Studies On Bell's Theorem

Veli Ugur Guney

Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Physics Commons

Recommended Citation

https://academicworks.cuny.edu/gc_etds/958

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
STUDIES ON BELL’S THEOREM

by

VELİ UĞUR GÜNÉY

A dissertation submitted to the Graduate Faculty in Physics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2015
This manuscript has been read and accepted for the Graduate Faculty in Physics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Date
Prof. Mark Hillery
Chair of Examining Committee

Date
Prof. Igor L. Kuskovsky
Executive Officer

Prof. János A. Bergou

Prof. Christopher C. Gerry

Prof. Daniel M. Greenberger

Prof. Neepa T. Maitra

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract

STUDIES ON BELL’S THEOREM

by

Uğur Güney

Adviser: Professor Mark Hillery

In this work we look for novel classes of Bell’s inequalities and methods to produce them. We also find their quantum violations including, if possible, the maximum one.

The Jordan bases method that we explain in Chapter 2 is about using a pair of certain type of orthonormal bases whose spans are subspaces related to measurement outcomes of incompatible quantities on the same physical system. Jordan vectors are the briefest way of expressing the relative orientation of any two subspaces. This feature helps us to reduce the dimensionality of the parameter space on which we do searches for optimization. The work is published in [24].

In Chapter 3, we attempt to find a connection between group theory and Bell’s theorem. We devise a way of generating terms of a Bell’s inequality that are related to elements of an algebraic group. The same group generates both the terms of the Bell’s inequality and the observables that are used to calculate the quantum value of the Bell expression. Our results are published in [25][26].

In brief, Bell’s theorem is the main tool of a research program that was started by Einstein, Podolsky, Rosen [19] and Bohr [8] in the early days of quantum mechanics in their discussions about the core nature of physical systems. These debates were about a novel type of physical states called superposition states, which are introduced by quantum mechanics and manifested in the apparent inevitable randomness in measurement outcomes of identically prepared systems.

Bell’s huge contribution was to find a means of quantifying the problem and hence of opening the way to experimental verification by rephrasing the questions as limits on certain combinations of correlations between measurement results of spatially separate systems [7]. Thanks to Bell, the fundamental questions related to the nature of quantum mechanical systems became quantifiable [6].

According to Bell’s theorem, some correlations between quantum entangled systems that involve incompatible quantities are not allowed by classical mechanics, a feature that is called as “quantum nonlocality”.
An experimental observation of those correlations, in other words, a violation of the limits imposed by classical physics, implies the correctness of quantum description and invalidates the classical, local realistic models.

The first Bell experiments were proposed by Clauser, Horne, Shimony, and Holt, who invented the most famous Bell’s inequality [13]. Later, the Aspect experiments were satisfactory enough for the physics community to be conclusive about the validation of quantum mechanics [1][3][4][2].

Ekert’s work on applications of quantum nonlocality to communication resulted in the new field of quantum communication and cryptography, and turned the research program into a practical one [20].

Pitowsky showed a method to find all expressions of limitations due to local realism, all Bell’s inequalities, for a given physical scenario. He also proved that the problem is, unfortunately, NP-complete and hence as the scenarios get more complex, they also become computationally intractable [33][34]. Therefore, different methods for the solution of special cases of the problem are necessary.

Inequalities found for those special cases can be called classes of Bell’s inequalities. For example, Werner and Wolf [41] and Collins, Gisin, Linden, Massar, and Popescu [16] found classes that cover a wide range of scenarios.

Our work is a similar kind of effort to produce and study new types of Bell’s inequalities.
Acknowledgements

I am deeply grateful to my adviser Mark Hillery. He provided me the coolest research problems and helped and allowed me to study a field that I find immensely interesting. Hard to believe, but thanks to his grant applications I studied a highly impractical topic and still got paid for it. His way of applying detailed and much necessary rigor to all aspects of research is going to be a guide for the rest of my life. He evaluated my proposals carefully with respect, shared his experienced feedback and answered my unending questions patiently. Each visit to his office was a valuable lesson and provided lots of material to think about. Prof. Hillery proved to me that the pen is mightier than the computer so many times. He also gave me enough freedom. He is definitely not a person who sees his students as research robots. Sharing similar views about contemporary topics made our meetings fun too. Thank you Professor!

I appreciate my dissertation committee members’ efforts and evaluations of this work. János A. Bergou, Christopher C. Gerry, Daniel M. Greenberger and Neepa T. Maitra shared their time to listen to my research proposals and oversaw the research process.

Thanks to Andi Shehu and Vadim Yerokhin for making the office environment a nice and warm one. We had so many conversations that I would like to continue in the future. Also thanks to Jorge Colon, Kay Hiranaka, Malory Gobet, Zehra Cevher, Aruna Dhawan, Arthur Parzygnat, Thomas Proctor, Vasilios Deligiannakis and other friends from Graduate Center physics department for their support, valuable discussions and fun times.

Thanks to Axel and Rey Cortes brothers for great musical collaboration and sharing their talent with me.

Thanks to my brother Murat Güney for reading my introduction to give me a non-physicist’s view of my topic and make it more understandable, and also taking care of me while I am under stress.

Thanks to my girlfriend Rae Ann Winkelstein for being so supportive with my whole PhD journey, being encouraging and editing some parts of the dissertation.

This work is dedicated to my mother Zehra Kocabas Güney whose dreams of getting a graduate level education abroad were thwarted by conservative and oppressive traditions that took away women’s freedom. Though, no matter what the obstacles are, now, some part of you achieved your old goal. I am proud of being a vehicle.
“Zerrdir belki, ama yok denilmaz”

(“Perhaps it’s just a mote, yet undeniable”)

Replikas - Zerre
## Contents

1 Introduction ............................................... 1

1.1 A Verbal Description of Bell’s Theorem .................. 1

1.1.1 The Superposition State ............................ 1

1.1.2 Superposition State When Dealing with a Single Quantity ...... 3

1.1.3 Superposition State When Dealing with Two Quantities, Sharpness of a Probability Distribution, Uncertainty Principle ... 4

1.1.4 Two Spatially Separated Systems ....................... 8

1.1.5 A Bell Experiment Example, CHSH Experiment ............. 10

1.1.6 Einstein, Podolsky, Rosen Arguments .................. 12

1.1.7 Visual summary of minimum necessary Bell setup .......... 15

1.1.8 Quantum nonlocality and indeterminism .................. 16

1.1.9 Disclaimer ........................................... 17

1.2 Mathematical Description of Bell’s Theorem ............... 17

1.2.1 Random Variables ................................... 17

1.2.2 Joint Events ......................................... 18

1.2.3 Conditional Probability ................................ 19

1.2.4 Statistical Dependence ................................ 20

1.2.5 Correlation .......................................... 22

1.2.6 Composite Events ..................................... 23

1.2.7 Generalization to Higher Number of Outcomes .............. 23

1.2.8 Correlation and Statistical Dependence .................. 28

1.2.9 Boole’s Conditions of Probable Experience ............... 29

1.2.10 Probable Conditions for the Simplest Case and Their Geometric Interpretation ............... 31

1.2.11 Convex Polytopes as Sets of Probable Conditions, Computational Complexity of Finding All Bell Inequalities for a Given Scenario ............... 34

1.2.12 Wigner Bell Inequality from \( N_E = 3, N_J = 3 \) Polytope ....... 36

1.2.13 Generic Bell Scenario .................................. 38

1.3 A selection of known Bell Inequalities ................... 39

1.3.1 Classes of Inequalities ................................ 39

1.3.2 Clauser, Horne, Shimony, Holt (CHSH) .................. 39

1.3.3 Greenberger, Horne, Zeilinger (GHZ) .................... 40
3.2.6 The Regular Representation .......................................................... 106
3.2.7 Character Tables .............................................................................. 107
3.2.8 Invariant Subspaces and Subrepresentations ............................... 108
3.2.9 Group Actions .................................................................................. 109
3.3 Bell Inequalities from Group Actions of Single-Generator Groups .......... 110
3.3.1 Reproduction of CH Using a Cyclic Group ...................................... 110
3.3.2 Classical Bound of Reproduced CH ................................................ 121
3.3.3 Non-local Game Properties of the Reproduced CH System ............. 122
3.3.4 Generalization to Qudits and More Than Two Measurements ........ 125
3.3.5 N = 2, M = 2, K = 3 Scenario .......................................................... 133
3.3.6 N = 2, M = 3, K = 2 Scenarios .......................................................... 135
3.3.7 N = 2, M ≥ 3, K = 2 Scenarios .......................................................... 137
3.3.8 Numerical Results for N = 2 Parties ................................................ 138
3.4 Bell Inequalities from Group Actions: N = 3 Parties ......................... 139
3.5 Bell Inequalities from Group Actions: Non-Abelian Groups ............... 144
3.5.1 Reproduction of CH Using a Cyclic Group (Reprise) ...................... 145
3.5.2 Properties of Dihedral Group ............................................................. 150
3.5.3 A Bell Violation Using D3 ................................................................. 154
3.5.4 Nonlocal Game Properties of a Z6 System ...................................... 161
3.5.5 Nonlocal Game Properties of the D3 System ................................... 163
3.5.6 A Bell Violation using D6 ................................................................. 164

