.6.2. *Let $\lambda : \mathcal{B} \to \mathcal{A}^0_*$ be an entanglement of the state $\sigma(B) = \mathrm{Tr}[\lambda(B)]$ to $(\mathcal{A}^0, \rho^0)$ with $\mathcal{A}^0 \subseteq \mathcal{L}(\mathcal{G}_0)$, $\varrho^0 = \lambda(I)$ on $\mathcal{B}$, and $\pi = \mathrm{K}_*\lambda$ be entanglement to the state $\rho = \rho^0\mathrm{K}$ on $\mathcal{A} \subseteq \mathcal{G}$ defined as the composition of λ with the predual operator $\mathrm{K}_* : \mathcal{A}^0_* \to \mathcal{A}_*$ normal completely positive unital map $\mathrm{K} : \mathcal{A} \to \mathcal{A}^0$.*

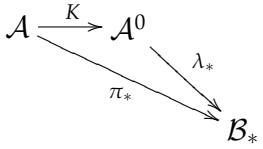
Then the following monotonicity inequality holds

$$\mathcal{I}_{\mathcal{A},\mathcal{B}}(\pi) \leq \mathcal{I}_{\mathcal{A}^0,\mathcal{B}}(\lambda). \tag{4.6.3}$$

Proof. This follows from the commutativity of the following diagrams:



Commutative diagram for entanglement π



Dual commutative diagram for entanglement π_*

Applying the monotonicity property of our new quantum relative entropy on $\mathcal{M} = \mathcal{A} \otimes \mathcal{B}$ with respect to the predual map $\omega_0 \mapsto (\mathrm{K}_* \otimes \mathrm{Id})(\omega_0)$ corresponding to $\varpi_0 \mapsto \varpi_0(\mathrm{K} \otimes \mathrm{Id})$ as the ampliation $\mathrm{K} \otimes \mathrm{Id}$ of a normal completely positive unital map $\mathrm{K} : \mathcal{A} \to \mathcal{A}^0$. □

Definition 4.6.3. *The maximal quantum mutual entropy $\mathcal{J}_{\tilde{\mathcal{B}},\mathcal{B}}(\pi_q)$ as the supremum*

$$H_{\mathcal{B}}(\varsigma) = \sup_{\pi^*(I)=\sigma} \mathcal{I}_{\mathcal{B},\mathcal{A}}(\pi^*) = \mathcal{J}_{\mathcal{B},\tilde{\mathcal{B}}}(\pi_q^*) \tag{4.6.4}$$

over all entanglements π^* of any $(\mathcal{A},\varrho)$ to $(\mathcal{B},\varsigma)$ is achieved on $\mathcal{A}^0 = \tilde{\mathcal{B}}$, $\varrho^0 = \tilde{\varsigma}$ by the standard quantum entanglement $\pi_q^*(A) = \sigma^{1/2}\tilde{A}\sigma^{1/2}$ for a fixed $\varsigma(B) = \mathrm{Tr}_{\mathcal{H}}[B\sigma]$ is named as entangled, or true quantum entropy of each type of the state ς.

Similar definition for Araki-Umegaki type can be found in [16–19], for Belavkin-Staszewski type in [21].

Definition 4.6.4. *We call the positive difference*

$$H_{\mathcal{B}|\mathcal{A}}(\pi) = H_{\mathcal{B}}(\varsigma) - \mathcal{I}_{\mathcal{A},\mathcal{B}}(\pi) \tag{4.6.5}$$

entangled (or true quantum) conditional entropy respectively of each type on $\mathcal{B}$ with respect to $\mathcal{A}$.

Similar definition for Araki-Umegaki type can be found in [16–19], for Belavkin-Staszewski type in [21].

4.7 Quantum Communication Channel

Entanglement-assisted quantum channel capacity, or entangled quantum channel capacity is extensively researched, for example, entangled quantum capacity [16–19] via a common framework, entanglement-assisted quantum capacity [24, 25] via entangled quantum mutual information.

We further discuss entangled quantum channel capacity via quantum mutual information upon our quantum relative entropy and its additivity property.

4.7.1 Quantum Channel Capacity

Let $\mathcal{B} \subseteq \mathcal{L}(\mathcal{H})$ be the W^*-algebra of operators in a (not necessarily finite dimensional unitary) Hilbert space $\mathcal{H}$.

Generally we denote the set of states, i.e. positive unit trace operators in $\mathcal{B}(\mathcal{H})$ by $\mathcal{S}(\mathcal{H})$, the set of all m-dimensional projections by $\mathcal{P}_m(\mathcal{H})$ and the set of all projections by $\mathcal{P}(\mathcal{H})$.

Definition 4.7.1. *A quantum channel Γ is a normal unital completely positive linear map (UCP) of $\mathcal{B}$ into the same or another algebra $\mathcal{B}^0 \subseteq \mathcal{B}(\mathcal{H}^0)$.*

These maps admit the Kraus decomposition, which is usually written in terms of the dual map $\Gamma^* : \mathcal{B}_*^0 \to \mathcal{B}_*$ as $\Gamma^*(\sigma^0) = \sum_k A_k \sigma^0 A_k^* \equiv \Gamma_*(\sigma^0)$ ([93], [68], [44]), $\Gamma(B) = \sum_k A_k^* B A_k$, for A_k are operators $\mathcal{H}^0 \to \mathcal{H}$ satisfying $\sum_k A_k^* A_k = I^0$.

For example, quantum noiseless channel in the case $\mathcal{B} = \mathcal{L}(\mathcal{H})$, $\mathcal{B}^0 = \mathcal{L}(\mathcal{H}^0)$ is described by a single isometric operator $Y : \mathcal{H}^0 \to \mathcal{H}$ as $\Gamma(B) = Y^* B Y$. See for example [44, 69] for the simple cases $\mathcal{B} = \mathcal{L}(\mathcal{H})$, $\dim(\mathcal{H}) < \infty$.

A noisy quantum channel sends input pure states $\varsigma_0 = \varrho_0$ on the algebra $\mathcal{B}^0 = \mathcal{L}(\mathcal{H}^0)$ into mixed states described by the output densities $\sigma = \Gamma^*(\sigma^0)$ on $\mathcal{B} \subseteq \mathcal{L}(\mathcal{H})$ given by the predual $\Gamma_* = \Gamma^* \mid \mathcal{B}_*^0$ to the normal completely positive unital map $\Gamma : \mathcal{B} \to \mathcal{B}^0$ which can always be written as

$$\Gamma(B) = \mathrm{Tr}_{\mathcal{F}_+}[Y^\dagger B Y]. \tag{4.7.1}$$

Here Y is a linear operator from $\mathcal{H}^0 \otimes \mathcal{F}_+$ to $\mathcal{H}$ with $\mathrm{Tr}_{\mathcal{F}_+}[Y^\dagger Y] = I$, and $\mathcal{F}_+$ is a separable Hilbert space of quantum noise in the channel.

Each input mixed state ς^0 is transmitted into an output state $\varsigma = \varsigma^0 \Gamma$ given by the density operator

$$\Gamma^*(\varsigma^0) = Y(\sigma^0 \otimes I_+)Y^\dagger \in \mathcal{B}_* \tag{4.7.2}$$

for each density operator $\sigma^0 \in \mathcal{B}_*^0$, the identity operator $I_+ \in \mathcal{F}_+$.

We follow [16–19] to denote $\mathcal{K}_q$ the set of all normal completely positive maps $\kappa : \mathcal{A} \to \mathcal{B}^0$ with any probe algebra $\mathcal{A}$, normalized as $\mathrm{Tr}\kappa(I) = 1$, and $\mathcal{K}_q(\varsigma^0)$ be the subset of $\kappa \in \mathcal{K}_q$ with $\kappa(I) = \sigma^0$.

We take the standard entanglement π_q^0 on $(\mathcal{B}^0, \varsigma^0) = (\mathcal{A}_0, \varrho^0)$, where $\varrho_0(A_0) = \mathrm{Tr}[A_0 \rho_0]$ given by the density operator $\rho_0 = \sigma^0$, and denote by K a normal unital completely positive map $\mathcal{A} \to \mathcal{A}^0 = \widetilde{\mathcal{A}}_0$ that decomposes κ as $\kappa(A) = \rho_0^{1/2} \widetilde{K(A)} \rho_0^{1/2}$.

It defines an input entanglement $\kappa^* = K_* \pi_q^0$ on the input of quantum channel as transpose-completely positive map on $\mathcal{A}_0 = \mathcal{B}^0$ into $\mathcal{A}_*$ normalized to $\rho = K_* \rho^0$, $\rho^0 = \widetilde{\rho}_0$.

The channel Γ transmits this input entanglement as a true quantum encoding into the output entanglement $\pi = K_* \pi_q^0 \Gamma \equiv K_* \lambda$ mapping $\mathcal{B}$ via the channel Γ into $\mathcal{A}_*$ with $\pi(I) = \rho$. The mutual entangled information, transmitted via the channel for quantum encoding κ is therefore $\mathcal{J}_{\mathcal{A},\mathcal{B}}(\kappa^* \Gamma) = \mathcal{J}_{\mathcal{A},\mathcal{B}}(K_* \pi_q^0 \Gamma) = \mathcal{J}_{\mathcal{A},\mathcal{B}}(K_* \Gamma)$, where $\Gamma = \pi_q^0 \Gamma$ is the standard input entanglement $\pi_q^0(B) = \sigma_0^{1/2} \tilde{B} \sigma_0^{1/2}$ with $\sigma_0 = \widetilde{\sigma}^0$, transmitted via the channel Γ.

Theorem 4.7.2. *Given a quantum channel $\Gamma : \mathcal{B} \to \mathcal{B}^0$, and an input state ς^0 on $\mathcal{B}^0$,*

the entangled input-output quantum information via a channel $\Gamma : \mathcal{B} \to \mathcal{B}^0$ *achieves the maximal value*

$$\mathcal{J}(\varsigma^0, \Gamma) = \sup_{\kappa \in \mathcal{K}_q(\varsigma^0)} \mathcal{J}(\kappa^* \Gamma) = \mathcal{I}_{\mathcal{A}^0, \mathcal{B}}(\lambda), \tag{4.7.3}$$

where $\lambda = \pi_q^0 \Gamma$ *is given by the corresponding extremal input entanglement* π_q^0 *mapping* $\mathcal{B}^0 = \tilde{\mathcal{A}}^0$ *into* $\mathcal{A}^0 = \tilde{\mathcal{B}}^0$ *with* $\mathrm{Tr}[\pi_q(B)] = \varsigma^0(B)$ *for all* $B \in \mathcal{B}^0$.

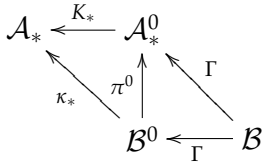
Note that simliar theorem for Araki-Umegaki type can be found in [16–19], for Belavkin-Staszewski type in [21].

Proof. Given to the monotonicity

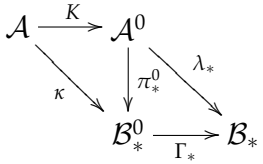
$$\mathcal{R}(\varpi_{01}(\mathrm{K} \otimes \Gamma); \varrho_0 \mathrm{K} \otimes \mathrm{Tr}) \leq \mathcal{R}(\varpi_{01}(I \otimes \Gamma); \varrho_0 \otimes \mathrm{Tr}), \tag{4.7.4}$$

the supremum of $\mathcal{J}(\kappa^* \Gamma)$ over all $\kappa \in \mathcal{K}_q(\varsigma^0)$ is achieved on the standard entanglement $\mathcal{B} \to \mathcal{A}^0$ given by $\kappa^* = \pi_q^0 \equiv \kappa^0$. □

The following definition depends on the commutativity of diagrams:



Commutative diagram for quantum channel Γ with standard entanglement $\pi^0 = \pi_q^0$ for $\mathcal{A} = \widetilde{\mathcal{B}}^0$



Dual commutative diagram for quantum channel Γ with standard entanglement π_*^0 for $\mathcal{A}_*^0 = \widetilde{\mathcal{B}}_*^0$

Definition 4.7.3. *Given a quantum channel* $\Gamma : \mathcal{B} \to \mathcal{B}^0$*, and a input state* ς^0 *on* $\mathcal{B}^0$*, we can define the input-output quantum entropy as the maximal quantum mutual information*

$$\mathcal{J}(\varsigma^0, \Gamma) = \mathcal{I}_{\mathcal{B}^0, \mathcal{B}}(\pi_q^0 \Gamma) \tag{4.7.5}$$

for input standard entanglement of the state ς^0 *to the state* $\varrho^0 = \tilde{\varsigma}^0$.

Note that similar definition for Araki-Umegaki type can be found in [16–19], for Belavkin-Staszewski type in [21].

4.7.2 Additivity of Quantum Channel Capacity

Here and below for notational simplicity we implement the agreements $\mathcal{A}_0^i = \mathcal{B}_i^0$, $\varrho_0^i = \varsigma_i^0$, $\mathcal{A}_0^{\otimes} = \otimes_{i=1}^n \mathcal{B}_i^0$, $\varrho_0^{\otimes} = \otimes_{i=1}^n \varsigma_i^0$ such that $\varsigma_0^{\otimes} = \otimes_{i=1}^n \rho_i^0$ is transposed

input state $\widetilde{\varrho}_0^{\otimes} = \otimes_{i=1}^n \widetilde{\varsigma}_i^0$ on $\mathcal{B}_0^{\otimes} = \otimes_{i=1}^n \mathcal{A}_i^0$ with $\widetilde{\mathcal{B}}_i^0 = \mathcal{A}_i^0 \equiv \mathcal{B}_0^i = \widetilde{\mathcal{A}}_0^i$, $\widetilde{\varsigma}_i^0 = \varrho_i^0 \equiv \varsigma_0^i = \widetilde{\varrho}_0^i$.

Let Γ_i be channels respectively from the algebra $\mathcal{B}_i$ on $\mathcal{H}_i$ to $\mathcal{B}_i^0$ on $\mathcal{H}_i^0$ for $i = 1, 2, ..., n$, and let $\Gamma^{\otimes} = \otimes_{i=1}^n \Gamma_i$ be their tensor product.

We have the additivity property of this entangled input-output quantum entropy upon the monotonicity property.