4 Conclusion ............................................................................................... 170

5 Appendix ................................................................................................. 171
5.1 Computer Calculations ......................................................................... 171
5.1.1 Calculations for the Jordan Angle Method with N parties ............... 171
5.1.2 Group Theoretical Calculations in Computer Algebra System SAGE 179
5.1.3 Calculation of Classical Bounds ...................................................... 181
5.1.4 Search for Bell violations ................................................................. 183

References .................................................................................................. 185
### List of Tables

1. Occurrence probabilities of two events ............................................. 19
2. Symmetry operations corresponding to symmetries of equilateral triangle .... 93
3. The sum of probabilities for different $M$ and $K$ values .......................... 138
4. Amount of Bell violations for different scenarios .................................. 138
5. Comparison of quantum value with the mutual information for increasing dimensionality ... 139
6. Winning conditions for the non-local game of $N = 3, N = 2, K = 3$ scenario .... 144
7. Winning conditions for the $D_3$ non-local game .................................. 164
8. Probabilities generated by orbits for $D_6$ ............................................. 167
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sharp and unsharp discrete distributions</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Sharp and unsharp continuous distributions</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>Variance</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>Uncertainty principle</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Probability distributions of joint events</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>CHSH experimental setup</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>Visual summary of minimum necessary Bell setup</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>Probability space $U$ for two events $E_1$ and $E_2$, and their joint occurrence $E_1 &amp; E_2$</td>
<td>31</td>
</tr>
<tr>
<td>9</td>
<td>Different extreme cases of two events on probability space</td>
<td>32</td>
</tr>
<tr>
<td>10</td>
<td>Urn Polytope</td>
<td>33</td>
</tr>
<tr>
<td>11</td>
<td>Venn Diagram of 3 Events</td>
<td>38</td>
</tr>
<tr>
<td>12</td>
<td>Venn diagram of a Wigner-Bell inequality</td>
<td>38</td>
</tr>
<tr>
<td>13</td>
<td>General Bell scenario</td>
<td>39</td>
</tr>
<tr>
<td>14</td>
<td>Geometric depiction of Jordan angles</td>
<td>48</td>
</tr>
<tr>
<td>15</td>
<td>Size of maximum quantum violation of inequality versus number of parties</td>
<td>70</td>
</tr>
<tr>
<td>16</td>
<td>Symmetries of a triangle</td>
<td>93</td>
</tr>
<tr>
<td>17</td>
<td>Great orthogonality theorem</td>
<td>102</td>
</tr>
<tr>
<td>18</td>
<td>States generated by $D_3$ group actions</td>
<td>155</td>
</tr>
<tr>
<td>19</td>
<td>Cayley graph of $D_3$</td>
<td>156</td>
</tr>
<tr>
<td>20</td>
<td>States generated by $D_6$ group actions</td>
<td>166</td>
</tr>
</tbody>
</table>
1 Introduction

Bell’s theorem [7] is at the heart of the distinction between the two physical models of reality: classical physics and quantum physics. By performing experiments related to a Bell inequality which involve measuring correlations between random variables belonging to distant subsystems, one can tell whether the physical system at hand can be explained inside the paradigm of classical mechanics or whether quantum mechanics is necessary. If the results of those measurements violate the Bell inequality, this implies the existence of nonlocality, a non-classical property of physical systems that is unique to quantum mechanics and has a key role in understanding why classical mechanics fails in the explanation of such systems.

In this chapter we explain what these two paradigms, classical and quantum mechanics, are, what a Bell Theorem is, how Bell experiments are designed after Bell inequalities, how the outcomes of Bell experiments can result in violations of the Bell inequality, and other key concepts such as nonlocality and contextuality and different interpretations of their existence. First, we give a verbal explanation of a Bell experiment and why it is constructed that way, then we give more precise mathematical explanations.

1.1 A Verbal Description of Bell’s Theorem

1.1.1 The Superposition State

From the point of view of our work, the main difference between classical and quantum mechanics is that in classical mechanics, all quantities of a system have certain, precise values (even if we do not know what they are), namely all variables representing a quantity are assigned real numbers (with some units). In contrast, in quantum mechanics (a framework that is more general), there can be incompatible quantities whose variables cannot be assigned single real numbers simultaneously (a feature that is really unintuitive).

If two quantities are incompatible, then the value of at least one of them has to have a non-sharp probability distribution (which corresponds to a state that cannot be explained by a single real number). The probability of an event is the tendency of its occurrence, which is expressed as a number between zero and one. The probability of a value is the tendency of getting that value when the related quantity is measured (namely, the occurrence tendency of the event that “that value is the outcome of the measurement”). The probability distribution of a quantity is the set of all probabilities corresponding to each possible outcome value. A sharp probability distribution means that the tendency of only one possible outcome is one, that is, the measurement outcome will certainly be that value. A non-sharp probability distribution means at least two of the possible outcomes have non-zero probabilities; the outcome value is not certain. Later, we will be explain in more detail what it means to have non-sharp probability distributions for certain quantities in an operational setting (via experimental procedures of measurements of the quantities) and also in comparison
between Kolmogorovian (classical) probability theory and quantum probability theory.

A Bell theorem pertains to physical systems for which whether or not their quantities are incompatible leads to measurable differences. If quantities are compatible (as classical mechanics requires) and hence can have arbitrarily sharp distributions (a sharper distribution has fewer outcomes with non-zero probabilities), then there will be certain bounds (expressed as inequalities that we call Bell inequalities) on the results of Bell experiments, whereas the quantum bounds on the same experiments are higher (thanks to the fact that in quantum mechanics quantities are allowed to be incompatible).

Bell experiments have direct implications for the fundamental nature of quantities in general. Doing a Bell experiment and getting a result that is higher than the classical bound tells us that some of the quantities we measured were incompatible. In this sense, according to Shimony, the Bell theorem is applied-metaphysics [17].

If one believes that the superposition is actually a way of being, then the Bell theorem is applied-ontology. However, if one believes that the superposition state sets the limit on what can be known about physical systems, then Bell theorem is, according to Cavalcanti, applied-epistemology [11].

Superposition is one of the idiosyncratic features of quantum mechanics. When a system is in a superposition state of “the quantity $X$ has the value $x_1$” and “$X$ has the value $x_2$” (in short, a superposition of “$X = x_1$” and “$X = x_2$”), the meaning of this state in English is “when the quantity $X$ of the system is measured, the outcome can be either $x_1$ or $x_2$”. What was the value of $X$ before the measurement? $X$ had no definite value but it was simply in a “superposition of $x_1$ and $x_2$”, which is different than “$X$ was $x_1$”, “$X$ was $x_2$” and “$X$ was both $x_1$ and $x_2$”. Superposition is neither of these options.

The definition of superposition is directly linked to the meaning of measurement. Measurement breaks down the superposition (so-called “collapse of the wavefunction”) and leaves the quantity in a classical state of one of the possible outcomes. After measurement, the state of $X$ is either $X = x_1$ or $X = x_2$. Also, the measurement process happens without any intervention of a conscious observer.

The physical model given by the axioms of quantum mechanics is probabilistic in nature. We say that the outcomes are sampled from a probability distribution which is determined by their state. Therefore, superposition leads to an indeterministic connection between the initial state and the measurement outcomes.

This, by itself, is not a radical deviation from the classical model. There are probabilistic frameworks based on classical mechanics, such as statistical mechanics, in which one can envisage ensembles of identical systems with certain distributions. These probabilistic models can describe deterministic systems with much higher efficiency and success than deterministic models of the same system.

For example, instead of dealing with the positions and momentums of each individual molecule in a gas (in a macroscopic system the number of parameters will be at the order of $10^{23}$) and solving Newton’s equation of
motion for all particles, statistical mechanics uses a statistical model that deals with possible configurations of these molecules (a configuration is a choice of position and momentum values for all molecules). This model does not take into account how the transition between these possible configurations happen mechanically. Instead, it derives the possibilities of macroscopically distinct configurations.

Therefore, the main distinction between classical and quantum models is not about determinism vs. indeterminism. Also, historically, in the earlier era of quantum mechanics, before physicists were able to manipulate individual systems, they were not sure whether the mathematical formulation of quantum mechanics that gives statistical predictions was about individual particles or ensembles of particles, as in statistical mechanics [3][22]. With present-day advances in technology, we can manipulate single degrees of freedom [27] and it has become evident that quantum states are associated with individual physical systems and not their ensembles.

1.1.2 Superposition State When Dealing with a Single Quantity

When we are dealing with only one quantity, then classical and quantum mechanics are equivalent in terms of the set of all possible probability distributions that can be generated by the chosen mechanics. A probability distribution is a set of numbers associated with the occurrence tendencies of some measurement outcomes that add up to one. “\( p(X = x_i) \)” means the probability of getting the outcome \( x_i \) when the quantity \( X \) is measured. Then the set \( P = \{ p(X = x_1), p(X = x_2) \} \) is a probability distribution. For example, \( P = \{0, 1\} \) defines a system in which \( X = x_2 \) with certainty and \( P = \{1/2, 1/2\} \) describes a fair coin toss system, etc.

The set of all possible sets \( P \) allowed by quantum mechanics, \( P_Q \), is the same as the set of all possible sets \( P \) allowed by classical mechanics, \( P_C \), which are \( \{0, 1\}, \{1, 0\} \) and all possibilities in between them that are in the form \( \{t, 1-t\} \) where \( t \) can be any number between 0 and 1.

To measure a probability, say the probability of getting the outcome \( x_1 \) after measuring \( X \), the number \( p(X = x_1) \), we measure the quantity \( X \) many times. Then, we divide the number measurements that result with the output \( x_1 \) by the total number of measurements. This will give a number between zero and one. (A ratio of one means the measurement outcome will be \( x_1 \) for certain. A ratio of zero means \( x_1 \) never occurs. Any value in between will be a non-certain probability.) The more measurements we make, the better the frequency of occurrence of \( x_1 \) will be an approximation of the tendency of occurrence. (Statisticians have methods to tell us with how many data points how certain we can be about our ratio.)
1.1.3 Superposition State When Dealing with Two Quantities, Sharpness of a Probability Distribution, Uncertainty Principle

The first real difference between classical and quantum mechanics becomes apparent when we are dealing with two quantities (in general, more than one quantity) at the same time. Quantum mechanics introduces a novel relationship type between quantities. In quantum mechanics, two quantities can be compatible or incompatible. In classical mechanics, all quantities are compatible. Incompatibility means that there is no state in which both quantities have definite values that can be expressed by real numbers, in other words, no state allows both quantities to have sharp distributions.

A sharp distribution means that the probability of one measurement outcome is 1, which is to say certain, and the probabilities of the rest of the outcomes are 0. Say $Z$ is a quantity that can have 3 different values $z_1$, $z_2$, and $z_3$ (a coin or die with three sides). The probability distribution associated with the outcomes of $Z$-measurements is $P_Z = \{p(Z = z_1), p(Z = z_2), p(Z = z_3)\}$. Among all possible distributions, only three of them are sharp: $P = \{1, 0, 0\}$, $\{0, 1, 0\}$ and $\{0, 0, 1\}$. Any other distributions, such as $\{1/4, 1/4, 1/2\}$, $\{1/2, 0, 1/2\}$, etc., are non-sharp distributions; they have a non-zero occurrence possibility for more than one outcome.

See Figure 1 for a visual description of sharpness for a quantity that can take only 5 different values, and Figure 2 for a quantity that can have infinitely many different outcomes.

![Figure 1: Sharp and unsharp discrete distributions. The height of each bar indicates the probability of the quantity $Z$ having the value under it. For each diagram, the sum of the heights of all bars is 1.](image)

A measurement of a quantity makes it certain that the quantity has one of its possible values, which is represented by the corresponding sharp distribution. For example, a system in the state corresponding to the outcome distribution $\{1/4, 1/4, 1/2\}$ will have the outcome distribution $\{0, 0, 1\}$ with the probability of $1/2$. 
The probability distributions of incompatible quantities are connected in such a way that manipulating the state of the system to change the distribution of one quantity affects the distribution of the other quantity. This property of connectedness of incompatible quantities manifests itself in terms of uncertainty relationships. The general name for all uncertainty relations is “the uncertainty principle”.

The most famous examples of incompatible quantities are 1) the position of a particle in a direction and the velocity of the same particle in the same direction, 2) the spin of a particle in one direction and the spin of the same particle in an orthogonal direction, and 3) the direction of the linear polarization of a photon and the direction of the circular polarization of the same photon.

A measure of the unsharpness of a distribution is called variance. For example, $P = \{0, 1, 0, 0\}$ is sharp (zero variance), $\{0, \frac{1}{2}, \frac{1}{2}, 0\}$ is unsharp (some variance) and $\{\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\}$ is more unsharp (higher variance). See Figure 3 for a visual description of a quantity with infinitely many different possible outcomes.

The uncertainty principle says that when two quantities are incompatible, there is no quantum state
for which $\Delta X = \Delta Y = 0$. For all possible physical states there is a lower bound for the products of their variances $\Delta X \Delta Y \geq$ some minimum limit.

Figure 4: A demonstration of the uncertainty principle for two discrete quantities. Here we have two quantities: 1) the magnitude of spin in $x$-direction, $\sigma_x$, and 2) the magnitude of spin in $z$-direction, $\sigma_z$. Each choice of the parameter $\theta$ corresponds to a different physical state. According to the plot, there is no physical state in which the variances of both quantities are zero for all choices of the parameter. On the contrary, when one of them is sharp (say, for $\theta = \pi/4$ the variance of $\sigma_z$ is 0), the other one is a flat distribution and hence has the highest value of variance (for the same parameter the variance of $\sigma_x$ is 1.)

(Here we do a non-exhaustive search in the space of possible states by varying the parameter $\theta$ for the class of states of the form $|\psi(\theta)\rangle = \cos \theta |0\rangle + \sin \theta |1\rangle$)

For example, according to the experiments, the value of the spin of a spin-$\frac{1}{2}$ particle in a chosen direction can be “up” or “down”. If the particle is put in a state in which its spin in $z$-direction has the value “up”, namely the distribution for $z$-spin $P_Z = \{p(z\text{-spin}=\text{up}), p(z\text{-spin}=\text{down})\}$ is $P_Z = \{1, 0\}$. Then the distribution for the spin in $x$ direction $P_X = \{p(x\text{-spin}=\text{up}), p(x\text{-spin}=\text{down})\}$ becomes $P_X = \{\frac{1}{2}, \frac{1}{2}\}$. There is no physical state for which these two distributions are simultaneously sharp, i.e. $P_Z = \{1, 0\}$ and $P_X = \{0, 1\}$, and there is a limit on how small the product of their variances can be.

At first, quantum mechanics seems more restrictive than classical mechanics. But actually, the quantum model allows compatibility where distributions related to two different quantities are independent, just like classical mechanics, and in addition to that, it introduces the concept of dependent distributions of different quantities. For example, the position of a particle in a direction is compatible with its velocity in a perpendicular direction. The distributions related to those quantities can be made arbitrarily sharp simultaneously, as can be done in classical mechanics.

However, when it comes to incompatible quantities whose incompatibility is experimental fact observed from nature, it is classical mechanics that cannot represent what is actually happening. Classical mechanics can only approximate the relationship between incompatible quantities via compatible probability distributions. The Bell theorem is about finding where classical mechanics fails.
Just as in the case of a single quantity, in the case of two compatible variables quantum mechanics and classical mechanics are equivalent in terms of the possible probability distributions that can be generated. When we investigate a case of more than one quantity, we talk about the joint occurrences of event and joint probabilities.