Theorem 4.7.4. *Let $\Gamma^{\otimes}$ be product channel from the algebra $\mathcal{B}^{\otimes} = \otimes_{i=1}^n \mathcal{B}_i$ to $\mathcal{A}_0^{\otimes} = \otimes_{i=1}^n \mathcal{A}_0^i$, and let $\varrho_0^{\otimes} = \otimes_{i=1}^n \varrho_0^i$ be the tensor product of input states ς_0^i on $\mathcal{B}_0^i$.*

Then

$$\mathcal{J}(\varrho_0^{\otimes}, \Gamma^{\otimes}) = \sum_{i=1}^n \mathcal{J}(\varrho_0^i, \Gamma_i). \tag{4.7.6}$$

Proof. Take $\Gamma_{i*} : \mathcal{B}_{i*}^0 \to \mathcal{B}_{i*}$, and $\varrho_0^i \in \mathcal{B}_{i*}^0$, $\varsigma_i = \Gamma_{i*}(\varrho_0^i) \in \mathcal{B}_{i*}$, and $K_*^{(n)} : \mathcal{A}_*^{\otimes} \to \mathcal{A}_*^{(n)}$, where $\mathcal{A}_{0*}^{\otimes} = \otimes_{i=1}^n \mathcal{B}_{i*}^0$, but $\mathcal{A}_*^{(n)}$ is predual to a general, not necessarily product algebra $\mathcal{A}^{(n)} \subseteq \mathcal{L}(\mathcal{G}^{(n)})$.

For $\pi^{(n)} = K_*^{(n)} \pi_q^{0\otimes} \Gamma^{\otimes}$, we consider quantum mutual entropy $\mathcal{I}_{\mathcal{A}^{(n)}, \mathcal{B}^{\otimes}}(\pi^{(n)})$ as quantum relative entropy

$$\mathcal{R}((K_*^{(n)} \otimes \Gamma_*^{\otimes}) \widetilde{\omega}_0^{\otimes}; K_*^{(n)}(\sigma_0^{\otimes}) \otimes \Gamma_*^{\otimes}(\varrho_0^{\otimes})), \tag{4.7.7}$$

where $\widetilde{\omega}_0^{\otimes} = \otimes_{i=1}^n \widetilde{\omega}_0^i$ is the density operator of the standard compound state $\otimes_{i=1}^n \omega_0^i$ with $\omega_0^i(A_i \otimes B_i) = \omega_i^0(A_i \otimes B_i) = \mathrm{Tr}[B_i \sqrt{\rho_i^0} \widetilde{A}_i \sqrt{\rho_i^0}]$ for $A_i \in \widetilde{\mathcal{B}}_i^0, B_i \in \mathcal{B}_i^0$, corresponding to $\sigma_i^0 = \rho_0^i$.

Applying monotonicity property of quantum relative entropy to the probe system $(\mathcal{G}^{(n)}, \mathcal{A}^{(n)})$ for this given ϱ_0^i and Γ_i, we obtain

$$\mathcal{R}((K_*^{(n)} \otimes \Gamma_*^{\otimes}) \widetilde{\omega}_0^{\otimes}; K_*^{(n)}(\sigma_0^{\otimes}) \otimes \Gamma_*^{\otimes}(\rho_0^{\otimes})) \tag{4.7.8}$$

$$\leq \mathcal{R}((Id^{\otimes} \otimes \Gamma^{\otimes}) \widetilde{\omega}_0^{\otimes}; Id^{\otimes}(\sigma_0^{\otimes}) \otimes \Gamma_*^{\otimes}(\rho_0^{\otimes})) \tag{4.7.9}$$

$$= \sum_{i=1}^n \mathcal{R}((Id \otimes \Gamma_{i*})(\widetilde{\omega}_0); Id(\sigma_0^i) \otimes \Gamma_{i*}(\rho_0^i)), \tag{4.7.10}$$

where $\sigma_0^i = \rho_i^0 = \widetilde{\rho}_0^i$, $\rho_0^i = \sigma_i^0 = \widetilde{\sigma}_0^i$.

The suprema over $K^{(n)}$ is achieved on $K^{(n)} = Id^{\otimes}$ identically mapping $\mathcal{A}^{(n)} = \otimes_{i=1}^n \mathcal{A}_0^i$ to $\mathcal{B}_{0*}^{\otimes} = \otimes_{i=1}^n \mathcal{B}_0^i$, where $\mathcal{B}_0^i = \widetilde{\mathcal{B}}_i^0$, coinciding with such $\mathcal{A}^{(n)}$ due to $\mathcal{A}_0^i = \widetilde{\mathcal{B}}_i^0$.

Thus $\mathcal{J}(\varrho_0^{\otimes}, \Gamma^{\otimes}) = \sum_{i=1}^n \mathcal{J}(\varrho_0^i, \Gamma_i)$. □

Let $\Gamma^{\otimes}$ be product channel from the algebra $\mathcal{B}^{\otimes} = \otimes_{i=1}^{n}\mathcal{B}_i$ to $\mathcal{A}_0^{\otimes} = \otimes_{i=1}^{n}\mathcal{B}_i^0$.

The additivity problem for quantum channel capacity via entanglement is if it is true that

$$\mathcal{C}_q(\Gamma^{\otimes}) = \sum_{i=1}^{n}\mathcal{C}_q(\Gamma_i). \tag{4.7.11}$$

In the spirit of [16–19], we prove this additivity property upon the monotonicity property of our quantum relative entropy.

Theorem 4.7.5. *Let $\Gamma^{\otimes}$ be product channel from the algebra $\mathcal{B}^{\otimes} = \otimes_{i=1}^{n}\mathcal{B}_i$ to $\mathcal{A}_0^{\otimes} = \otimes_{i=1}^{n}\mathcal{B}_i^0$. Then*

$$\mathcal{C}_q(\Gamma^{\otimes}) = \sum_{i=1}^{n}\mathcal{C}_q(\Gamma_i). \tag{4.7.12}$$

Proof. It simply follows from the additivity (4.7.6). Indeed,

$$C_q(\Gamma^{\otimes}) = \sup_{\kappa\in\mathcal{K}_q^{(n)}}\mathcal{I}_{\mathcal{A}^{(n)},\mathcal{B}}(\kappa^*\Gamma^{\otimes}) = \sup_{\varrho_0^{\otimes}}\mathcal{J}(\varrho_0^{\otimes},\Gamma^{\otimes}) = \sup_{\varrho_0^{\otimes}}\sum_{i=1}^{n}\mathcal{J}(\varrho_0^{i},\Gamma_i) \tag{4.7.13}$$

Therefore by further taking suprema over $\varrho_0^{\otimes}$ as over independently for each $i = 1, 2, ..., n$, thus we have

$$\mathcal{C}_q(\Gamma^{\otimes}) = \sum_{i=1}^{n}\sup_{\varrho_0^{\otimes}}\mathcal{J}(\varrho_0^{i},\Gamma_i) = \sum_{i=1}^{n}\mathcal{C}_q(\Gamma_i), \tag{4.7.14}$$

which is the additivity property of entangled quantum channel capacity due to encodings via entanglement obviously. □

Remark 4.7.6. *Though there are such additivity for our entangled quantum channel capacity in above chapter, but there exists no such additivity for Holevo channel capacity for a arbitrary channel $\Gamma : \mathcal{B} \to \mathcal{B}^0$.*

Indeed, this capacity is defined as the supremum

$$C_d(\Gamma) = \sup_{\kappa\in\mathcal{K}_d}\mathcal{I}_{\mathcal{A},\mathcal{B}}(\kappa^*\Gamma) \tag{4.7.15}$$

over the smaller class $\mathcal{K}_d \subseteq \mathcal{K}_q$ of the diagonal (semiclassical) encodings $\kappa : \mathcal{A} \to \mathcal{B}_*^0$ corresponding to the Abelian algebra $\mathcal{A}$.

This supremum cannot in general be achieved on the standard entanglement of $\mathcal{A}^0 = \widetilde{\mathcal{B}}^0 \equiv \mathcal{B}_0$ if $\mathcal{A}^0$ is non Abelian corresponding to the non Abelian input algebra $\mathcal{B}^0$.

Therefore the supremum $\mathcal{C}_d(\Gamma^{\otimes}) \leq \sum_{i=1}^{n}\mathcal{C}_d(\Gamma_i)$ can be achieved not on a product Abelian algebra $\mathcal{A}^{(n)}$ as is was in the true quantum case where we could take $\mathcal{A}^{(n)} = \otimes_{i=1}^{n}\mathcal{B}_0^i$ with non Abelian $\mathcal{B}_0^i = \widetilde{\mathcal{B}}_i^0$.

CHAPTER 5

Quantum State Identification

5.1 Introduction

In quantum information and quantum computation, it is a necessary to cognise quantum states and operations, which takes the terminology of quantum state estimation in quantum mechanics or quantum statistics. Here, we use the terminology of quantum state identification to distinguish between quantum state estimation in quantum mechanics and quantum statistics, since we investigate identifying state in the frame of game theory.

In quantum statistics, since Helstrom [41] initiated the optimal testing of two quantum statistical hypotheses and optimal estimation of a single unknown parameter of a quantum state, A. S. Holevo [43] studied the optimization problem for quantum measurements and inferences. V. P. Belavkin had systematic explorations, for example, the optimal estimation for several parameters of non-commuting quantum states [4, 5], the quantum Bayesian problem for the linear-Gaussian case by generalized Heisenberg inequality [6, 8]. Also the optimal multiple hypothesis was tested for several non-orthogonal quantum pure states [9–11]. Right Cramer-Rao bound, uncertainty relation for shift parameter estimation, and the efficiency were investigated by applying symmetrical and right quantum Cramer-Rao inequality [12]. And the recursive solutions were obtained for quantum optimal estimation of Markov chains [13], and wave pattern recognition [15].

In quantum computation, a game theoretic perspective was presented in [64] on quantum state estimation, and the legitimacy of universal machines and the different measures of success was explained.

Different from above, in the sense of game theory, this chapter is to apply

Kelly's criterion [52] to identify quantum state, during which a practical game theoretic interpretation is given to abstract classical information theory, especially relative entropy, mutual information, and asymptotic information, upon which curious financial stock may be designed (or suitably "quantum" stock in physical implementation).

Let $\{X \in \mathbb{R}^{m\times m} : X_{ij} \geq 0, i, j = 1, 2, ..., n\}$ denote a set of matrix-valued awards, where X_{ij} is the ratio of the award element at the end of one shot game to the award element at the beginning, and $B = \{b \in \mathbb{C}^{m\times m} : b \geq 0, \mathrm{Tr}b = 1\}$ be the set of all quantum states, where $b \geq 0$ refer to nonnegative matrix b, and $\mathrm{Tr}(.)$ is the trace.

Therefore the resulting utility is taken as

$$S = \mathrm{Tr}[bX], \tag{5.1.1}$$

where we take $S_0 = 1$.

Applying Kelly's criterion [52] to identify quantum state, we maximize the expectation $\mathbb{E}\log(\mathrm{Tr}[bX])$.

Thus, we will find following things.

The decrement in the doubling rate achieved with true knowledge of the distribution F of the awards over that achieved with incorrect knowledge G is bounded above by $\mathcal{R}(F;G)$ (the entropy of F relative to G).

The increment Δ in the doubling rate resulting from side information Y is less than or equal to the mutual information $\mathcal{I}$.

A good sequence of identifications of the true quantum state leads to asymptotically optimal growth rate of utility.

And applying the asymptotic behavior of classical relative entropy (between the n-fold product of a given member of a parameterized family of distributions p_θ and a mixture of products of such distributions M_n) to sequential identification, the utility of the Bayesian strategy is lower bounded in terms of the optimal utility.

The structure of this chapter is organized as follows.

We will introduce basic terminology and results of repeated game in the second section, basic definitions on classical mutual information in the third section, and Kelly criterion in the section four.

In the fifth section, we will consider one spot identification of quantum state, and find the decrement of the doubling rate based on incorrect distributions,

and its information bound, also the increment of the doubling rate with side information and its information bound.

The section six is for sequential identification, a good sequence of identifications of the true quantum state leads to asymptotically optimal growth rate of utility, etc., and applying the asymptotic behavior of classical relative entropy to sequential identification, the utility of the Bayesian strategy is bounded below in terms of the optimal utility.

5.2 Repeated Game

Composed in some number of repetitions of some one-spot game, a repeated game is formed in an extensive form. We usually take the well known 2-person games as one-spot game.

In the repeated game, one person will play the game again with the same person, therefore, different equilibrium properties exist for the real threat of retaliation.

This section introduces repeated game in extensive form. See [81], for example, in detail for reference.

5.2.1 Extensive Form Game

When playing a game, we can exactly specify (1) the physical order of play; (2) the choices available to a player when it is his turn to move; (3) the rules for determining whose move it is at any point; (4) the information a player has once it is his turn to move; (5) the payoffs to the players as functions of the moves they select; and (6) the actions of nature or the initial conditions that begin the game. Similar to Kuhn's extensive game [57], we call this as an extensive game.

Below we follow [54] to define extensive game in mathematics, and discuss its properties.

Definition 5.2.1. *([54]) We call the collection* $\{T, \prec; A, \alpha; I, \iota; H\}$ *an extensive form if consisting of following things (1)-(4).*

(1) The physical order of play is given by a finite set T *of nodes together with a binary relation* $\prec$ *on* T *representing precedence. Precedence is indicated by arrows-one node precedes another if there is a sequence of arrows pointing from the first to the second.*

Remark 5.2.2. *([54]) The binary relation* $\prec$ *is a partial order, and* $(T, \prec)$ *form an*

arborescence: the relation $\prec$ totally orders the predecessors of each member of T, preventing cycles from appearing in the order of play, and each node can be reached by one and only one path from an initial node.

In general, the terminal nodes are described in Z, the decision nodes in X, and the initial nodes in W (possibly more than one element of W).

Remark 5.2.3. *([54]) Beginning at one of the initial nodes (determined by nature), the game proceeds along some path from node to immediate successor, terminating when a terminal node is reached. The various paths imply the various possible orders of play.*

(2) With a finite set A of actions, a function $\alpha : T\backslash W \to A$ labels each non-initial node with the last action taken to reach it, which gives the choices available to players at decision nodes.

Remark 5.2.4. *([54]) $\alpha(S(x))$ is the set of feasible actions at the decision node x with α one-to-one on the set $S(x)$ of immediate successors of x.*

(3) For a finite set I of players, a function $\iota : X \to I$ assigns to each decision node the player whose turn it is, therefore, we can specify the rules for determining whose move it is at a decision node.

(4) We represent information possessed by players by a partition H of X dividing the decision nodes into information sets.

The cell $H(x)$ of H, containing x, is the decision nodes that player $\iota(x)$ cannot distinguish from x upon the information he has available when it is his turn to choose an action at x.

A player knows when it is his turn to choose and which actions are feasible: if $x \in H(x')$, then $\iota(x) = \iota(x')$ and $\alpha(S(x)) = \alpha(S(x'))$, thus write $\iota(h)$ and partition H into sets $H^i = \iota^{-1}(i)$, also write $A(h)$ for $\alpha(S(h))$, the set of actions feasible at information set h.

Assume α is onto, for each $a \in A$, $A^{-1}(a)$ is a singleton set, that is, each action can be taken only in a single information set, thus we can partition A into sets $A' = \{a : A^{-1}(a) \subseteq H'\}$ for $i \in I$.

Assume each player has perfect recall, each player knows whether he chose previously, that is,

$$x \in H(x') \Longrightarrow x \prec x', \tag{5.2.1}$$

also knows whatever he knew previously, including his previous actions, that is, if $x, x', x'' \in \iota^{-1}(i)$, $x \prec x'$, and $H(x') = H(x'')$, therefore, $H(x)$ consists some pre-

decessor of x'' at which the same action was chosen as was chosen at x; thus $P(x'') \cap H(x) = \{x_0\}$, and if $x = p_n(x')$ and $x^0 = p_m(x'')$, then $\alpha(p_{n-1}(x')) = \alpha(p_{m-1}(x''))$.

To obtain an extensive game, we still need further specify the players' utilities assigned to the terminal nodes and the probabilities assigned to the initial nodes [54].

Definition 5.2.5. *([54]) We call the collection $\{T, \prec; A, \alpha; I, \iota; H; u; \rho; \pi\}$ an extensive game if also consisting of following things (5)-(7).*

(5) We assign payoff function $u^i : Z \to \mathbb{R}$, for each player i, and specify the payoffs as $u = (u^i(z)) \in \mathbb{R}^{I \times Z}$.

(6) We specify player i's initial assessment ρ^i as a probability measure on the set W of states or initial nodes. Furthermore, we assume the players' initial assessments strictly positive and all the same: $\rho^i = \rho \gg 0$. Initial assessments are taken by recording the probability $\rho(\omega)$ in braces next to the node ω.

To specify what action player i will take each time it is his turn to choose on the information that he possesses, a pure strategy for player i is taken as an assignment $\sigma^i : H' \to A$ such that $\sigma^i(h) \in A(h)$.

A mixed strategy for player i is therefore defined as a probability distribution over the set of his pure strategies.

(7) We assign a strategy $\pi^i : A^i \to [0, 1]$ for player i to each information set $h \in H^i$ a probability measure on the set $A(h)$. Thus $\sum_{a \in A(h)} \pi^i(a) = 1$, for each $h \in H^i$.

We denote $\prod^i$ the set of strategies for player i, and $\prod = \times_{i \in I} \prod^i$ the set of strategies for the game. Each strategy $\pi \in \prod$ thus induces a probability measure P^π on the set Z of outcomes satisfying the formula

$$P^\pi(z) = \rho(p_{l(z)}(z)) \prod_{l=1}^{l(z)} \pi^{lp_l(z)}(\alpha(p_{l-1}(z))), \tag{5.2.2}$$

where expectation operator $\mathbb{E}^\pi[.]$ is denoted by P^π. In particular, $\mathbb{E}^\pi[u^i(z)]$ is player i's expected utility from the strategy π.

A subform of extensive game can be introduced as follows.

Definition 5.2.6. *([54]) A subform of an extensive form is a collection of nodes $\hat{T} \subseteq T$, together with $\prec, l, A, \alpha$ and H all defined on $\hat{T}$ by restriction, closure under succession and preservation of information sets: $S(x) \subseteq \hat{T}$ and $H(x) \subseteq \hat{T}$ if $x \in \hat{T}$.*

Remark 5.2.7. *([54]) For every node $x \in X$, there exists a minimal subform $\hat{T}(x)$ containing x, where x need not be an initial node in $\hat{T}(x)$.*

In particular, a proper subform is a subform $\hat{T}$ consisting only of some node x and its successors, where we call x the root of $\hat{T}$.

Remark 5.2.8. *([54]) Given a proper subform $\hat{T}$ with root x, there exists a well-defined proper subgame starting with x as the unique initial node, that is, the game is formed by $\hat{T}$ and all the structure that $\hat{T}$ inherits from the original form, the payoffs u restricted to $\hat{T} \cap Z$, and the initial assessment $\hat{\rho}(x) = 1$.*

On the contrary, in nonproper subforms, a subgame is not always well-defined, once the initial assessment $\hat{\rho}$ is lacking.

Principally we can introduce a strategy as Nash Equilibrium [54] (the weakest criterion for equilibrium originally [73]) if each player's strategy is an optimal response to the other players' strategies as follows.

$\pi \in \prod$ is a Nash equilibrium if, for each player $i \in I$, for every strategy $\bar{\pi} \in \prod$, such that $\bar{\pi}^j = \pi^j$ for $j \neq i$, $\mathbb{E}^{\pi}[u^i(z)] \geq \mathbb{E}^{\bar{\pi}}[u^i(z)]$.

A threat exists to this criterion: if players arrive at some "agreed-upon" mode of behavior, then necessarily this behavior constitutes a Nash equilibrium; otherwise, it is advantageous for some player to defect from the agreement.

In fact, in any proper subgame, we can speak of each player's expected utility in that subgame, and thus apply the Nash criterion. And it is a natural restriction for any "agreed-upon" behavior, otherwise the agreement would not hold up if the subgame were reached. Accordingly some players might defect from the agreement and cause the subgame to be reached, anticipating a breakdown of the agreement favorable to himself.

Therefore considering games in extensive form leads to other criterion, more stringent necessary conditions for "agreed-upon" behavior, and the following subgame perfection criterion can make sense [90].

Definition 5.2.9. *([90]) Strategy π is subgame perfect if for every proper subgame the strategy π restricted to the subgame is a Nash equilibrium for the subgame.*

5.2.2 Repeated Game

As an extensive form game, repeated games can be repeated finitely or infinitely many times (so called supergames elsewhere).

The repeated games of a possibly infinite number of times are most widely studied. We usually model the repeated game by applying a discount factor

to each future stage, for two primary considerations: First, at each stage there may be some finite probability that the game ends; Second, each individual may care slightly less about each successive future stage.

The subgame perfection criterion fails sometimes, since there may be not a proper subgame starting from one node. But we can formulate general criterion as follows [54]: A strategy π should be such that for any information set h that is a singleton, player $l(h)$ should not be able to change his strategy unilaterally and thereby improve this expected utility starting from h.

Mathematically we need restrict this criterion to singleton information sets h, such that player $l(h)$'s expected utility starting from h can be calculated. Thus we formulate the criterion of sequential rationality [54]: the strategy of each player starting from each information set must be optimal starting from there according to some assessment over the nodes in the information set and the strategies of everyone else.

To describe an equilibrium, for each information set h, we denote the assessment made by player $l(h)$ over the nodes in h if h is reached.

Definition 5.2.10. *([54]) We define a system of beliefs by a function $\mu : X \to [0,1]$ such that*

$$\sum_{x \in h} \mu(x) = 1 \tag{5.2.3}$$

for each $h \in H$, where $\mu(x)$ as the probability assigned by $l(h)$ to $x \in h$ if h is reached.

Definition 5.2.11. *([54]) An assessment is taken as a pair (μ, π) consisting of a system of beliefs μ and a strategy π.*

Definition 5.2.12. *([54]) Given an assessment (μ, π), for each $h \in H$, "conditional" probability $P^{\mu,\pi}(. \mid h)$ over Z is taken [54] as follows.*

(1) If $z \in Z(h)^c$, then $P^{\mu,\pi}(z \mid h) = 0$;

(2) If $z \in Z(h)$, for example, $p_n(z) \in h$, then $P^{\mu,\pi}(z \mid h) = \mu(p_n(z)) \prod_{m=1}^{n} \pi p_{m-1}(z)$.

Conditional expectations are denoted by $\mathbb{E}^{\mu,\pi}(. \mid h)$, and the assessment is sequentially rational [54] if, for all $h \in H$, for all $\bar{\pi}$ such that, for $j \neq l(h)$,

$$\bar{\pi}^j = \pi^j, \mathbb{E}^{\mu,\pi}(u^{l(h)}(z) \mid h) = \mathbb{E}^{\mu,\bar{\pi}}(u^{l(h)}(z) \mid h). \tag{5.2.4}$$

Remark 5.2.13. *([54]) It is rough to take a sequential equilibrium only as a sequentially rational assessment (μ, π), since we may impose consistency conditions, for example, assessments obey Bayesian rule when it applies, but an assessment will or will not be*

a sequential equilibrium. Therefore, we require an equilibrium specify beliefs as well as strategies.

Let $\prod^0$ be the set of all strictly positive strategies, i.e., $\pi \in \prod^0$ if $\pi(a) > 0$ for all $a \in A$. If $\pi \in \prod^0$, then $P^\pi(x) > 0$ for all x, and we define beliefs μ associated with strategy π by Bayesian rule

$$\mu(x) = \frac{P^\pi(x)}{P^\pi(H(x))}. \tag{5.2.5}$$

Let Ψ^0 be the subset of the set of assessments (μ, π), where $\pi \in \prod^0$ and μ is defined from ρ and π by Bayesian rule.

Definition 5.2.14. *([54]) An assessment (μ, π) is consistent if*

$$(\mu, \pi) = \lim_{n\to\infty}(\mu_n, \pi_n), \tag{5.2.6}$$

for some assessment sequences $\{\mu_n, \pi_n\} \subseteq \prod^0$.

We denote the set of consistent assessments by Ψ. Therefore, a sequential equilibrium is an assessment (μ, π) both consistent and sequentially rational.

Similarly, we can consider the existence of sequential equilibrium for any extensive game and the relation between sequential equilibrium and subgame perfect Nash equilibrium.

Theorem 5.2.15. *([54]) For every extensive game, there is at least one sequential equilibrium.*

Theorem 5.2.16. *([54]) If (μ, π) is a sequential equilibrium, then π is a subgame perfect Nash equilibrium.*

See [54], for example, for proof in detail.

5.3 Kelly Criterion

Initialed since [52], the desirability of maximizing $\mathbb{E}\log X_n$ (Kelly criterion) was investigated in general [29]. This section simply introduces Kelly criterion after coin tossing and minimizing the probability of ruin, among which the advantages of Kelly criterion will be seen. See [94], for example, for systematic introduction.

In repeated independent trials with finitely many outcomes $I = (1, ..., s)$ for each trial, we denote the probability of each outcome by $P(i) = p_i$, $i = 1, ..., s$,

and $\{A_1, ..., A_r, \}$ a collection of betting subsets of I, that each i is in some A_k, and payoff odds o_k corresponding to the A_k.

Amounts $B_1, ..., B_r$ are bet on the respective A_k. And if outcome i occurs, we have the total payoff $\sum_{j:i\in A_j} B_j o_j$. Thus we have this betting in a game model.

5.3.1 Coin Tossing

This section will follow [94] to introduce coin tossing.

An infinitely adversary matches all bets that we make on repeated independent trials of a biased coin, and two outcomes may happen: "heads" and "tails". Finite money is denoted by X_0, B_i is bet on the outcome of the i-th trial, and X_i is denoted our money after the i-th trial with the probability p of heads where $\frac{1}{2} < p < 1$. The problem is thus to decide how much to bet at each trial.

In a classic criterion, we choose B_i to maximize the expected gain $\mathbb{E}(X_i - X_{i-1})$ at each trial, that is, maximize $\mathbb{E}X_n$ for all n.

If the j-th trial results in success, we take $T_j = 1$, otherwise, $T_j = -1$ if the j-th trial results in failure. Thus $X_j = X_{j-1} + T_j B_j$, for $j = 1, 2, ...$, and $X_n = X_0 + \sum_{j=1}^{n} T_j B_j$. If T_j, X_j, and B_j are random variables on a suitable sample space Ω, for example, B_j is a function of $X_0, X_1, ..., X_{j-1}$ in the common gambling systems, then B_j is a function of $X_0, T_1, T_2, ...$ We can take the underlying sample space as the space of all sequences of successes and failures, with the usual product measure.

If $B_j \leq 0$, $-B_j \geq 0$ as a nonnegative bet by a player succeeding when $-T_j = 1$, with probability q. The $-T_j$ are independent, thus we have the trials with success probabilities q. Particularly, we can write the payoff $B_j T_j$ from trial j as $(-B_j)(-T_j)$.

When $B_j > X_{j-1}$, the player is betting more than he has, thus asking for credit. When $B_j < 0$, the player is making a "negative" bet, "backing" the other side of the game, thus to taking the role of the "other" player.

Therefore we can simply suppose $0 \leq B_j \leq X_{j-1}$, and B_j independent of T_j, the amount bet on the j-th outcome is independent of that outcome.

Definition 5.3.1. *([94]) A betting strategy is a family $\{B_j\}$ such that $0 \leq B_j \leq X_{j-1}$, for $j = 1, 2, ...$*

Since $X_n = X_0 + \sum_{j=1}^{n} B_j T_j$, we have expression

$$\mathbb{E}(X_n) = X_0 + \sum_{j=1}^{n} \mathbb{E}(B_j T_j) = X_0 + \sum_{j=1}^{n} (p-q)\mathbb{E}(B_j). \tag{5.3.1}$$

To maximize the expected gain, we have the following result.

Theorem 5.3.2. *([94])*

(1) The strategies $B_j = X_{j-1}$ when $p > \frac{1}{2}$;

(2) $B_j = 0$ when $p < \frac{1}{2}$;

(3) B_j arbitrary when $p = \frac{1}{2}$,

which are precisely maximizing $\mathbb{E}(X_j)$ for each j.

Proof. (1) If $p - q > 0$, that is, $p > \frac{1}{2}$, $\mathbb{E}(B_j)$ should be maximized, when $B_j = X_j$, to maximize $\mathbb{E}(X_n)$;

(2) If $p - q < 0$, that is, $p < \frac{1}{2}$, $\mathbb{E}(B_j)$ should be minimized to maximize the j-th term, that is, $B_j = 0$;

(3) If $p - q = 0$, that is, $p = \frac{1}{2}$, $\mathbb{E}(B_j)$ does not affect $\mathbb{E}(X_n)$. □

Remark 5.3.3. *[94] We must bet total resources at each trial to maximize the expected gain. When losing once, we will ruin, with the probability of this $1 - p^n \to 1$, thus it is undesirable to maximize the expected gain .*

5.3.2 Minimizing the Probability of Ruin

This section will still follow [94] to introduce minimizing the probability of ruin.