For example, when the quantity $X$ and the quantity $Y$ are measured simultaneously, $X$ has the value $x_1$ and $Y$ has the value $y_1$, “$X = x_1$ and $Y = y_1$”. There are three other possible outcomes, “$X = x_1$ and $Y = y_2$”, “$X = x_2$ and $Y = y_1$”, “$X = x_2$ and $Y = y_2$”. The outcome of the joint measurement will be one out of these four joint outcomes. Now we can talk about a probability distribution with four elements $P_{XY} = \{p(X = x_1, Y = y_1), p(X = x_1, Y = y_2), p(X = x_2, Y = y_1), p(X = x_2, Y = y_2)\}$ where four elements should add up to 1. In short, express $P_{XY}$ as $\{p_1, p_2, p_3, p_4\}$. $\mathcal{P}_C$ is the set of all possible $P_{XY}$ where $0 \leq p_1, p_2, p_3, p_4 \leq 1$ and $p_1 + p_2 + p_3 + p_4 = 1$. See Figure 5 for a visual description of a joint probability distribution.

$$P(X,Y)$$

![Figure 5: Visualization of a discrete probability distribution with 2 arguments, $P(X,Y)$. 3D bars correspond to (from short to tall) $P(X = x_1, Y = y_1)$, $P(X = x_2, Y = y_1)$, $P(X = x_1, Y = y_2)$ and $P(X = x_2, Y = y_2)$. Marginals $P(X)$ and $P(Y)$ are shown using darker 2D bars at the axes, $P(X)$ on x-axis with $P(X = x_1) = P(X = x_1, Y = y_1) + P(X = x_1, Y = y_2)$ and $P(X = x_2) = P(X = x_2, Y = y_1) + P(X = x_2, Y = y_2)$. And similarly for $P(Y = y_1)$ and $P(Y = y_2)$.](image-url)

In quantum mechanics, the system can be in a superposition state of some of these joint outcomes. For example, the system can be in a superposition of “$X = x_1, Y = y_1$” and “$X = x_2, Y = y_2$”. This means that when both quantities are measured, either the outcome of $X$ will be $x_1$ and the outcome of $Y$ will be $y_1$,
or the outcome of $X$ will be $x_2$ and the outcome of $Y$ will be $y_2$. This also means that if only one of the quantities is measured, say $Y$, it will have the value of either $y_1$ or $y_2$. If it has the value $y_1$, then $X$ has the value $x_1$, and if it has the value $y_2$, then $X$ has the value $x_2$.

When $X$ and $Y$ are compatible, the set of all possible probability distributions $P_{XY}$, in classical mechanics $P_C$ and in quantum mechanics $P_Q$ are the same. Any distribution that can be produced by quantum mechanics can be produced by classical mechanics too (and vice versa). $P_C = P_Q$.

On the other hand, if $X$ and $Y$ are incompatible, then the elements $p(X = x_i, Y = y_j)$ do not have a meaning in quantum mechanics. For example, the probability of $X$ having the value $x_1$ and $Y$ having the value $y_2$ corresponds with the joint event of $X = x_1$ and $Y = y_2$. As we said before, after the measurement the system will be in one of the sharp distribution states. For our example this will be a state for which $p(X = x_1) = 1$ and $p(Y = y_2) = 1$. These are both sharp distributions for both quantities, which is not allowed by quantum mechanics. Therefore, joint probabilities of joint outcome events of incompatible quantities are meaningless in quantum mechanics.

Even though we know that there is a fundamental difference between quantum mechanics and classical mechanics when quantities are incompatible, we do not have a way of comparing the two models if we only have two incompatible quantities, because classical mechanics cannot model incompatible quantities and quantum mechanics has no joint probability distributions of them. The equality of $P_C = P_Q$ does not take the existence of incompatible quantities into account.

To achieve a mean of comparison we need a more complicated setup.

1.1.4 Two Spatially Separated Systems

Two quantities that belong to different physical systems that are spatially separated always commute. Say $X$ and $Y$ are two quantities so that $X$ of a particle and $Y$ of the same particle are incompatible. However, when $X$ is measured on one particle and $Y$ is measured on another (identical) particle that has a different position, then they are compatible.

Call the property of $X$ of particle-1 $X_1$, the property $X$ of particle-2 $X_2$, the property $Y$ of particle-1 $Y_1$ and the property $Y$ of particle-2 $Y_2$. $X_1$ and $X_2$ can have values $x_1$ and $x_2$. And $Y_1$ and $Y_2$ can have the values $y_1$ and $y_2$.

In this case the joint occurrences of $X_1$ having a value and $Y_2$ having a value are meaningful in both classical and quantum mechanics because $X_1$ and $Y_2$ are compatible. This allows us to talk about joint probability distributions in the form of $p(X_m = x_i, Y_n = y_j)$ where $m, n, i, j$ can be 1 or 2.

This scenario involves four probability distributions according to choice of our quantities to measure. We can measure the pairs $(X_1, X_2)$, $(X_1, Y_2)$, $(Y_1, X_2)$ and $(Y_2, X_2)$. We cannot jointly measure $(X_1, Y_1)$ and
(X₂, Y₂) because X₁ and Y₁ are not compatible.

The four probability distributions that correspond to allowable pairs will have elements in these forms:

\[ p(X₁ = x₁, X₂ = x₂), ~ p(X₁ = x₁, Y₂ = y₂), ~ p(Y₁ = y₁, X₂ = x₂) \] and \[ p(Y₁ = y₁, Y₂ = y₂) \] with each distribution having 4 possible outcomes.

Because all quantities belonging to the first particle are compatible with all quantities belonging to the second particle, the joint system of two particles can be in a superposition state over the choice of quantities. Say it can be in the superposition of “Y₁ = y₁, Y₂ = y₁” and “Y₁ = y₂, Y₂ = y₂”. This means, if we measure the quantity Y₁ on particle-1 and get the outcome y₂, then the measurement Y₂ on particle-2 gives the outcome of y₂ too.

In total there are 16 different joint outcomes that can be achieved by measuring four allowed pairs:

“X₁ = x₁, X₂ = x₂”, “X₁ = x₁, X₂ = x₂”, “X₁ = x₁, Y₂ = y₂”, “X₁ = x₂, X₂ = x₁”,

“X₁ = x₂, X₂ = x₂”, “X₁ = x₂, Y₂ = y₂”, “Y₁ = y₁, X₂ = x₁”, “Y₁ = y₁, X₂ = x₂”,

“Y₁ = y₁, Y₂ = y₂”, “Y₁ = y₁, Y₂ = y₂”, “Y₁ = y₂, X₂ = x₁”, “Y₁ = y₂, X₂ = x₂”, “Y₁ = y₂, Y₂ = y₁”,

“Y₁ = y₂, Y₂ = y₂”. We can prepare two particles in a certain state and then measure the probability of occurrence of any of these 16 outcomes corresponding to that state. Both classical and quantum mechanics has methods of prediction of these 16 numbers.

These probabilities corresponding to this 16 possible outcomes are the simplest building blocks to find an experimentally verifiable difference between the classical and quantum mechanics. We never need to jointly measure the incompatible quantities (X₁, Y₁) and (X₂, Y₂), but we are allowed to measure (X₁, X₂) and (X₁, Y₂) where X₂ and Y₂ are incompatible. The incompatibility of X₂ and Y₂ creates an uncertainty relation between the distributions of X₂ and Y₂ that we cannot analyze by jointly measuring them. But their uncertainty relation affects these 16 probabilities that we can are allowed to measure jointly. Hence the analysis is achieved indirectly.