Ruin occurs after the j-th outcome if $X_j = 0$. We play minimizing the probability of eventual ruin. Without further restriction on B_j, many strategies minimize the probability of ruin.

Without generality, it suffices to choose $B_j < \frac{X_{j-1}}{2}$ and $0 < C \leq B_j$ with a non-zero constant C. We further require $B_j = C$ if $0 < X_{j-1} < a$ and $B_j = 0$ if $X_{j-1} \leq 0$ or $X_{j-1} \geq a$, and C divides both $a - z$ and z, where $z = X_0$.

In the ruin situation: $X_0 = z$, $B_j = 1$ if $0 < X_{j-l} < a$; and $B_j = 0$ if $X_{j-1} = 0$ or $X_{j-1} = a$, where a and z are integers. Take r a positive number (necessarily rational) such that zr and ar are integers, and $R(r)$ the ruin probability when z and a are replaced by zr and ar respectively. Thus we have the ruin probability [94]

$$R(r) = \frac{\theta^{ar} - \theta^{zr}}{\theta^{ar} - 1}, \tag{5.3.2}$$

where $0 < p \neq \frac{1}{2}$ and $\theta = \frac{q}{p}$.

Remark 5.3.4. *([94]) For $0 < z < a$, $x > 0$. If $0 < \theta < 1$, the function $f(x) = \frac{\theta^{zx} - \theta^{ax}}{1-\theta^{ax}}$ is strictly decreasing as x increases, $x > 0$. While if $\theta > 1$, the function $f(x)$ is strictly increasing as x increases, $x > 0$.*

From above remark, we can obtain the following result.

Theorem 5.3.5. *([94])*

(a) If $0 < p < \frac{1}{2}$, $R(r)$ is a strictly increasing function of r;

(b) If $\frac{1}{2} < p < 1$, $R(r)$ is a strictly decreasing function of r.

Therefore in a favorable game, the chance of ruin is decreased by decreasing stakes. Noticing that if $p > \frac{1}{2}$, that is, $\theta < 1$, $R(r) \to 0$ as $r \to \infty$, and the chance of ruin can be arbitrarily small if stakes sufficiently small. At least in the limited strategies to which it applies, we can minimize ruin probability by a minimum bet on each trial.

Remark 5.3.6. *([94]) Obviously, the strategy, which minimizes ruin probability, also minimizes the expected gain.*

5.3.3 Kelly Criterion

This section will still follow [94] to introduce Kelly criterion.

For Bernoulli trials with $\frac{1}{2} < p < 1$ and $B_j = fX_{j-1}$ with a constant $0 \leq f \leq 1$, we have [94]

$$X_n = X_0(1+f)^{S_n}(1-f)^{F_n}, \tag{5.3.3}$$

where S_n and F_n are the number of successes and failures respectively in n trials.

If $f < 1$, no chance exists for $X_n = 0$, for each $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(X_n \geq \varepsilon) \neq 0. \tag{5.3.4}$$

Thus in the sense of the gambler's ruin problem, ruin cannot occur. Therefore it suffices to assume $0 < f < 1$.

In Bernoulli trials, the min-max criterion in game theory is not appropriate, since for a positive integer B_j for all j, the maximum loss (or ruin) is always possible and all strategies have the same maximum possible loss, thus all are equivalent, therefore, minimizing ruin or of maximizing expectation likewise do not make desirable distinctions between those strategies.

Notice that the quantity

$$\log(\frac{X_n}{X_0})^{\frac{1}{n}} = \frac{S_n}{n}\log(1+f) + \frac{F_n}{n}\log(1-f) \tag{5.3.5}$$

measures the rate of increase per trial, according to Kelly's choice [52], we can maximize the exponential rate of growth [94]

$$\log(\frac{X_n}{X_0})^{\frac{1}{n}} = p\log(1+f) + q\log(1-f) \equiv G(f). \tag{5.3.6}$$

Remark 5.3.7. *([94]) If $\frac{1}{2} < p < 1$, function*

$$G(f) = p\log(1+f) + q\log(1-f) \tag{5.3.7}$$

has a unique maximum at $f^* = p - q$, $0 < f^* < 1$ with

$$G(f^*) = p\log p + q\log q + \log 2 > 0. \tag{5.3.8}$$

There exists a unique fraction $f_c > 0$ with $G(f_c) = 0$, and $f^* < f_c < 1$.

Moreover, $G(f) > 0$ for $0 < f < f_c$; $G(f) < 0$ for $f > f_c$, and $G(f)$ strictly increasing from 0 to $G(f^*)$ on $[0, f^*]$, but $G(f)$ strictly decreasing from $G(f^*)$ to $-\infty$ on $[f^*, 1]$.

Theorem 5.3.8. *([94])*

(a) If $G(f) > 0$, then for each M,

$$\mathbb{P}(\underline{\lim} X_n > M) = 1; \tag{5.3.9}$$

(b) If $G(f) < 0$, then for each $\varepsilon > 0$,

$$\mathbb{P}(\overline{\lim} X_n < \varepsilon) = 1; \tag{5.3.10}$$

(c) If $G(f) = 0$, then

$$\mathbb{P}(\underline{\lim} X_n > M) = 1, \tag{5.3.11}$$

and

$$\mathbb{P}(\overline{\lim} X_n < \varepsilon) = 1. \tag{5.3.12}$$

Thus (1) for $0 < f < f_c$, the player's utility will eventually permanently exceed any fixed bounds with probability one;

(2) For $f = f_c$, it will almost surely oscillate wildly between 0 and $+\infty$;

(3) If $f > f_c$, ruin occurs almost surely.

Proof. (a) According to the strong law of large numbers,

$$\lim \log(\frac{X_n}{X_0})^{\frac{1}{n}} = G(f) > 0 \tag{5.3.13}$$

with probability 1.

Therefore, almost surely, for $\omega \in \Omega$, where Ω is the space of all sequences of Bernoulli trials, there is $N(\omega)$ such that for $n > N(\omega)$,

$$\log(\frac{X_n}{X_0})^{\frac{1}{n}} \geq \frac{G(f)}{2} > 0. \tag{5.3.14}$$

Thus $X_n \geq X_0 e^{\frac{nG(f)}{2}}$, then $X_n \to \infty$.

(b) According to the strong law of large numbers,

$$\lim \log(\frac{X_n}{X_0})^{\frac{1}{n}} = G(f) < 0 \tag{5.3.15}$$

with probability 1.

Therefore, almost surely, for $\omega \in \Omega$, where Ω is the space of all sequences of Bernoulli trials, there exists $N(\omega)$ such that for $n > N(\omega)$,

$$\log(\frac{X_n}{X_0})^{\frac{1}{n}} \geq \frac{G(f)}{2} < 0. \tag{5.3.16}$$

Thus $X_n \leq X_0 e^{\frac{nG(f)}{2}}$, then $X_n \to 0$.

(c) Given any M,

$$\overline{\lim} S_n \geq np + M + 1, \tag{5.3.17}$$

and

$$\underline{\lim} S_n \geq np - M - 1. \tag{5.3.18}$$

Then (1) If $S_n \geq np + M$,

$$\lim \log(\frac{X_n}{X_0})^{\frac{1}{n}} \geq \frac{np+M}{n} \log(1+f) + \frac{n-(np+M)}{n} \log(1-f) \tag{5.3.19}$$

$$= G(f) + \frac{M}{n} \log \frac{1+f}{1-f} = \frac{M}{n} \log \frac{1+f}{1-f}, \tag{5.3.20}$$

thus $X_n \geq X_0 (\frac{1+f}{1-f})^M$.

Since $S_n \geq np + M$ almost surely, $\overline{\lim} X_n \geq X_0 (\frac{1+f}{1-f})^M$ almost surely.

Thus $\overline{\lim} X_n = \infty$ almost surely, for we can choose $X_0 (\frac{1+f}{1-f})^M$ arbitrarily large.

(2) If $S_n \leq np - M$,

$$\lim \log(\frac{X_n}{X_0})^{\frac{1}{n}} \leq \frac{np-M}{n} \log(1+f) + \frac{n-(np-M)}{n} \log(1-f) \tag{5.3.21}$$

$$= G(f) + \frac{M}{n}\log\frac{1-f}{1+f} = \frac{M}{n}\log\frac{1-f}{1+f}, \tag{5.3.22}$$

thus $X_n \leq X_0(\frac{1-f}{1+f})^M$.

Since $S_n \leq np - M$ almost surely, $\underline{\lim} X_n \leq X_0(\frac{1-f}{1+f})^M$ almost surely.

Thus $\underline{\lim} X_n = 0$ almost surely, for we can choose $X_0(\frac{1-f}{1+f})^M$ arbitrarily small.

□

Theorem 5.3.9. *([94]) If $G(f_1) > G(f_2)$. Then $\lim \frac{X_n(f_1)}{X_n(f_2)} = \infty$ almost surely.*

Proof. From above, we have the formula

$$\log[\frac{X_n(f_1)}{X_0}]^{\frac{1}{n}} - \log[\frac{X_n(f_2)}{X_0}]^{\frac{1}{n}} = \log[\frac{X_n(f_1)}{X_n(f_2)}]^{\frac{1}{n}} \tag{5.3.23}$$

$$= \frac{S_n}{n}\log\frac{1+f_1}{1+f_2} + \frac{F_n}{n}\log\frac{1-f_1}{1-f_2}. \tag{5.3.24}$$

According to the strong law of large numbers,

$$\lim \log[\frac{X_n(f_1)}{X_n(f_2)}]^{\frac{1}{n}} \to G(f_1) - G(f_2) > 0 \tag{5.3.25}$$

with probability 1.

Therefore, almost surely, for $\omega \in \Omega$, where Ω is the space of all sequences of Bernoulli trials, there exists $N(\omega)$ such that for $n > N(\omega)$,

$$\log[\frac{X_n(f_1)}{X_n(f_2)}]^{\frac{1}{n}} \to \frac{G(f_1) - G(f_2)}{2} > 0. \tag{5.3.26}$$

Then

$$\frac{X_n(f_1)}{X_n(f_2)} \geq \frac{X_0(f_1)}{X_0(f_2)} e^{\frac{G(f_1)-G(f_2)}{2}}. \tag{5.3.27}$$

Thus

$$\frac{X_n(f_1)}{X_n(f_2)} \to \infty. \tag{5.3.28}$$

□

If one player applies f^* and another applies any other fraction strategy f. Then we have the following proposition.

Proposition 5.3.10. *([94])*

$$\lim \frac{X_n(f^*)}{X_n(f)} = \infty \tag{5.3.29}$$

with probability 1, which is the criterion to maximize $\mathbb{E}\log X_n$.

Moreover, f^* not only maximizes $\mathbb{E}\log X_n$ within the class of all fraction betting strategies, but also in the class of "all" betting strategies [29].

For a series of independent trials, the gain on one unit bet on the i-th outcome is taken as the random variable Q_i. Then

$$X_n = \prod_{i=1}^{n} \frac{X_i}{X_{i-1}}, \tag{5.3.30}$$

$$\mathbb{E}(\log X_n) = \sum_{i=1}^{n} \mathbb{E}\log \frac{X_i}{X_{i-1}}. \tag{5.3.31}$$

Thus, $X_i = X_{i-1} + B_i Q_i$ and $\frac{X_i}{X_{i-1}} = 1 + \frac{B_i}{X_{i-1}} Q_i$.

Each term is written as $\mathbb{E}\log(1 + F_i Q_i)$, where the random variable F_i depends on the first $i-1$ trials, and independent random variable Q_i on the i-th trial. Thus subject to the constraint $0 \leq F_i \leq 1$, it is free to choose the F_i to maximize $\mathbb{E}\log(1 + F_i Q_i)$.

Theorem 5.3.11. *([94]) For each i, there exists f_i with $0 < f_i < 1$ for well-defined positive $\mathbb{E}\log(1 + f_i Q_i)$ (where $Q_i \neq 0$ with probability 1 for each i). Then, for each i, there exists a number f_i^* such that $\mathbb{E}\log(1 + F_i Q_i)$ reaches its unique maximum at $F_i = f_i^*$ with probability 1.*

Proof. Obviously, let $a_i = \min(1, b_i)$ and $b_i = \sup\{f_i : \mathbb{P}(f_i Q_i > 0) = 1\} > 0$, an interval $[0, a_i)$ or $[0, a_i]$ is the domain of definition of $\mathbb{E}\log(1 + f_i Q_i)$.

Any maximum of $\mathbb{E}\log(1 + f_i Q_i)$ is unique, since $-\mathbb{E}\frac{Q_i^2}{(1+f_i Q_i)^2}$, the second derivative with respect to f_i of $\mathbb{E}\log(1 + f_i Q_i)$, is negative.

If the continuous function $\mathbb{E}\log(1 + f_i Q_i)$ (on its domain) is defined at a_i, there is a maximum.

If it is not defined at a_i, $\lim_{f_i \to a_i} \mathbb{E}\log(1 + f_i Q_i) = -\infty$, and there still exists a maximum.

Let $F_i(s_1)$ and $Q_i(s_2)$ be functions on a product measure space $S_1 \times S_2$, for the independence of F_i and Q_i. Then

$$\mathbb{E}\log(1 + F_i Q_i) = \int_{S_1}\int_{S_2} \log(1 + F_i(s_1) Q_i(s_2)) = \mathbb{E}\mathbb{E}\log(1 + F_i(s_1) Q_i) \tag{5.3.32}$$

$$\leq \mathbb{E}\log(1 + f_i^* Q_i), \tag{5.3.33}$$

where equality holds if and only if

$$\mathbb{E}\log(1 + F_i(s_1) Q_i) = \mathbb{E}\log(1 + f_i^* Q_i) \tag{5.3.34}$$

with probability 1, that is, $f_i^* Q_i = F_i(s_1) Q_i$ with probability 1, thus almost surely either $f^* = F_i$ or $Q_i = 0$ for their independence, therefore, $f_i^* = F_i$ with probability 1. □

Remark 5.3.12. *([94]) For Bernoulli trials with success probability p_i on the i-th trial and $\frac{1}{2} < p < 1$, we can maximize $\mathbb{E} \log X_n$ by choosing $f_i^* = p_i - q_i$, which in fact maximizes $\mathbb{E} \log(1 + f_i Q_i)$.*

5.4 Classical Mutual Information

This section introduces classical mutual information, just for its basic definition. See [79, 91], for example, for reference.

Let X, Y be two random variables with joint distribution P_{XY}. In general, the relative entropy between the conditional distribution $P_{X|Y}$ and the marginal distribution P_X is defined as mutual information $\mathcal{I}(X;Y)$ of X and Y.

Definition 5.4.1. *(**Mutual Information $\mathcal{I}(X;Y)$ of random variables X and Y**, [79])*

$$\mathcal{I}(X;Y) \equiv \int \mathcal{R}(P_{X|Y=y}; P_X) P_Y(dy) = \mathcal{R}(P_{XY}; P_X P_Y), \tag{5.4.1}$$

where P_X, P_Y are the marginal distributions of joint distribution P_{XY} for two random variables X, Y.

Many interpretations exist for mutual information. For example, if we interpret the relative entropy $\mathcal{R}(F;G)$ as the error exponent for the hypothesis test F versus G, $\mathcal{I}(X;Y)$ is the error exponent for the hypothesis test (X,Y) independent versus (X,Y) dependent.

Let $H(X)$ be the Shannon entropy, and $H(X|Y)$ the conditional entropy, we can write the mutual information $\mathcal{I}(X;Y)$ as follows.

$$\mathcal{I}(X;Y) = H(X) - H(X|Y). \tag{5.4.2}$$

Therefore, $\mathcal{I}$ is the amount that the entropy of X is decreased by the knowledge of Y.

5.5 One Spot Identification

Let us denote $F(X)$ the probability distribution function of the awards X, to consider efficiency, we define the doubling rate $W(X)$ by the maximum of the

expectation $\mathbb{E}\log(\mathrm{Tr}[bX])$.

Definition 5.5.1. *(The doubling rate $W(X)$ for the awards X with the probability distribution function $F(X)$)*

$$W(X) \equiv W(F) \equiv \max_{b\in B} \int \log(\mathrm{Tr}[bX])dF(X), \tag{5.5.1}$$

where all logarithms are to the base 2.

Let $b^* = b^*(F)$ achieve $W(F)$ for the convexity of B. Then

$$W(X) \equiv W(F) \equiv \int \log(\mathrm{Tr}[b^*X])dF(X). \tag{5.5.2}$$

We generalize the Kuhn-Tuker Conditions [55, 56, 58, 77, 81] to our situation characterizing $b^*(F)$. We have the following result.

Theorem 5.5.2. *Necessary and sufficient conditions for b to maximize the expectation $\mathbb{E}\log\mathrm{Tr}[bX]$ are*

$$\mathbb{E}\frac{X}{\mathrm{Tr}[bX]} = \mathrm{I}. \tag{5.5.3}$$

Proof. Necessary and sufficient conditions for b to maximize the expectation $\mathbb{E}\log\mathrm{Tr}[bX]$ are

$$\max_b \mathbb{E}\log\mathrm{Tr}[bX] \tag{5.5.4}$$

subject to

$$\mathrm{Tr}b = 1, \tag{5.5.5}$$

which are

$$L'_b = 0, \tag{5.5.6}$$

and

$$L'_\lambda = 0, \tag{5.5.7}$$

where

$$L(b,\lambda) = \mathbb{E}\log\mathrm{Tr}[bX] - \lambda\mathrm{Tr}b. \tag{5.5.8}$$

After some calculation, $L'_b = 0$, $L'_\lambda = 0$, and $\mathrm{Tr}b = 1$ are equivalent to

$$\mathbb{E}\frac{X}{\mathrm{Tr}[bX]} = \mathrm{I}. \tag{5.5.9}$$

Therefore necessary and sufficient conditions for b to maximize the expectation $\mathbb{E}\log\mathrm{Tr}[bX]$ are

$$\mathbb{E}\frac{X}{\mathrm{Tr}[bX]} = \mathrm{I}. \tag{5.5.10}$$

□

In one spot identification of quantum state, we may mistake the independent and identical distribution of $\{X\}$, or identify quantum state with available side information Y, for example, world events, the behavior of a correlated state identification, or the past information on previous outcomes X.

For those reasons, the following first subsection is for the incorrect distribution, and the second subsection for state identification upon side information.

5.5.1 Decrement with Incorrect Distributions

Suppose that $\{X\}$ are independent and identically distributed according to G, while in fact $\{X\}$ are independent and identically distributed according to F.

Therefore instead of correct $b^*(F)$, the incorrect $b^*(G)$ is used in the doubling rate. We write the doubling rate $W(b^*(F), F)$ associated with distribution F as follows.

$$W(b^*(F), F) = \int \log(\mathrm{Tr}[b^*(F)X])dF(X), \tag{5.5.11}$$

and we write the doubling rate $W(b^*(G), F)$ associated with incorrect distribution G by

$$W(b^*(G), F) = \int \log(\mathrm{Tr}[b^*(G)X])dF(X). \tag{5.5.12}$$

And further we define decrement $\Delta W(F, G)$ of the doubling rate in applying $b^*(G)$ as follows.

$$\Delta W(F, G) \equiv W(b^*(F), F) - W(b^*(G), F), \tag{5.5.13}$$

Thus we have the following result.

Theorem 5.5.3.

$$0 \leq \Delta W(F, G) \leq \mathcal{R}(F; G). \tag{5.5.14}$$

Proof. On the one hand, $b^*(F)$ is the optimal for the distribution F, according to the definition of decrement Δ, therefore, $\Delta W(F, G) \geq 0$.

On the other hand, if $\mathcal{R}(F; G) = \infty$, trivially $\Delta W(F, G) \leq \mathcal{R}(F; G)$; otherwise, whence $F << G$, $\mathcal{R}(F; G)$ is finite.

Let $S_1^* = \mathrm{Tr}[b^*(F)X]$ and $S_2 = \mathrm{Tr}[b^*(G)X]$ be the utility factors corresponding to the optimal utilities with respect to F and G.

According to our generalized Kuhn-Tucker conditions, the utility factor S_2 is strictly positive with probability one with respect to G, hence strictly positive with probability one with respect to F since $F << G$.

Let F and G have densities f and g with respect to some dominating measure.

Since $F << G$, the set $A = \{S_2 > 0, f > 0, g > 0\}$ has probability one with respect to F.

Then we have the following relations

$$\Delta W(F,G) = \int_A \log \frac{S_1^*}{S_2} dF \tag{5.5.15}$$

$$= \int_A \log(\frac{S_1^*}{S_2}\frac{g}{f}\frac{f}{g})dF \tag{5.5.16}$$

$$= \int_A \log(\frac{S_1^*}{S_2}\frac{g}{f})dF + \mathcal{R}(F;G), \tag{5.5.17}$$

$$\leq \log \int_A \frac{S_1^*}{S_2} dG + \mathcal{R}(F;G) \tag{5.5.18}$$

by the concavity of the logarithm, i.e., $\int_A \log(\frac{S_1^*}{S_2}\frac{g}{f})dF \leq \log \int_A \frac{S_1^*}{S_2} dG$,

$$\leq \mathcal{R}(F;G) \tag{5.5.19}$$

by our generalized Kuhn-Tucker conditions for the optimality of $b^*(G)$ for the distribution G. □

5.5.2 Increment with Side Information

Upon available side information Y (say, world events, the behavior of a correlated state identification, or the past information on previous outcomes X), the utility is taken as $\mathbb{E}\mathrm{Tr}[b(Y)X]$, where the b depends on Y, and we take the doubling rate $W(X;Y)$ for side information Y as follows.

$$W(X;Y) \equiv \max_{b(Y)} \int\int \log(\mathrm{Tr}[b(Y)X])dF(X,Y), \tag{5.5.20}$$

where $F(X,Y)$ is the joint distribution of the awards X and Y.

Theorem 5.5.4. *If $b^*(Y) = b^*(F_{X|Y})$ achieves $W(X;Y)$. Then $b^*(Y)$ maximizes the conditional expected logarithm of the utility*

$$\mathbb{E}[\log \mathrm{Tr}[b^*X]|Y]. \tag{5.5.21}$$

Proof. Since $F(X,Y)$ is the joint distribution of the awards X and Y, we have

$$F(X,Y) = F(X|Y)F(Y). \tag{5.5.22}$$

Therefore

$$W(X;Y) \equiv \max_{b(Y)} \int\int \log(\mathrm{Tr}[b(Y)X])dF(X,Y) \tag{5.5.23}$$

$$= \max_{b(Y)} \int dF(Y) \int \log(\mathrm{Tr}[b(Y)X]) dF(X|Y). \tag{5.5.24}$$

Since $b^*(Y) = b^*(F_{X|Y})$ achieves $W(X;Y)$, thus $b^*(Y)$ maximizes the conditional expected logarithm of the utility

$$\mathbb{E}[\log \mathrm{Tr}[b^* X]|Y]. \tag{5.5.25}$$

□

We define the difference between the maximum expected logarithm of utility with side information Y and without side information Y as follows.

$$\Delta \equiv W(X;Y) - W(X). \tag{5.5.26}$$

Therefore, in the state identification, we have the following formulas for Δ and $\mathcal{I}$.

$$\Delta = \mathbb{E} \log \frac{\mathrm{Tr}[b^*(Y)X]}{\mathrm{Tr}[b^* X]}, \tag{5.5.27}$$

which involves utility and depends on the values X takes on.

$$\mathcal{I} = \mathbb{E} \log \frac{f(X,Y)}{f(X)f(Y)}, \tag{5.5.28}$$

where $\{(X,Y)\}$ are independent and identically distributed according to $F(X,Y)$, which involves information and depends on X and Y only through the density $f(X,Y)$.

It will be shown that the increment Δ in the doubling rate resulting from side information Y is less than or equal to the mutual information $\mathcal{I}$.

Theorem 5.5.5.

$$0 \leq \Delta \leq \mathcal{I}(X;Y). \tag{5.5.29}$$

Proof. For any side information Y, let $P_{X|Y}$ be the conditional distribution for X given Y, P_X the marginal distribution for X, and $b^{**} = b^*(P_{X|Y})$.

Applying Theorem 5.5.3, respectively replacing F and G by $P_{X|Y}$ and P_X, that is, at first, $b^*(P_{X|Y})$ is the optimal for the distribution $P_{X|Y}$, therefore,

$$\mathbb{E}[\log \frac{\mathrm{Tr}[b^{**}X]}{\mathrm{Tr}[b^* X]}|Y] \geq 0. \tag{5.5.30}$$

Secondly, assume that $P_{X|Y} << P_X$, $\mathcal{R}(P_{X|Y}; P_X)$ is finite, let $S_1^* = \mathrm{Tr}[b^*(P_{X|Y})X]$ and $S_2 = \mathrm{Tr}[b^*(P_X)X]$ be the utility factors corresponding to the optimal utilities with respect to $P_{X|Y}$ and P_X.

According to our generalized Kuhn-Tucker conditions, the utility factor S_2 is strictly positive with probability one with respect to G, hence strictly positive with probability one with respect to $P_{X|Y}$ since $P_{X|Y} << P_X$.

Let $P_{X|Y}$ and P_X have densities f and g with respect to some dominating measure. Since $P_{X|Y} << P_X$, the set $A = \{S_2 > 0, f > 0, g > 0\}$ has probability one with respect to $P_{X|Y}$.

Then we have the following relations

$$\mathbb{E}[\log \frac{\mathrm{Tr}[b^{**}X]}{\mathrm{Tr}[b^{*}X]}|Y] = \int_A \log(\frac{S_1^*}{S_2})dF \tag{5.5.31}$$

$$= \int_A \log(\frac{S_1^*}{S_2}\frac{g}{f}\frac{f}{g})dP_{X|Y} \tag{5.5.32}$$

$$= \int_A \log(\frac{S_1^*}{S_2}\frac{g}{f})dP_{X|Y} + \mathcal{R}(P_{X|Y}; P_X) \tag{5.5.33}$$

$$\leq \log \int_A \frac{S_1^*}{S_2}dP_X + \mathcal{R}(P_{X|Y}; P_X) \tag{5.5.34}$$

by the concavity of the logarithm, i.e., $\int_A \log(\frac{S_1^*}{S_2}\frac{g}{f})dP_{X|Y} \leq \log \int_A \frac{S_1^*}{S_2}dP_X$,

$$\leq \mathcal{R}(P_{X|Y}; P_X) \tag{5.5.35}$$

by our generalized Kuhn-Tucker conditions for the optimality of $b^*(P_X)$ for the distribution P_X.

Therefore we obtain the result

$$0 \leq \mathbb{E}[\log \frac{\mathrm{Tr}[b^{**}X]}{\mathrm{Tr}[b^{*}X]}|Y] \leq \mathcal{R}(P_{X|Y}; P_X). \tag{5.5.36}$$

Averaging with respect to the distribution of Y, we obtain the following result

$$0 \leq \Delta \leq \mathcal{I}(X; Y). \tag{5.5.37}$$

Thus Δ is the increment in doubling rate due to the side information Y. □

Remark 5.5.6. *If $\Delta = 1$, then the side information Y yields an additional doubling of the utility in each game period, in which sense, we call W the doubling rate.*

5.6 Sequential Identification

Against award sets $\{X_1\}, \{X_2\}, ..., \{X_n\}$, we can consider sequential identification of quantum state, but there are several cases according to the distribution of $\{X_1\}, \{X_2\}, ..., \{X_n\}$:

(1) $\{X_1\}, \{X_2\}, ..., \{X_n\}$ are independent and identically distributed;

(2) $\{X_1\}, \{X_2\}, ..., \{X_n\}$ are joint probability distributed;

(3) $\{X_1\}, \{X_2\}, ..., \{X_n\}$ are parameterized distributed.

The following subsections respectively introduce those three cases.

5.6.1 Independent and Identical Distribution

Against the award set $\{X_1\}, \{X_2\}, ..., \{X_n\}$ with independent and identical distribution $F(X)$, current utility can be reallocated according to b^* without side information.

Then, instead of one-spot identification, since we take $S_0^* = 1$, we take the utility S_n^* at time n by

$$S_n^* = \prod_{i=1}^{n} \mathrm{Tr}[b^* X_i]. \tag{5.6.1}$$

According to the strong law of large numbers, we have the following property.

Lemma 5.6.1.

$$(S_n^*)^{1/n} = 2^{1/n \sum_{i=1}^{n} \log \mathrm{Tr}[b^* X_i]} \longrightarrow 2^{W(X)} \tag{5.6.2}$$

with probability one.