Construct the set which is made of the four probability distributions \( P_{X₁X₂}, ~ P_{X₁Y₂}, ~ P_{Y₁X₂}, ~ P_{Y₁Y₂} \).

\[
P_{X₁X₂} = \{ p(X₁ = x₁, X₂ = x₂), ~ p(X₁ = x₁, X₂ = x₂), ~ p(X₁ = x₁, X₂ = x₁), ~ p(X₁ = x₂, X₂ = x₂) \}.
\]

\[
P_{X₁Y₂} = \{ p(X₁ = x₁, Y₂ = y₁), ~ p(X₁ = x₁, Y₂ = y₂), ~ p(X₁ = x₁, Y₂ = y₂), ~ p(X₁ = x₂, Y₂ = y₁) \}.
\]

\[
P_{Y₁X₂} = \{ p(Y₁ = y₁, X₂ = x₁), ~ p(Y₁ = y₁, X₂ = x₂), ~ p(Y₁ = y₂, X₂ = x₁), ~ p(Y₁ = y₂, X₂ = x₂) \}.
\]

\[
P_{Y₁Y₂} = \{ p(Y₁ = y₁, Y₂ = y₁), ~ p(Y₁ = y₁, Y₂ = y₂), ~ p(Y₁ = y₂, Y₂ = y₁), ~ p(Y₁ = y₂, Y₂ = y₂) \}.
\]

The set that is made of the element of these four distribution sets is

\[ P_{X₁X₂Y₁Y₂} = \{ \ldots p(Y₁ = y₂, Y₂ = y₁), ~ p(Y₁ = y₂, Y₂ = y₂) \} \] Define the set of all possible \( P \) sets as \( \mathcal{P} \). Bell theorem says that \( \mathcal{P}_C \neq \mathcal{P}_Q \). More precisely, quantum set is a bigger set than classical one \( \mathcal{P}_Q \supset \mathcal{P}_C \). There are combinations of these 16 numbers that can be produced by quantum mechanics which cannot be produced by classical mechanics.
To conduct a Bell experiment in this 2 particles, 2 quantities per particle, 2 outcomes per quantity scenario, first, we take two particles. Then we choose two incompatible properties for each particle. We call them $X_1$ and $Y_1$ for particle-1 and $X_2$ and $Y_2$ for particle-2. The incompatibility of $X$’s and $Y$’s is necessary. Then we prepare the combined state of this two-particle system so that when these 16 joint probabilities are measured we get a set of numbers, $P$, that can not be produced by classical mechanics. In other words $P \notin \mathcal{P}_C$ but $P \in \mathcal{P}_Q$. The type of quantum states that give non-classical distributions like this are called “entangled” states.

1.1.5 A Bell Experiment Example, CHSH Experiment

As a concrete example of a Bell experiment we can look at the famous experiment proposal by Clauser-Horne-Shimony-Holt (CHSH) [13]. The quantities they measure are the orientation of polarizations of photons. For two photons that are sent in opposite directions, their linear polarizations in chosen directions (which are determined by chosen angles) are measured. $X_1$ is the polarization of photon 1 in $22.5^\circ$, $Y_1$ is the polarization of photon 2 in $67.5^\circ$, $X_2$ is the polarization of photon 1 in $157.5^\circ$, and $Y_2$ is the polarization of photon 2 in $202.5^\circ$. See Figure 6 for a diagram of CHSH setup.

For a choice of an angle for a photon, a polarizing filter is put on the path of that photon, and a photon detector behind the polarizing filter. If the detector clicks, it means the photon passed the filter, hence the outcome $x_1$ (or $y_1$) is occurred. If the detector does not click, it means the photon did not pass the filter, hence the outcome $x_2$ (or $y_2$) is occurred.

![Figure 6: Illustration of CHSH setup. Two entangled photons go in opposite, $+z$ and $-z$, directions. For each photon 2 possible directions are chosen to measure the polarization. For example, for the photon in $+z$ direction, its polarization can be either measured in $Y_1$ direction or $Y_2$ direction. The choice of measurement can be made even after the photons left their source.](image)

Then they send entangled photons towards detectors and measure the probabilities of joint occurrences.

A way of deciding whether the 16 numbers measured in the experiment is in $\mathcal{P}_C$ or not is to calculate a combination of them. Quantum mechanics and classical mechanics will have different bounds for some
Combinations. For example the combination that CHSH chose is this:

\[ B = p(X_1 = x_1, X_2 = x_1) - p(X_1 = x_1, X_2 = x_2) - p(X_1 = x_2, X_2 = x_1) + p(X_1 = x_2, X_2 = x_2) \]

\[ + p(X_1 = x_1, Y_2 = y_1) - p(X_1 = x_1, Y_2 = y_2) - p(X_1 = x_2, Y_2 = y_1) + p(X_1 = x_2, Y_2 = y_2) \]

\[ - p(Y_1 = y_1, X_2 = x_1) + p(Y_1 = y_1, X_2 = x_2) + p(Y_1 = y_2, X_2 = x_1) - p(Y_1 = y_2, X_2 = x_2) \]

\[ + p(Y_1 = y_1, Y_2 = y_1) - p(Y_1 = y_1, Y_2 = y_2) - p(Y_1 = y_2, Y_2 = y_1) + p(Y_1 = y_2, Y_2 = y_2) \]

\[ B \] is a certain combination of 16 numbers where some of them are added and others subtracted. Basically each line corresponds to the correlation between chosen quantities, \( E(X_i, Y_j) \), when +1 is assigned to detector click, and −1 is assigned to no click, namely \( x_1 = 1 \) and \( x_2 = -1 \). In other words \( B = E(X_1, X_2) + E(X_1, Y_2) - E(Y_1, X_2) + E(Y_1, Y_2) \).

Classical mechanics puts a bound on the possible values that \( B \) can have which can be expressed as an inequality.

\[-2 \leq B \leq 2\]

These inequalities are called “Bell Inequalities”. Any measurement of \( B \) in the laboratory that gives a number bigger than 2 is called a “Bell violation”. The maximum value for \( B \) allowed by quantum mechanics is \( 2\sqrt{2} \approx 2.83 \). Values \( B > 2 \) are observed in experiments.

This violation concludes that classical mechanics and quantum mechanics are fundamentally different in terms of the probabilities that they can produce. Therefore we have to take the idea of superposition seriously. Superposition is unavoidable in the presence of incompatible quantities which is manifested in uncertainty relations and in the inevitability of probabilistic measurement outcomes, namely indeterminism.

However the essential point of Bell arguments is not refutation of determinism in favor of indeterminism (which happens as a side effect). Indeterministic systems are conceivable in the framework of classical mechanics (such as statistical mechanics) too. One of the main differences is in the nature of how probabilities arise.

In the classical model, when the value of a quantity is probabilistic, when we get different values after measuring the same quantity from identical systems, we think that the quantity had certain definite value before we measure it, but we do not know what that value is. Probabilities are due to our lack of knowledge on the system and measurements reveal pre-existing realities.

In the quantum model, when the value of a quantity is probabilistic due to superposition, it is not the case that the quantity had a certain definite value before the measurement and we did not know it. The
quantity did not have a certain value at all that can be known, in the first place. No pre-existing reality is revealed by the measurement. The measurement is the active process of converting a superposition state to a non-superposition state. This is what Bell’s Theorem says.

1.1.6 Einstein, Podolsky, Rosen Arguments

In 1935 Einstein, long before the discovery of Bell’s theorem, Podolsky and Rosen (EPR) came with an argument that describes the probabilities arise in quantum mechanics due to our lack of knowledge, namely, quantum mechanics is an incomplete theory [19]. According to EPR, quantum mechanics does not consider all possible causes that can affect measurement outcomes, therefore its predictions are probabilistic.

Even though their conclusions in their work is incorrect their chain of reasoning will help us to understand the concept of nonlocality.

Because EPR did not know that some collection of probability distributions that are predicted but quantum mechanics and observed in nature cannot be predicted by classical mechanics, which leads us to change our concept of quantities having values (in their terminology, our concept of reality), they start with an incorrect definition of reality: “If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.”

According to this definition incompatible quantities cannot have simultaneous physical realities. We will call their definition “classical reality”, and superposition states do not correspond to a classical reality.

They define completeness of a physical theory as “every element of the physical reality must have a counterpart in the physical theory”. According to this definition and their assumption that all quantities should have a classical reality quantum mechanics is incomplete, it does not predict all quantities with certainty simultaneously, hence it is not considering some aspects of physical reality. We will call their completeness as “classical completeness” and classical completeness is an ideal that cannot be achieved by physical systems due to uncertainty principles.

After these definitions they propose a Gedanken Experiment where two systems share an entangled quantum state and two incompatible quantities belonging to those particles are considered. The quantity $X$ can be measured on particle-1 and particle-2 ($X_1$ and $X_2$) and also the quantity $Y$ can be measured on particle-1 and particle-2 ($Y_1$ and $Y_2$). This is exactly the same system that is used in the CHSH setup. Interestingly EPR inspired the Bell experiments in their attempt of proving the opposite of Bell’s theorem (namely quantum mechanics is incomplete and its predictions are reproducible by classical mechanics).

In previous section we said that the type of quantum states belonging to the system of two parties that result in Bell violations are called entangled states. Entangled states generate correlations between some
quantities. Namely, even though the measurement results on single particles are still random, the results of both of them tend to have the same (or opposite values) so that it become possible to figure out the outcome of the second particle by knowing the outcome of the first particle (and vice versa).