Proof. Following the definition of $S_n^* = \prod_{i=1}^{n} \mathrm{Tr}[b^* X_i]$, we obtain

$$(S_n^*)^{1/n} = 2^{1/n \sum_{i=1}^{n} \log \mathrm{Tr}[b^* X_i]}. \tag{5.6.3}$$

Applying the strong law of large numbers, and the definition of $W(X)$, we obtain

$$1/n \sum_{i=1}^{n} \log \mathrm{Tr}[b^* X_i] \longrightarrow W(X) \tag{5.6.4}$$

with probability one.

Combining above two formulas, we have the result

$$(S_n^*)^{1/n} = 2^{1/n \sum_{i=1}^{n} \log \mathrm{Tr}[b^* X_i]} \longrightarrow 2^{W(X)} \tag{5.6.5}$$

with probability one. □

Furthermore, it may be conjectured that no other will achieve a higher exponent.

In the sequential identification against $\{X_1\}, \{X_2\}, ..., \{X_n\}$, where $\{(X_i, Y_i)\}$ are distributed independently and identically according to $F(X, Y)$.

Let $b^*(Y_i)$ be used at time i given side information Y_i. Then we take the utility S_n^{**} at time n by

$$S_n^{**} = \prod_{i=1}^{n} \mathrm{Tr}[b^*(Y_i)X_i]. \tag{5.6.6}$$

According to the strong law of large numbers, it also follows the property.

Lemma 5.6.2.

$$(S_n^{**})^{1/n} = 2^{1/n\sum_{i=1}^{n} \log \mathrm{Tr}[b^*(Y_i)X_i]} \longrightarrow 2^{W(X;Y)} \tag{5.6.7}$$

with probability one.

Proof. Following the definition of $S_n^{**} = \prod_{i=1}^{n} \mathrm{Tr}[b^*(Y_i)X_i]$, we obtain

$$(S_n^{**})^{1/n} = 2^{1/n\sum_{i=1}^{n} \log \mathrm{Tr}[b^*(Y_i)X_i]}. \tag{5.6.8}$$

Applying the strong law of large numbers, and the definition of $W(X;Y)$, we obtain

$$1/n \sum_{i=1}^{n} \log \mathrm{Tr}[b^*(Y_i)X_i] \longrightarrow W(X;Y) \tag{5.6.9}$$

with probability one.

Combining above two formulas, we have the result

$$(S_n^{**})^{1/n} = 2^{1/n\sum_{i=1}^{n} \log \mathrm{Tr}[b^*(Y_i)X_i]} \longrightarrow 2^{W(X;Y)} \tag{5.6.10}$$

with probability one. □

Therefore the ratio of the utility with side information Y to utility without side information Y has the limit $2^{W(X;Y)-W(X)}$.

Theorem 5.6.3.

$$(\frac{S_n^{**}}{S_n^{*}})^{1/n} \longrightarrow 2^{W(X;Y)-W(X)} \equiv 2^{\Delta} \tag{5.6.11}$$

with probability one, where Δ is the difference between the maximum expected logarithm of the utility with side information Y and without side information Y, that is,

$$\Delta \equiv W(X;Y) - W(X). \tag{5.6.12}$$

Proof. Applying (5.6.2) and (5.6.7), we can obtain

$$(\frac{S_n^{**}}{S_n^{*}})^{1/n} = 2^{1/n\sum_{i=1}^{n}(\log \mathrm{Tr}[b^*(Y_i)X_i] - \log \mathrm{Tr}[b^*X_i])} \tag{5.6.13}$$

$$\longrightarrow 2^{W(X;Y)-W(X)} \tag{5.6.14}$$

by the strong law of large numbers, and the definitions of $W(X;Y)$ and $W(X)$,

$$\equiv 2^{\Delta} \tag{5.6.15}$$

by the definition of Δ, the difference between the maximum expected logarithm of the utility with side information Y and without side information Y. □

5.6.2 Joint Probability Distribution

Generalizing Theorem 5.5.3 to sequential identification, we will find that a good sequence of identifications of the true quantum state leads to asymptotically optimal growth rate of utility.

Suppose $\{X_1\}, \{X_2\}, ..., \{X_n\}$ a sequence of random award sets with joint probability distribution P^n. Then the log-optimal sequential strategy employs the $b_i^* = b^*(P_{X_i|X_1,...,X_{i-1}})$, maximizing the conditional expected value of $\log \mathrm{Tr}[bX]$ given $X_1, ..., X_{i-1}$.

And instead of b^*, $\hat{b}_i = b^*(Q_{X_i|X_1,...,X_{i-1}})$ are optimal for an incorrect distribution Q^n for the sequence $X_1, X_2, ..., X_n$. Similarly, we can obtain the following result.

Theorem 5.6.4. *Let $P_{X_i|X_1,...,X_{i-1}}$ and $Q_{X_i|X_1,...,X_{i-1}}$ be the regular conditional distributions associated respectively with joint probability distributions P^n and Q^n.*

Comparing the resulting utility

$$\hat{S}_n = \prod_{i=1}^{n} \mathrm{Tr}[\hat{b}_i X_i] \tag{5.6.16}$$

with the utility

$$S_n^* = \prod_{i=1}^{n} \mathrm{Tr}[b_i^* X_i]. \tag{5.6.17}$$

Then

$$0 \leq \mathbb{E} \log \frac{S_n^*}{\hat{S}_n} \leq \mathcal{R}(P^n; Q^n), \tag{5.6.18}$$

where $\mathcal{R}(P^n; Q^n)$ is the entropy of P^n relative to Q^n.

Proof. Applying Theorem 5.5.3 to a sequential identification, respectively replacing F and G by $P_{X_i|X_1,...,X_{i-1}}$ and $Q_{X_i|X_1,...,X_{i-1}}$, i.e., $b^*(P_{X_i|X_1,...,X_{i-1}})X$ is the optimal for the distribution $P_{X_i|X_1,...,X_{i-1}}$, then

$$\mathbb{E}[\log \frac{\mathrm{Tr}[b^*(P_{X_i|X_1,...,X_{i-1}})X]}{\mathrm{Tr}[b^*(Q_{X_i|X_1,...,X_{i-1}})X]}|Y] \geq 0. \tag{5.6.19}$$

Since $b^*(Q_{X_i|X_1,...,X_{i-1}})$ is the optimal for the distribution $Q_{X_i|X_1,...,X_{i-1}}$, we will see

$$\mathbb{E}[\log \frac{\mathrm{Tr}[b^*(P_{X_i|X_1,...,X_{i-1}})X]}{\mathrm{Tr}[b^*(Q_{X_i|X_1,...,X_{i-1}})X]}|Y] \leq \mathcal{R}(P_{X_i|X^{i-1}}; Q_{X_i|X^{i-1}}). \tag{5.6.20}$$

Assume that $P_{X_i|X_1,...,X_{i-1}} << Q_{X_i|X_1,...,X_{i-1}}$, $\mathcal{R}(P_{X_i|X^{i-1}}; Q_{X_i|X^{i-1}})$ is finite.

Let $S_1^* = \mathrm{Tr}[b^*(P_{X_i|X_1,...,X_{i-1}})X]$ and $S_2 = \mathrm{Tr}[b^*(Q_{X_i|X_1,...,X_{i-1}})X]$ be the utility factors corresponding to the optimal utilities with respect to $P_{X_i|X_1,...,X_{i-1}}$ and $Q_{X_i|X_1,...,X_{i-1}}$.

According to our generalized Kuhn-Tucker conditions, the utility factor S_2 is strictly positive with probability one with respect to $Q_{X_i|X_1,...,X_{i-1}}$, hence strictly positive with probability one with respect to $P_{X_i|X_1,...,X_{i-1}}$ since $P_{X_i|X_1,...,X_{i-1}} << Q_{X_i|X_1,...,X_{i-1}}$.

Let $P_{X_i|X^{i-1}}$ and $Q_{X_i|X^{i-1}}$ have densities f and g with respect to some dominating measure. Since $P_{X_i|X^{i-1}} << Q_{X_i|X^{i-1}}$, the set $A = \{S_2 > 0, f > 0, g > 0\}$ has probability one with respect to $P_{X_i|X^{i-1}}$. Then, we have the following relations

$$\mathbb{E}[\log \frac{\mathrm{Tr}[b^*(P_{X_i|X_1,...,X_{i-1}})X]}{\mathrm{Tr}[b^*(Q_{X_i|X_1,...,X_{i-1}})X]}|Y] = \int_A \log(\frac{S_1^*}{S_2}) dP_{X_i|X^{i-1}} \tag{5.6.21}$$

$$= \int_A \log(\frac{S_1^*}{S_2}\frac{g}{f}\frac{f}{g}) dP_{X_i|X^{i-1}} \tag{5.6.22}$$

$$= \int_A \log(\frac{S_1^*}{S_2}\frac{g}{f}) dP_{X_i|X^{i-1}} + \mathcal{R}(P_{X_i|X^{i-1}}; Q_{X_i|X^{i-1}}) \tag{5.6.23}$$

$$\leq \log \int_A \frac{S_1^*}{S_2} dQ_{X_i|X^{i-1}} + \mathcal{R}(P_{X_i|X^{i-1}}; Q_{X_i|X^{i-1}}) \tag{5.6.24}$$

by the concavity of the logarithm, i.e., $\int_A \log(\frac{S_1^*}{S_2}\frac{g}{f}) dP_{X_i|X^{i-1}} \leq \log \int_A \frac{S_1^*}{S_2} dQ_{X_i|X^{i-1}}$,

$$\leq \mathcal{R}(P_{X_i|X^{i-1}}; Q_{X_i|X^{i-1}}) \tag{5.6.25}$$

by our generalized Kuhn-Tucker conditions for the optimality of $b^*(Q_{X_i|X^{i-1}})$ for the distribution $Q_{X_i|X^{i-1}}$.

Therefore, we result that

$$0 \leq \mathbb{E}[\log \frac{\mathrm{Tr}[b_i^* X_i]}{\mathrm{Tr}[\hat{b}_i X_i]}|X_1, ..., X_{i-1}] \leq \mathcal{R}(P_{X_i|X^{i-1}}; Q_{X_i|X^{i-1}}). \tag{5.6.26}$$

Averaging with respect to the distribution of $X^{i-1} \equiv (X_1, ..., X_{i-1})$, and summing from $i = 1$ to $i = n$, we yields the following result

$$0 \leq \sum_{i=1}^{n} \mathbb{E} \log \frac{\mathrm{Tr}[b^* X_i]}{\mathrm{Tr}[\hat{b}_i X_i]} \tag{5.6.27}$$

$$\leq \mathbb{E}\mathcal{R}(P_{X_i|X^{i-1}}; Q_{X_i|X^{i-1}}) \tag{5.6.28}$$

by chain rule,

$$\equiv \mathcal{R}(P^n; Q^n) \tag{5.6.29}$$

according to the definition of relative entropy. □

Theorem 5.6.5. *Let award sets $\{X_1\}, \{X_2\}, ..., \{X_n\}$ be independently and identically distributed according to P, and $\hat{P}_n$ a sequential identification of the true distribution P such that*

$$\mathbb{E}\mathcal{R}(P; \hat{P}_i) \longrightarrow 0. \tag{5.6.30}$$

We take

$$\hat{S}_n = \prod_{i=1}^{n} \mathrm{Tr}[\hat{b}_i X_i], \tag{5.6.31}$$

where $\hat{b}_i = b^*(\hat{P}_i)$. And suppose that

$$S_n^* = \prod_{i=1}^{n} \mathrm{Tr}[b^*(P) X_i] \tag{5.6.32}$$

is the optimal utility sequence.

Then

$$\hat{S}_n = S_n^* 2^{no(1)}, \tag{5.6.33}$$

where $o(1) \longrightarrow 0$ in probability.

Consequently, $\hat{S}_n$ has the same asymptotical exponent with S_n^*.

Proof. At first, by the Markov inequality, we observe that

$$\mathbb{P}(\{\frac{\hat{S}_n}{S_n^*} > 2^{n\epsilon}\}) \leq 2^{-n\epsilon} \mathbb{E}\frac{\hat{S}_n}{S_n^*} \tag{5.6.34}$$

$$\leq 2^{-n\epsilon} \tag{5.6.35}$$

by our generalized Kuhn-Tucker conditions for the optimality of b^*, i.e., $\mathbb{E}\frac{\hat{S}_n}{S_n^*} \leq 1$, where $\mathbb{P}(.)$ denotes the probability of one set.

Let $y^+ = \max\{0, y\}$, and $y^- = \max\{0, -y\}$. We observe that

$$\mathbb{E}(\log \frac{S_n^*}{\hat{S}_n})^- = \mathbb{E} \log \max\{\frac{\hat{S}_n}{S_n^*}, 1\} \tag{5.6.36}$$

$$\leq \mathbb{E} \log(1 + \frac{\hat{S}_n}{S_n^*}) \tag{5.6.37}$$

by the property of max,

$$\leq \log(1 + \mathbb{E}\frac{\hat{S}_n}{S_n^*}) \tag{5.6.38}$$

by the concavity of the logarithm,

$$\leq \log 2 = 1. \tag{5.6.39}$$

again by our generalized Kuhn-Tucker conditions for the optimality of b^*, i.e., $\mathbb{E}\frac{\hat{S}_n}{S_n^*} \leq 1$.

Therefore, we have the following result

$$\mathbb{P}(\{\frac{S_n^*}{\hat{S}_n} > 2^{n\epsilon}\}) = \mathbb{P}(\{\log \frac{S_n^*}{\hat{S}_n} > n\epsilon\}) \tag{5.6.40}$$

$$\leq \frac{1}{n\epsilon}\mathbb{E}(\log \frac{S_n^*}{\hat{S}_n})^+ \tag{5.6.41}$$

by the Markov inequality,

$$\leq \frac{1}{\epsilon}[\frac{1}{n}\mathbb{E}\log \frac{S_n^*}{\hat{S}_n} + \frac{1}{n}] \tag{5.6.42}$$

by inequality (5.6.39).

Thus, from the inequalities (5.6.35) and (5.6.42), we result that $\frac{\hat{S}_n}{S_n^*} = 2^{no(1)}$ in probability. □

5.6.3 Parameterized Probability Distribution

Let $\{p_\theta : \theta \in \Theta\}$ be a parameterized family of distributions on a measurable space with $\Theta \subset \mathbb{R}^d$, and suppose that $\{X_1\}, ..., \{X_n\}$ are independent and identically distributed random awards with respect to the distribution p_θ, where $\theta \in \Theta$.

We take the probability measures p_θ for probability density functions $p_\theta(X)$ with respect to a fixed sigma-finite measure $\lambda(dX)$. The outcomes of $\{X_n\}$ are denoted, in this section, by X_n and a sequence of n random awards $\{X^n\}$ with outcomes X^n, in convenience.

Let $\omega(\theta)$ be the prior density for θ with respect to Lebesgue measure. We take the Bayesian marginal density function for X^n with respect to λ^n as the mixture of the conditional densities

$$p^n(X^n|\theta) = \prod_{i=1}^{n} p^n(X_i|\theta) \tag{5.6.43}$$

by integrating with respect to the prior $\omega(\theta)$, that is,

$$m_n(X^n) = \int_\Theta \omega(\theta) p^n(X^n|\theta) d\theta. \tag{5.6.44}$$

The mixture distribution is denoted by M_n.

If we use the predictive density identifier

$$\hat{p}_n(X) = m(X_{n+1}|X_1, ..., X_n), \tag{5.6.45}$$

this Bayes's sequential identification strategy is optimal with respect to M_n.

In this section, we apply the asymptotic behavior of classical relative entropy (between the n-fold product of a given member of a parameterized family of distributions p_θ and a mixture of products of such distributions M_n) to sequentially identify quantum state.

For smoothly parameterized families and for priors assigning positive mass to neighborhoods of θ_0, we need the result that classical relative entropy increases in proportion to the logarithm of the sample size plus a constant [32].

At first we introduce related conditions [32] as follows.

Condition C_1: For almost every X, the density $p_\theta(X)$ is twice continuously differentiable at θ_0, and there exists a $\delta > 0$ so that for each j and k from 1 to d,

$$\mathbb{E} \sup_{\|\theta-\theta_0\|<\delta} |\frac{\partial^2}{\partial\theta_j\partial\theta_k} \log p(X|\theta)|^2 < \infty, \tag{5.6.46}$$

and

$$\mathbb{E}|\frac{\partial}{\partial\theta_j} \log p(X|\theta)|^2 < \infty. \tag{5.6.47}$$

Condition C_2 At θ_0, relative entropy $\mathcal{R}(p; p_{\theta_0})$ is twice continuously differentiable with J_{θ_0} positive definite, and the prior $\omega(\theta)$ is continuous and positive at θ_0, where the Fisher information $I(\theta_0)$ and the second derivative matrix $J(\theta_0)$ (for the informational divergence or relative entropy $\mathcal{R}(p; p_{\theta_0})$) are defined as follows.

$$I(\theta_0) = \mathbb{E}[\frac{\partial}{\partial\theta_j} \log p(X|\theta) \frac{\partial}{\partial\theta_k} \log p(X|\theta)]_{j,k=1,2,...,d} \tag{5.6.48}$$

and

$$J(\theta_0) = [\frac{\partial^2}{\partial\theta_j\partial\theta_k} \mathcal{R}(p; p_{\theta_0})]_{j,k=1,2,...,d}, \tag{5.6.49}$$

where in each case the derivatives are evaluated at $\theta = \theta_0$.

Condition C_3 Given X^n, the posterior distribution of θ asymptotically concentrates on neighborhoods of θ_0, except for X^n in a set of probability of order $o(1/\log n)$, that is,

$$\mathbb{P}^n(\{W(N^c \mid X^n) > \delta\}) = o(1/\log n), \tag{5.6.50}$$

for every open set N containing θ_0 and every $\delta > 0$, where $W(. \mid X^n)$ is the posterior distribution of θ given X^n.

To lower bound the utility of the Bayes' strategy in terms of the optimal utility, we need the following important formula in [32].

Lemma 5.6.6. *([32]) Suppose the parametric family p_θ and the prior $\omega(\theta)$ satisfy the smoothness Condition C_1 and Condition C_2 and posterior consistency Condition C_3 for θ_0 in the interior of Θ. Then*

$$\lim_{n\to\infty}\{\mathcal{R}(p_{\theta_0}^n; M_n) - \frac{d}{2}\log\frac{n}{2\pi}\} = \log\frac{1}{\omega(\theta_0)} + \frac{1}{2}\log\det J_{\theta_0} - \frac{1}{2}\mathrm{tr}(I_{\theta_0}J_{\theta_0}^{-1}). \tag{5.6.51}$$

Theorem 5.6.7. *Upon M_n, the Bayes' strategy achieves utility at least approximately, that is,*

$$S_n \geq S_n^* e^{-\mathcal{R}(p_{\theta_0}^n; M_n)} \tag{5.6.52}$$

$$= S_n^*(\frac{2\pi e}{n})^{\frac{d}{2}}\frac{\omega(\theta_0)}{\sqrt{\det I_{\theta_0}}} \tag{5.6.53}$$

under the condition C_1, condition C_2, and condition C_3, where

$$\hat{S}_n = \prod_{i=1}^n \mathrm{Tr}[\hat{b}_i X_i] \tag{5.6.54}$$

and

$$S_n^* = \prod_{i=1}^n \mathrm{Tr}[b_i^* X_i] \tag{5.6.55}$$

following the above section.

Factually, for any $\alpha \in (0,1)$ and any $\tau > 0$ with $\alpha > e^{-\tau}$, we have the following inequality

$$S_n \geq (\alpha - e^{-\tau})S_n^* e^{-\mathcal{R}(p_{\theta_0}^n; M_n) + \frac{1}{2}c(\alpha) - \tau} \tag{5.6.56}$$

except on a set with probability asymptotically less than or equal to $\alpha + e^{-\tau}$, as $n \to \infty$, where $c(\alpha)$ is the $1-\alpha$ quantile of a centered chi-square random variable with d freedom, that is,

$$\mathbb{P}(\{\chi_d^2 - \mathbb{E}\chi_d^2 > c\}) = \alpha. \tag{5.6.57}$$

Proof. Except on a set of probability

$$\mathbb{P}_{\theta_0}\left(\{X^n | \frac{S_n^*}{S_n}\frac{m(X^n)}{p(X^n|\theta_0)} \geq e^\tau\}\right) \leq e^{-\tau}\mathbb{E}_{\theta_0}\frac{S_n^*}{S_n}\frac{m(X^n)}{p(X^n \mid \theta_0)} \tag{5.6.58}$$

$$\leq e^{-\tau}\mathbb{E}_{m_n}\frac{S_n^*}{S_n} \tag{5.6.59}$$

$$\leq e^{-\tau} \tag{5.6.60}$$

by $\mathbb{E}_{m_n}\frac{S_n^*}{S_n} \leq 1$ for the optimality of S_n for the distribution M_n, we have the following result

$$S_n \geq S_n^* \frac{m(X^n)}{p(X^n \mid \theta_0)} \tag{5.6.61}$$

by the Markov inequality.

Applying above formula directly to (5.6.61) under the *condition* C_1, *condition* C_2, and *condition* C_3, we have the following result

$$S_n \geq S_n^* \exp\{-\mathcal{R}(p_{\theta_0}^n; M_n)\} \tag{5.6.62}$$

$$= S_n^* (\frac{2\pi e}{n})^{\frac{d}{2}} \frac{\omega(\theta_0)}{\sqrt{\det I_{\theta_0}}}. \tag{5.6.63}$$

Let A_n be any critical region $\{X^n | \frac{S_n^*}{S_n}\frac{m(X^n)}{p(X^n|\theta_0)} \geq e^\tau\}$, thus $\mathbb{P}_{\theta_0}(A_n) \leq e^{-\tau}$ from (5.6.60), and let B_n be the set

$$B_n = \{X^n | \log \frac{p(X^n|\theta_0)}{m(X^n)} \leq \mathcal{R}(p_{\theta_0}^n; M_n) - \frac{1}{2}c(\alpha)\}, \tag{5.6.64}$$

where $\alpha > e^{-\tau}$.

Recalling the fact

$$\log \frac{m(X^n)}{p(X^n|\theta_0)} + \mathcal{R}(p_{\theta_0}^n; M_n) \longrightarrow \frac{1}{2}(\chi_d^2 - d) \tag{5.6.65}$$

in distribution, where χ_d^2 has a centered chi-square distribution with d degrees of freedom, and

$$\mathbb{P}(\{\chi_d^2 - \mathbb{E}\chi_d^2 > c\}) = \alpha, \tag{5.6.66}$$

we obtain the result

$$\mathbb{P}_{\theta_0}^n(B_n) \longrightarrow \alpha. \tag{5.6.67}$$

Obviously, $M_n(A_n^c) \geq M_n(A_n^c \cap B_n)$, but according to the definition of A_n, we can obtain

$$M_n(A_n^c) \leq e^\tau \frac{S_n}{S_n^*}, \tag{5.6.68}$$

while, from the definition of B_n, we can obtain

$$M_n(A_n^c \cap B_n) \geq e^{-\mathcal{R}(p_{\theta_0}^n; M_n) + \frac{1}{2}c(\alpha)} \mathbb{P}_{\theta_0}^n(A_n^c \cap B_n) \tag{5.6.69}$$

$$\geq e^{-\mathcal{R}(p_{\theta_0}^n; M_n) + \frac{1}{2}c(\alpha)} (\mathbb{P}_{\theta_0}^n(A_n^c) - \mathbb{P}_{\theta_0}^n(B_n^c)). \tag{5.6.70}$$

Since $\mathbb{P}_{\theta_0}^n(A_n^c) - \mathbb{P}_{\theta_0}^n(B_n^c) \longrightarrow \alpha - e^{-\tau} > 0$, therefore

$$e^\tau \frac{S_n}{S_n^*} \geq (\alpha - e^{-\tau}) e^{-\mathcal{R}(p_{\theta_0}^n; M_n) + \frac{1}{2}c(\alpha)}, \tag{5.6.71}$$

thus

$$S_n \geq (\alpha - e^{-\tau}) S_n^* e^{-\mathcal{R}(p_{\theta_0}^n; M_n) + \frac{1}{2}c(\alpha) - \tau}, \tag{5.6.72}$$

except on a set $A_n \cup B_n$, with

$$\mathbb{P}_{\theta_0}^n(A_n^c \cup B_n) \leq \mathbb{P}_{\theta_0}^n(A_n^c) + \mathbb{P}_{\theta_0}^n(B_n) \longrightarrow \alpha + e^{-\tau}. \tag{5.6.73}$$

□

CHAPTER 6

Conclusions

So far, this project is, bridging entropy econometrics, game theory and information theory, a game theoretic approach to quantum information, extending classical econometrics to noncommutative data, and a game theoretic identification of quantum state, practically giving a game theoretic interpretation of abstract classical information theory, upon which curious financial stock may be designed (or suitably "quantum" stock in physical implementation).

In chapter 3, continuing [16–19] on quantum channel capacity for one-way communication via entanglement, we treated two types of quantum mutual information and corresponding quantum channel capacities via entanglement.

Upon the monotonicity property of quantum mutual information of Araki-Umegaki type and Belavkin-Staszewski typ introduced in [14, 16], we proved the additivity property of entangled quantum channel capacities, which therefore extended the results of V. P. Belavkin [16, 17] to products of arbitrary quantum channel and to quantum relative entropy of both Araki-Umegaki type and Belavkin-Staszewski type.

Well, quantum channel capacities have several different formulations when considering to send classical information or quantum information, one-way or two-way communication, prior or via entanglement, etc., in the form of different constraints on the encoding class $\mathcal{K}$. Anyway general quantum channel capacity with different constraints is still a big open and challenging research problem in quantum information. Much more open problems can be found in [92].

Another natural problem in this direction is to compare true quantum capacities in quantity for some interesting quantum channels with other smaller capacities under constraints, such as Holevo capacity, entanglement-assistant capac-

ity, etc., and find for which channels they coincide.

The third natural problem in this direction is to consider entangled quantum mutual information and corresponding quantum channel capacities for γ type since [39] studied this third and more general quantum relative entropy in quantum information, which also meet more natural axiomatic properties of relative entropy.

Chapter 4 began a game theoretic application to quantum information.

In the introduction to strategic game, we obtained a sufficient condition for minimax theorem, but it is still worth to explore the necessary condition of minimax theorem.

In the exploration of classical/quantum estimate, we found the existence of the minimax value of the game and its minimax strategy, but it is still interesting to discuss the existence of maximin value and its maximin strategy. One further problem is to give the general bounds on those values.

In the view of biconjugation in convex analysis, we arrived at one approach to quantum relative entropy, quantum mutual entropy, and quantum channel capacity and obtained the monotonicity of quantum relative entropy and the additivity of quantum channel capacity, and it is still worth to explore other properties and their bounds.

Applying Kelly's criterion to identify quantum state, in the chapter 5, we have given a practical game theoretic interpretation to classical relative entropy, mutual information, and asymptotical information, during which we find following results.

The decrement in the doubling rate achieved with true knowledge of the distribution F over that achieved with incorrect knowledge G is bounded by relative entropy $\mathcal{R}(F;G)$ of F relative to G.

The increment Δ in the doubling rate resulting from side information Y is less than or equal to the mutual information $\mathcal{I}(X,Y)$.

A good sequence of true quantum state identifications leads to asymptotically optimal growth rate of utility.

Applying the asymptotic behavior of classical relative entropy, the utility of the Bayes' strategy is bounded below in terms of the optimal utility.

However, several fundamental open problems exist in this part: practical game theoretic definitions are needy for relative entropy, mutual information, and

asymptotical information, even in the classical information theory; it is worth investigating the asymptotic behavior of quantum relative entropy.

All those wait further possible papers!

References

[1] L. Accardi, F. Frigerio and J. T. Lewis, Quantum Stochastic Process, *Publ. RIMS* **18** 97-133 (1982).

[2] H. Araki, Relative Entropy of states of von Neumann Algebras, *Publications RIMS, Kyoto University* **11** 809 (1976).

[3] Arnold, Vladimir Igorevich, *Mathematical Methods of Classical Mechanics* second edition (Springer, 1989).

[4] V. P. Belavkin, Optimal Estimation of Noncommutating Quantum Gaussian Variables under the Sequential Inaccurate Measurement, *Radio Eng Electron Physics* **17** no. 12, 527-2532 (1972).

[5] V. P. Belavkin, Linear Estimation of Non-commuting Observables by their Indirect Measurement, *Radio Eng Electron Physics* **17** no. 12, 2533-2540 (1972).

[6] V. P. Belavkin and B. Grishanin, Study of the Optimal Estimation Problem in Quantum Channels by a Generalised Heisenberg Inequality Method, *Problems of Information Transmission* **9** no. 3, 44-52 (1973).

[7] V. P. Belavkin and R. L. Stratonovich, Optimization of Quantum Information Processing Maximizing Mutual Information, *Radio Eng. Electron. Phys.* **19**, 1349 (1973).