For example, say $X_1$ and $X_2$ are perfectly correlated. Then $p(X_1 = x_1, X_2 = x_2) = 0$ and $p(X_1 = x_2, X_2 = x_1) = 0$. This means when we measure $X_1$ we either get $x_1$ or $x_2$ randomly. But when we look at the joint measurements of $X_1$ and $X_2$ we never get outcomes in which $X_1$ and $X_2$ have different values. Therefore we can infer the outcome of $X_2$ when we know the outcome of $X_1$. If $X_1$ is $x_1$ then $X_2$ is $x_1$ too. And if $X_1$ is $x_2$, then $X_2$ is $x_2$ too.

This is just one type of perfect correlation. There can be other types of correlations, such as anti-correlation (they always have the opposite values, and never have the same value) etc. The essential point of a perfect correlation is is that it makes it possible to perfectly infer the outcome of the other party by knowing the outcome of one party.

EPR uses an entangled state in which both positions and velocities of two particles are perfectly correlated. For example, if the velocity of the first particle is $V_1 = v$, the velocity of the second particle is $V_2 = -v$. And if the position of the first particle is $X_1 = x$, then the velocity of the second particle is $X_2 = x + \Delta x$. Even though this quantum state is so hard to implement and will not last long, theoretically it is possible to produce these perfect correlations.

Using this setup EPR tries to prove that incompatible quantities have classical realities. They accept that simultaneous measurements of incompatible quantities cannot be done on a single particle at the same time, because measuring one quantity affects the distribution of the second one. But they claim that one can measure a quantity on particle-2 indirectly, and without affecting particle-2’s state by measuring the same quantity on particle-1 and inferring the value on particle-2 using the perfect correlation relation due to entanglement. And because the particle-2 is not disturbed the inferred values are classical realities.

The broken part of this chain of reasoning is that measuring a quantity on particle-1 actually affects the state of particle-2 by a process called “collapse of the wave function”. Though, this effect does not change the probability distributions that will be produced by the particle-2. In other words, whether we measure $X_1$ or $Y_1$, the probabilities $p(X_2 = x_1), p(X_2 = x_2), p(Y_2 = y_1), p(Y_2 = y_2)$ do not change. The effect of the choice of measurement on particle-1 is not observable by only measuring the quantities on particle-2. (To observe the effect we need joint measurement outcomes from both particles. Namely, the 16 numbers we mentioned in the previous chapter).

The fact this effect can not be measured locally at the other particle is of high importance. Say if choosing $X_1$ to measure instead $Y_1$ on particle-1 affects the value of $p(X_2 = x_1)$. Say if $X_1$ is measured $p(X_2 = x_1) = 0.51$ and $p(X_2 = x_1) = 0.49$. Then an experimenter at the second measurement site could
measure $p(X_2 = x_1)$, the occurrence frequency of getting the outcome $x_1$ for $X_2$ measurements, and can tell whether the experimenter at the first measurement site chose to measure $X_1$ or $Y_1$.

This method could be used in sending messages in a way such as Morse code, say $X_1$ is dot and $Y_1$ is dash. These two particles can be arbitrarily far away from each other at the moment of measurements. Their distance can even be several light years. But still the indirect inference would allow them to communicate faster than light, or even instantly. This would be a direct violation of special relativity and would result inconsistencies. However, quantum mechanics do not allow the effect of the choice of measurement on one particle is not locally observable at the second particle, this effect and hence quantum mechanics is consistent with special relativity and does not allow superluminal communication.

Returning to EPR reasoning, measuring $X_1$ forces $X_2$ to have a classical reality, but $Y_2$ will be in superposition. Likewise, measuring $Y_1$ forces $Y_2$ to have a classical reality, but $X_2$ will be in a superposition. EPR agrees that by measuring them indirectly, $X_2$ and $Y_2$ cannot have simultaneous classical reality. “On this point of view, since either one or the other, but not both simultaneously, of the quantities $X_2$ and $Y_2$ can be predicted, they are not simultaneously real.” But EPR still insist on that $X_2$ and $Y_2$ share the same reality, even though after different choices for the measurement on particle-1 the quantum state of particle-2 will be different: “One would not arrive at our conclusion if one insisted that two or more physical quantities can be regarded as simultaneous elements of reality only when they can be simultaneously measured or predicted. This makes the reality of $X_2$ and $Y_2$ depend upon the process of measurement carried out on the first system in any way. No reasonable definition of reality could be expected to permit this.” Unfortunate for them, this “unreasonable” process is how reality works.

It took 30 years after EPR paper for physicists to figure out and implement an experiment to decide on different models of reality.
1.1.7 Visual summary of minimum necessary Bell setup

Figure 7: In this diagram each elliptical set represents a physical system. Dots represent quantities that can be measured on the system it belongs to. Solid connecting lines imply the compatibility of the quantities they connect, hence they can have sharp distributions simultaneously. Dashed lines mean that connected quantities are incompatible and hence cannot have both sharp distributions.

$P_C$ is the set of probabilities or joint probabilities that can be produced by the combined system. $P_Q$ is the one allowed by classical mechanics and $P_Q$ by quantum mechanics.

First diagram on the top of Figure 7 indicates that quantum and classical mechanics allow the same set of distributions for the measurements done on a single quantity. (Basically $0 \leq P(X) \leq 1$.)

Second diagram on the top is about the set of joint probabilities of joint measurement outcomes where two quantities are compatible and are on the same party. (Again quantum and classical sets are identical. $\sum_{x,y} P(X,Y) = 1$ and $0 \leq P(X,Y) \leq 1$.)

Third one from the top means that we can not have joint probability distributions of incompatible quantities. Because the event corresponding to the probability $P(X = x, Y = y)$, which is $E : X = x \& Y = y$, can not happen. $X$ and $Y$ can not have certain, sharp values simultaneously. Therefore $P_Q$ do not exist in this case, and we can not do a comparison.

Last one on top row is the case where two quantities of joint investigation are distributed over spatially distant parties. In this case any two quantities are compatible (even tough it could be the case that they
would be incompatible if they were on the same system) hence we get the same case as compatible observables.

The first diagram at the bottom row means that because local quantities on spatially distant parties always commute, this case is impossible to happen.

The middle one at the bottom row is the first setup where we have a difference between quantum and classical sets. In order to have incompatible observers (to introduce the quantum novelty) and joint probability distributions (to measure correlations, make data collection possible) at the same time, here we choose two incompatible quantities on each party, \( a_1, a_2 \) on \( A \), and \( b_1, b_2 \) on \( B \).

We, then, choose to measure one quantity from each party. All 4 possible choices \((a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\) are compatible pairs which allows us to measure their joint probability distributions.

The union of these 4 distributions is still affected by the incompatibility of \((a_1, a_2)\) and \((b_1, b_2)\). Therefore this is the simplest scenario where we can observe a difference in the set of all possible probabilities that can be assigned to 16 joint outcomes.

CHSH inequalities tell whether a given point is in the classical set or not for this setup.

The last diagram is an example of increased number of quantities per party. (We could have increased the number of parties too.) Any scenario that has equal or larger parameters than \((2, 2, 2)\) allows points in \( P_Q \) that are not in \( P_C \).

\[ P_Q \supset P_C \]

is what we mean by Bell’s theorem.

1.1.8 Quantum nonlocality and indeterminism

The common terminology for “classical realism” is “local realism”, and the violation of local realism by quantum mechanics is called the “quantum nonlocality”.

We showed that the essential point of Bell theorem is not the refutation of determinism in favor of indeterminism, but the refutation of local realism in favor of nonlocality. The physical systems that has nonlocality have correlations between its subsystems that cannot be explained or reproduced by local realistic models, namely classical physics.

Thanks to the indeterminism the nonlocal correlations cannot be used to send signals faster than light. Therefore we cannot have a causal relationship between the measurement choice on one particle and the measurement outcome on another particle that is arbitrarily distant. Even though nonlocality and causality seems to be irreconcilable at first sight, it is the inherent randomness of quantum mechanics, Popescu says, that reconciles nonlocality with relativity/causality [35].
In a deterministic system in which probabilities arise due to lack of the knowledge on the exact state of the physical systems, nonlocality would violate causality. If one starts with nonlocality and relativity as given axioms, then the indeterminism arise as a derived concept [35].

1.1.9 Disclaimer

By the time this dissertation is being written there is no universal consensus on which interpretation of quantum mechanics to accept, in other words the so-called measurement problem is not solved, yet. We do not know the exact mechanisms of how classical physics emerge as a limiting case of quantum mechanics.

The nonlocal effect that changes the quantum state of one particle according to the measurement done on another particle that is entangled to the first one is a consequence of the collapse of the wave function, namely, the quantum measurement process. Further enlightenment on quantum measurement will clarify the meaning of quantum nonlocality and Bell violations.

1.2 Mathematical Description of Bell’s Theorem

Previous section was written with a general audience in mind. Hence heavy wording was used rather than mathematical rigor and conciseness. In this section we will clarify the concepts.

By themselves Bell inequalities have nothing to do with quantum mechanics. They express certain limits on some correlations, which we call Bell expressions, between some random variables under some assumptions. (Bold terms will be explained).

1.2.1 Random Variables

A random variable is a type of variable which, when ‘sampled’, gives a value from a set of possible values. The occurrence tendency of a value is determined by its probability distribution. Say $X$ can have a value from the set of all possible values for $X$, $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$. $X \in \mathcal{X}$. The probability of $X$ having the value $x_n$ is shown as $p(X = x_n)$, the set of all probabilities $\{p(X = x_n) | n = 1..N\}$ is its probability distribution. $\sum_{n=1}^{N} p(X = x_n) = 1$ is the normalization condition that all probability distributions must obey.

Random variables can be used to model quantities of which measurement outcomes are not fixed, or cannot be predicted with certainty. Measuring a quantity in real life corresponds to sampling a random variable in a probabilistic model.

A random variable is a general case of a deterministic variable. A deterministic variable can be thought as a random variable that has the probability 1 for a certain value and 0 for the rest.
\[
p(X = x_n) = \begin{cases} 
1 & n = i \\
0 & n \neq i 
\end{cases}
\]
for \( n = 1..N \). In short \( X = x_i \).

A correlation is, in general, a type of statistical relationship about the tendencies of joint occurrences of different outcomes of random variables.

Say we have two random variables \( X \in \mathcal{X} = \{x_n|n = 1..N\} \) and \( Y \in \mathcal{Y} = \{y_n|n = 1..N\} \) and we sample them simultaneously. We choose one outcome for \( X \), \( x_i \), and one outcome for \( Y \), \( y_j \), and investigate their outcome probabilities \( p(X = x_i) \equiv p_1 \) and \( p(Y = y_j) \equiv p_2 \). Then the probability of the event “\( X \) having any value other than \( x_i \)” is

\[
p(X \neq x_i) = \sum_{n \in \{x_1..x_N\}\setminus\{x_i\}} p(X = x_n) \\
\sum_{n \in \{x_1..x_N\}} p(X = x_n) - p(X = x_i) \\
= 1 - p(X = x_i) \\
= 1 - p_1 \\
\equiv p_\bar{1}
\]

and similarly the probability of \( Y \) having any value other than \( y_j \) is \( p(Y \neq y_j) = 1 - p_2 \equiv p_\bar{2} \).

For now, the only constraints on these probabilities are

\[
p_1 + p_\bar{1} = 1 \\
p_2 + p_\bar{2} = 1
\]

and

\[
0 \leq p_1, p_\bar{1}, p_2, p_\bar{2} \leq 1.
\]

### 1.2.2 Joint Events

Correlation and statistical dependence are about joint events. The probability of the joint event “\( X \) having the value \( x_i \) and \( Y \) having the value \( y_j \)” can be shown as \( p(X = x_i, Y = y_j) \).

We can also talk about “\( X \) having the value \( x_i \) and \( Y \) not having the value \( y_j \)” \( p(X = x_i, Y \neq y_j) \), “\( X \)
There are several constraints on these joint probabilities. The normalization condition is

\[ p(X = x_i, Y = y_j) + p(X = x_i, Y \neq y_j) + p(X \neq x_i, Y = y_j) + p(X \neq x_i, Y \neq y_j) = 1. \]

And the marginal relations are

\[ p(X = x_i) = p(X = x_i, Y = y_j) + p(X = x_i, Y \neq y_j) \]
\[ p(X \neq x_i) = p(X \neq x_i, Y = y_j) + p(X \neq x_i, Y \neq y_j) \]
\[ p(Y = y_j) = p(X = x_i, Y = y_j) + p(X \neq x_i, Y = y_j) \]
\[ p(Y \neq y_j) = p(X = x_i, Y \neq y_j) + p(X \neq x_i, Y \neq y_j). \]

where the probabilities of single events are expressed as combinations of joint events.

We can describe these relations with the Table 1 where the columns add up to individual event probabilities at the header row, and the rows add up to the individual event probabilities at the header column.

<table>
<thead>
<tr>
<th>( ) ( X )</th>
<th>( p(X = x_i) )</th>
<th>( p(X \neq x_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = y_j )</td>
<td>( p(X = x_i, Y = y_j) )</td>
<td>( p(X \neq x_i, Y = y_j) )</td>
</tr>
<tr>
<td>( Y \neq y_j )</td>
<td>( p(X = x_i, Y \neq y_j) )</td>
<td>( p(X \neq x_i, Y \neq y_j) )</td>
</tr>
</tbody>
</table>

Table 1: Occurrence probabilities of two events
\( X = x_m \) given \( Y = y_n \) already happened is shown as

\[ p(X = x_m | Y = y_n). \]

The conditional probability expressed in terms of the joint probability and the individual probability of occurred event is

\[ p(E | F) = \frac{p(E \& F)}{p(F)} \]

which is also called the Bayes’ theorem.

The relation in opposite direction is

\[ p(F | E) = \frac{p(E \& F)}{p(E)} \]

and from both of them we see that

\[ p(E | F) p(F) = p(E \& F) = p(F | E) p(E). \]

### 1.2.4 Statistical Dependence

Two random variables \( X \) and \( Y \) (or two random events) are statistically independent if they satisfy

\[ p(X = x_m, Y = y_n) = p(X = x_m) p(Y = y_n) \]

for all \( m, n \). It is shown as

\[ X \perp \perp Y \]

For example two successive rolls of dice are statistically independent events. Therefore, if \( p(D_1 = i) \) and \( p(D_2 = j) \) mean that the probability of getting the value \( i \) from the first roll and \( j \) from the second roll, successively, where \( i, j \in \{1, 2, 3, 4, 5, 6\} \), then

\[ p(D_1 = i, D_2 = j) = p(D_1 = i) p(D_2 = j) \]

\[ \frac{1}{36} = \frac{1}{6} \times \frac{1}{6}. \]
To verify their independence we can first measure \( p(D_1 = i) \) for all \( i \) values for the first dice (and find \( \frac{1}{6} \) for each \( i \), if the dice is fair, and a different distribution if not), then measure \( p(D_2 = j) \) for all \( j \). Then throw them together and measure \( p(D_1 = i, D_2 = j) \). If the expression of statistical independence holds for all \( i \) and \( j \) then these random events, hence the random variables that represent them are statistically independent.

There is no unique measure of statistical dependence though. For sure, if \( p(X = x_m, Y = y_n) \neq p(X = x_m) p(Y = y_n) \) then \( X \) and \( Y \) are statistically dependent. And we can define \( S \) as

\[
S = p(X = x_i, Y = y_j) - p(X = x_i) p(Y = y_j) = p_{12} - p_1 p_2.
\]

Note that here we choose specific outcome for \( X \) and \( Y \), the events of \( X \) being \( x_i \), \( E \), and \( Y \) being \( y_j \), \( F \).

- \( S = 0 \) means that \( E \) and \( F \) are independent from each other. The occurrence of one has no relationship to the occurrence of other. Knowing that \( E \) occurred will not improve our prediction of the probability of \( F \).

- \( S > 0 \) means that \( p(E \& F) > p(E) p(F) \), in other words \( p(E \& F) / p(F) = p(E|F) > p(E) \), which means the probability of “event \( E \) occurring knowing that \( F \) already occurred” is higher than the probability of “event \( E \) occurring” (with no prior knowledge on anything else). Which means the chance of \( E \) happening is higher when \( F \) was happened. Implies a positive statistical dependence.

- \( S < 0 \) means that \( p(E \& F) < p(E) p(F) \), in other words \( p(E \& F) / p(F) = p(E|F) < p(E) \), which means the probability of “event \( E \) occurring knowing that \( F \) already occurred” is lower than the probability of “event \( E \) occurring” (with no prior knowledge on anything else). Which means the chance of \( E \) happening is lower when \( F \) was happened. Implies a negative statistical dependence.