[8] V. P. Belavkin, Optimal Observation of Boson Signals in Quantum Gaussian Channels, *Problems of Control and Information Theory* no. 4, 241-257 (1975).

[9] V. P. Belavkin, Optimal Discrimination of Non-orthogonal Quantum Signals, *Radio Eng Electron Physics* **20** no. 6, 1177-1185 (1975).

[10] V. P. Belavkin, Optimal Multiple Quantum Statistical Hypothesis Testing, *Stochastics* no. 1, 315-345 (1975).

[11] V. P. Belavkin, Hypothesis Testing for Quantum Optical Fields, *Radio Eng Electron Physics* **21** no. 6, 95-104 (1976).

[12] V. P. Belavkin, Generalized Heisenberg Uncertainty Relations and Efficient Measurements in Quantum Systems, *Theoret Math Physics* no. 3, 316-329 (1976).

[13] V. P. Belavkin, Optimal Quantum Filtration of Markovian Signals, *Problems of Control and Information Theory* **7** no. 5, 345-360 (1978).

[14] V. P. Belavkin and P. Staszewski, C^*-algebraic generalization of relative entropy and entropy, *Ann. Inst. Henri Poincare* **37** Sec. A 51-58 (1982).

[15] V. P. Belavkin and V. Maslov, Design of Optimal Dynamic Analyzer: Mathematical Aspects of Wave Pattern Recognition, *Mathematical Aspects of Computer Engineering, Ed V. Maslov* 146-237 (Mir, Moscow, 1987).

[16] V. P. Belavkin, On Entangled Quantum Capacity, *Quantum Communication, Computing, and Measurement.* **3** 325-333 (2001).

[17] V. P. Belavkin, On Entangled Information and Quantum Capacity, *Open Sys. and Information Dyn.* **8**1-18 (2001).

[18] V. P. Belavkin and M. Ohya, Quantum Entropy and Information in Discrete Entangled States, *Infinite Dimensional Analysis, Quantum Probability And Related Topics.* **4** 137-160 (2001).

[19] V. P. Belavkin and M. Ohya, Entanglement, Quantum Entropy and Mutual Information, *Proc. R. Soc. Lond.* **A 458** 209 - 231 (2002).

[20] V. P. Belavkin, Quantum stochastic positive evolutions: characterization, construction, dilation, *Commun. Math. Phys.* **184** 533-566 (1997).

[21] V. P. Belavkin and X. Dai, Additivity of Entangled Channel Capacity given Quantum Input State, *Proceedings of Quantum Probability, Quantum Information, and Quantum Control 2006*, *quant-ph/0702098* (2007).

[22] R. Bellman and R. Kalaba, On the role of dynamic programming in statistical communication theory, *IRE Trans. of the professional group on information theory* **IT-3**, no. 3, 197-203 (1957).

[23] C. H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres and W. K. Wootters, Teleporting an unknown quantum state via dual classic and Einstein-Podolsky-Rosen channels, *Phys. Rev. Lett.* **70** 1895-1899 (1993).

[24] C. H. Bennett, P. W. Shor, J. A. Smolin and A. V. Thapliyal, Entanglement assisted classical capacity of noisy quantum channels, *Phys. Rev. Lett.* **83** 3081-3084 (1999).

[25] C. H. Bennett, P. W. Shor, J. A. Smolin and A. V. Thapliyal, Entanglement-Assisted Capacity of a Quantum Channel and the Reverse Shannon Theorem, *quant-ph/0106052* (2001).

[26] B. Blackadar, *Operator algebras* (Springer, 2005).

[27] E. Borel, The Theory of Play and Integral Equations with Skew Symmetric Kernels, *Econometrica* **21** 97-100 (1953).

[28] M. Born, W. Heisenberg and P. Jordan, Zur Quantenmechanik II, *Zeitschrift für Physik***35** 557-615 (1925).

[29] L. Breiman, Optimal gambling systems for favorable games, *Fourth Berkeley Symposium on probability and statistics* **I** 65-78 (1961).

[30] N. Cerf and G. Adami, Von Neumann capacity of noisy quantum channels, *Phys. Rev.* **A 56**, 3470-3483 (1997).

[31] Y. S. Chow and Teicher Henry, *Probability Theory: Independence, Interchangeability, Martingales* (Springer-Verlag, New York, 1978).

[32] B. S. Clarke and A. R. Barron, Information-Theoretic Asymptotic of Bayes Methods, *IEEE Transactions on Information Theory* **36** no. 3, 453-471 (1990).

[33] A. Connes, *Non-commutative geometry* (Academic Press, 1994).

[34] X. Dai and V. P. Belavkin, A Game Theoreic Approach to Quantum Information, *arXiv:0710.0556* (2007).

[35] J. Dixmier, *Von Neumann algebras* A translation of J. Dixmier (1957) (1981).

[36] A. Einstein, B. Podolsky and N. Rosen, Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.* **47** 777-780 (1935).

[37] A. Ekert, Quantum cryptography based on Bell's Theorem, *Phys. Rev. Lett.* **67** 661-663 (1990).

[38] Peter D. Grünwald and A. Philip Dawid, Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory, *Ann. Statist.* **32** no. 4, 1367-1433 (2004).

[39] S. J. Hammersley and V. P. Belavkin, Information Divergence for Quantum Channels, Infinite Dimensional Analysis, *Quantum Information and Computing* World Scientific, Quantum Probability and White Noise Analysis, VXIX 149-166 (2006).

[40] Harsanyi, John C., An equilibrium point interpretation of stable sets, *Management Science* **20**, 1472-1495 (1974).

[41] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[42] A. S. Holevo, Information theoretical aspects of quantum measurements, *Probl. Inform. Transm.* **9**, no. 2, 31-42 (1973).

[43] A. S. Holevo, On asymptotically optimal hypotheses testing in quantum statistics, *Theor. Probab. and Appl.* **23** no. 2, 429-432 (1978).

[44] A. S. Holevo, Quantum coding theorems, *Russian Math. Surveys.* **53**, 1295-1331 (1998).

[45] A. S. Holevo, the capacity of the quantum channel with general signal states, *IEEE Trans. Inform. Theory* **44**, no. 1, 269-273 (1998).

[46] Michal Horodecki, Jonathan Oppenheim, and Andreas Winter, Partial quantum information, *Nature* **436** 673-676 (2005).

[47] Jacob Bernoulli, *The Art of Conjecture*, by Nicholas (1713).

[48] E. T. Jaynes, Information Theory and Statistical Mechanics, *Phys. Rev.* **106** 620 (1957).

[49] E. T. Jaynes, Information Theory and Statistical Mechanics II, *Phys. Rev.* **108** 171 (1957).

[50] R. Jozsa and B. Schumacher, A new proof of the quantum noiseless coding theorem, *J. Mod. Opt.* **41** 2343-2350 (1994).

[51] S. Kakutani, A generalization of Brouwer's fixed point theorem, *Duke Mathematical Journal* **8** 457-459 (1941).

[52] J. Kelly, New interpretation of information rate, *Bell Syste. Tech. J.* **35** 917-926 (1956).

[53] J. Kovacs and J. Szcs, Ergodic type theorems in von Neumann algebras, *Acta Sci. Math. Szeged* **27** 233-246 (1966).

[54] David M. Kreps and Robert Wilson, Sequential Equilibria, *Econometrica* **50**, no. 4, 863-894 (1982).

[55] Harold W. Kuhn and Albert William Tucker (eds), *Contributions to the Theory of Games* **1** *Annals of Mathematics Studies* no. 24, 19-26, 81-96 (Princeton University Press, Princeton, 1950).

[56] Harold W. Kuhn and Albert William Tucker (eds), *Contributions to the Theory of Games* **2** *Annals of Mathematics Studies* no. 28, 5-12 (Princeton University Press, Princeton, 1953).

[57] H. Kuhn, Extensive Games and the Problem of Information, *Contributions to the Theory of Games* **2** ed. by H. Kuhn and A. Tucker 193-216 (Princeton University Press, Princeton, 1953).

[58] Harold W. Kuhn, Lectures on the Theory of Games, *Annals of Mathematics Studies* **37** 21-24 (Princeton University Press, Princeton and Oxford, 2003).

[59] S. Kullback, and R. A. Leibler, On information and sufficiency, *Annals of Mathematical Statistics* **22** 79-86 (1951).

[60] S. Kullback, Certain Inequalities in Information Theory and the Cramer-Rao Inequality, *The Annals of Mathematical Statistics* **25** 745-751 (1954).

[61] S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959).

[62] Burkhard Kümmerer and Hans Maassen, Elements of Quantum Probability, *Quantum Probability Communications* **X** 73-100 (1998). eds. R.L. Hudson, J.M. Lindsay, (World Scientific, Singapore, 1998).

[63] L. D. Landau, and E. M. Lifschitz, *Quantum Mechanics (Non-Relativistic Theory)*, 3rd ed. (Oxford, England: Pergamon Press, 1977).

[64] C. F. Lee and N. F. Johnson, Game-theoretic discussion of quantum state estimation and cloning, *Physics Letters A***319** no. 5, 429-433 (2003).

[65] Peter Levay, The geometry of entanglement: metrics, connections and the geometric phase, *J. Phys.* **A37** 1821-1842 (2004).

[66] Peter Levay, The twistor geometry of three-qubit entanglement, *quant-ph/0403060* (2004).

[67] D. V. Lindley, On a measure of information provided by an experiment, *The Annals of Mathematical Statistics* **27** 986-1005 (1956).

[68] G. Lindblad, Entropy, Information and Quantum Measurements, *Comm. In Math. Phys.* **33** 305-322 (1973).

[69] G. Lindblad, Quantum entropy and quantum measurements, *Lect. Notes Phys.* **378**, 71-80, (Springer-Verlag, Berlin, 1991).

[70] John Maynard Smith, *Evolution and the Theory of Games*, (Cambridge University Press, 1982).

[71] P. A. Meyer, Quantum Probaility for Porbabilitists, *Lecture Notes in Mathematics* **1538** (Springer-Verlag, Berlin Heidelberg, 1993).

[72] M. Nakamura and H. Umegaki, A note on entropy for operator algebras, *Proc. Japan Acad.* **37** 149-154 (1961).

[73] J. F. Nash, Equilibrium Points in N-Person Games, *Proceedings of the National Academy of Sciences of United States of America* **36** 48-49 (1950).

[74] J. F. Nash, Non-Cooperative Games, *Annals of Mathematics***54** 286-295 (1951).

[75] J. von Neumann, On the Theory of Games of Strategy, *Contributions to the Theory of Games* **5** 13-42 (1959).

[76] J. von Neumann, Zur Algebra der Funktionaloperationen und Theorie der normalen Operatoren, *Math. Ann.* **102** 370-427 (1929).

[77] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, 1944).

[78] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, (Princeton University Press, 1955).

[79] Michael A. Nielsen and Isaac L. Chuang, *Quantum Computation and Quantum Information*, (Cambridge University Press, 2000).

[80] M. Ohya and D. Petz, *Quantum Entropy and its Use*, (Springer-Verlag, Berlin, 1993).

[81] Martin J. Osborne and Ariel Rubinstein, *A Course in Game Theory* (MIT Press, 1994).

[82] R. Penrose, *Report in NCG workshop in Newton Institute* (Cambridge, Sept., 2006).

[83] Ralph Tyrell Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, 1970)

[84] S. Sakai, *C^*-algebras and W^*-algebras* (Springer, 1971).

[85] B. Schumacher, Quantum Coding, *Phys. Rev.* **A 51**, 2738-2747 (1995).

[86] B. Schumacher, Sending entanglement through noisy quantum channels, *Phys. Rev.* **A 54**, 2614-2628 (1996).

[87] Benjamin Schumacher and Michael D. Westmorel, Quantum mutual information and the one-time pad, *quant-ph/0604207* (2006).

[88] Jacob T. Schwartz, *W^*-algebras* (New York: Gordon and Breach, 1967).

[89] I. E. Segal, A note on the concept of entropy, *J. Math. Mech.* **9** 623-629 (1960).

[90] R. Selten, Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetragheit, *Zeitschrift fur die gesamte Staatswissenschaft* **121** 301-324 (1965).

[91] C. E. Shannon, a mathematical theory of communication, *The Bell System Technical Journal* **27** 379-423, 623-656 (1948).

[92] Peter Shor, Quantum Information Theory: Results and Open Problems, *Geom. Funct. Anal., Special Volume-GAFA2000*, 816-838 (2000).

[93] W. F. Stinespring, Positive functions on C^*-algebras, *Proc. Amer. Math. Soc.***6** 211 (1955).

[94] E. O. Thorp, Optimal gambling systems for favorable games, *Review of the International Statistical Institute* **37** 3 (1969).

[95] R. James Tomkins, Another Proof of Borel's Strong Law of Large Numbers, *The American Statistician* **38** no. 3, (1984).

[96] R. James Tomkins, General Approaches to Strong Limit Theorems, *Selecra Srarisrica Canadians* no. 5, 173-200 (1979).

[97] H. Umegaki, Conditional expectation in an operator algebra. IV. Entropy and information, *Kodai Math. Sem. Rep.* **14** 59 (1962).

[98] A. Wehrl, General properties of entropy, *Rev. Mod. Phys.* **50** 221 (1978).

[99] David H. Wolpert, Information Theory-The Bridge Connecting Bounded Rational Game Theory and Statistical Physics, *CoRR cond-mat/0402508* (2004).

[100] A. Zellner, Causality and Causal Laws in Economics, *Journal of Econometrics* **39**, 7-21 (1988).

[101] A. Zellner, Time series analysis, forecasting and econometric modeling: The structural econometric modeling, time series analysis approach, *Journal of Forecasting* **13**, 215-233 (1994).

[102] A. Zellner, The Bayesian Method of Moments: Theory and Applications, *Advances in Econometrics* **12**, 85-105 (1997).

[103] A. Zellner, J. Tobias and H. Ryu, Bayesian Method of Moments (BMOM) Analysis of Parametric and Semiparametric Regression Models, *South African Statistical Journal* **31**, 41-69 (1999).

[104] R. A. L. Carter and A. Zellner, The ARAR Model for Time Series, *Proceedings of Section of the Bayesian Statistical Science of the American Statistical Association (1996)*, 226-331 (1997).

[105] A. Zellner and J. Tobias, Further Results on Bayesian Method of Moments Analysis of the Multiple Regression Model, *International Economic Review* **42** no. 1, 121-140 (2001).

[106] A. Zellner, Remarks on a 'Critique' of the Bayesian Method of Moments, *Journal of Applied Statistics* **28** no. 6, 775-778 (2001).