For variables of several outcomes, \( S \) can be different for events corresponding to different outcomes, i.e. \( S(x_i, y_j) = p(X = x_i, Y = y_j) - p(X = x_i) p(Y = y_j) \) and \( S(x_{i'}, y_{j'}) = p(X = x_{i'}, Y = y_{j'}) - p(X = x_{i'}) p(Y = y_{j'}) \).

\[
S(x_i, y_i) \neq S(x_{i'}, y_{j'})
\]

in general.

Another way of expressing \( S \) can be
\[ S = \ p_{12} - p_1 p_2 \]
\[ = \ p_{12} - (p_{12} + p_{1\bar{2}})(p_{1\bar{2}} + p_{1\bar{2}}) \]
\[ = \ p_{12} - (p_{1\bar{2}}^2 + p_{12}p_{1\bar{2}} + p_{12}p_{1\bar{2}} + p_{1\bar{2}}p_{1\bar{2}}) \]
\[ = \ p_{12} (1 - p_{12} - p_{1\bar{2}} - p_{1\bar{2}}) \]
\[ = \ p_{12}p_{1\bar{2}} - p_{1\bar{2}}p_{1\bar{2}} \]

where the final form looks like the determinant of \( \mathcal{P} \), \( p_{12}p_{1\bar{2}} - p_{1\bar{2}}p_{1\bar{2}} = \det(\mathcal{P}) = S. \)

### 1.2.5 Correlation

Given

<table>
<thead>
<tr>
<th>( Y \setminus X )</th>
<th>( p_1 )</th>
<th>( p_{\bar{1}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_2 )</td>
<td>( p_{12} )</td>
<td>( \bar{p}_{1\bar{2}} )</td>
</tr>
<tr>
<td>( p_{\bar{2}} )</td>
<td>( \bar{p}_{12} )</td>
<td>( p_{1\bar{2}} )</td>
</tr>
</tbody>
</table>

Let us define a correlation function between two events as

\[ C = (p_{12} + p_{1\bar{2}}) - (p_{1\bar{2}} + p_{12}). \]

Again this event can be a random variable having a certain outcome, \( p_1 = p(X = x_i) \) and the complementary event can be that random variable having a different outcome, \( \bar{p}_1 = p(X \neq x_i) \). Say \( E \) is \( X = x_i \) and \( F \) is \( Y = y_j \).

Let us analyze different cases of

\[ \mathcal{P} = \begin{pmatrix} p_{12} & p_{1\bar{2}} \\ p_{\bar{1}2} & p_{\bar{1}\bar{2}} \end{pmatrix}. \]

If \( \mathcal{P} = \begin{pmatrix} p_{12} & 0 \\ 0 & p_{1\bar{2}} \end{pmatrix} \) with \( p_{12} + p_{1\bar{2}} = 1 \), then it means that the events either happen at the same time, or they never occur. This will correspond to a \( C = 1 \), and is called a **perfect correlation**.

If \( \mathcal{P} = \begin{pmatrix} 0 & p_{1\bar{2}} \\ p_{1\bar{2}} & 0 \end{pmatrix} \) with \( p_{12} + p_{1\bar{2}} = 1 \), then it means that either \( E \) occurs and \( F \) does not occur, or vice versa. They never occur or do not occur simultaneously. This will correspond to a \( C = -1 \), and is called a **perfect anti-correlation**.

In the case where \( p_{12} + p_{1\bar{2}} = p_{1\bar{2}} - p_{1\bar{2}} = \frac{1}{2} \), \( C \) becomes 0. This case is called **no correlation**.
\[ C = (p_{12} + p_{1\bar{2}}) - (p_{\bar{1}2} + p_{\bar{1}\bar{2}}) \]

means that we want to measure the likelihood of cases when two events happen at the same time \((E \& F)\) or both of them do not occur \((\bar{E} \& \bar{F})\) and unlikelihood of the cases where only one of them occurs and the other not \((E \& \bar{F} \text{ or } \bar{E} \& F)\).

Therefore, unlike the statistical dependence, just having a joint probability distribution is not enough to define a correlation. A correlation needs a choice of similarity between the outcomes of different random variables. Correlations involves actual values of outcomes, not only the labels of outcomes. In this example we associated the \(X = x_i\) outcome of the first quantity with the \(Y = y_j\) outcome of the second quantity.

### 1.2.6 Composite Events

However, the events could have been chosen arbitrarily, and we can associate any event \(E = X \in \mathcal{X}_i\) and \(F = Y \in \mathcal{Y}_j\) where \(\mathcal{X}_i \subset \mathcal{X}\) and \(\mathcal{Y}_j \subset \mathcal{Y}\). \(\mathcal{X}\) and \(\mathcal{Y}\) are the set of possible outcomes of \(X\) and \(Y\), and \(\mathcal{X}_i\) and \(\mathcal{Y}_j\) are any subset of them.

<table>
<thead>
<tr>
<th>(Y \setminus X)</th>
<th>(p(X \in \mathcal{X}_i))</th>
<th>(p(X \notin \mathcal{X}_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p(Y \in \mathcal{Y}_j))</td>
<td>(p(X \in \mathcal{X}_i, Y \in \mathcal{Y}_j))</td>
<td>(p(X \notin \mathcal{X}_i, Y \in \mathcal{Y}_j))</td>
</tr>
<tr>
<td>(p(Y \notin \mathcal{Y}_j))</td>
<td>(p(X \in \mathcal{X}_i, Y \notin \mathcal{Y}_j))</td>
<td>(p(X \notin \mathcal{X}_i, Y \notin \mathcal{Y}_j))</td>
</tr>
</tbody>
</table>

Any correlation in terms of subsets of all possible pairs can be expressed as a combination of joint probabilities of individual outcome pairs

\[
p(X \in \mathcal{X}_i, Y \in \mathcal{Y}_j) = \sum_{x_m \in \mathcal{X}_i} \sum_{y_n \in \mathcal{Y}_j} p(X = x_m, Y = y_n).
\]

### 1.2.7 Generalization to Higher Number of Outcomes

Let us investigate correlations and statistical independence in the case of more than one event (and its complement) per quantity to see how these concepts generalize.

In the previous section we were interested in events \(E\) where \(X \in \mathcal{X}_i\) and \(F\) where \(Y \in \mathcal{Y}_j\) where \(\mathcal{X}_i \subset \mathcal{X}\) and \(\mathcal{Y}_j \subset \mathcal{Y}\). And the complementary events \(\bar{E}\) where \(X \notin \mathcal{X}_i\) and \(\bar{F}\) where \(Y \notin \mathcal{Y}_j\). The corresponding probabilities were

\[
p_1 = p(E) = p(X \in \mathcal{X}_i) = \sum_{x_k \in \mathcal{X}_i} p(x_k)
\]

and similar expressions for \(p_2\) and \(p_{\bar{2}}\).
Now, we will deal with more than 2 events per quantity. Say the event $E_1$ corresponds to $X$ having an outcome from the subset $\mathcal{X}_1 \subset \mathcal{X}$, $E_2 : X \in \mathcal{X}_2$ and $E_3 : X \in \mathcal{X}_3$ where $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \mathcal{X}_3$. Similarly $F_j$ are events for the second quantity $F_j : Y \in \mathcal{Y}_j$ and $\bigcup_j \mathcal{Y}_j$ where $j = 1, 2, 3$.

To make the analysis even simpler without loosing generality, do not use composite events but events corresponding to individual outcomes. $E_1 : X = x_1$, $E_2 : X = x_2$ and $E_3 : X = x_3$ where $\mathcal{X} = \{x_1, x_2, x_3\}$ and similarly $F_j : Y = y_j$ where $\mathcal{Y} = \{y_1, y_2, y_3\}$. Again we can talk about occurrence probabilities of single events and joint events. The table of joint probability distribution is

<table>
<thead>
<tr>
<th>$Y \setminus X$</th>
<th>$p(X = x_1)$</th>
<th>$p(X = x_2)$</th>
<th>$p(X = x_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(Y = y_1)$</td>
<td>$p(X = x_1, Y = y_1)$</td>
<td>$p(X = x_2, Y = y_1)$</td>
<td>$p(X = x_3, Y = y_1)$</td>
</tr>
<tr>
<td>$p(Y = y_2)$</td>
<td>$p(X = x_1, Y = y_2)$</td>
<td>$p(X = x_2, Y = y_2)$</td>
<td>$p(X = x_3, Y = y_2)$</td>
</tr>
<tr>
<td>$p(Y = y_3)$</td>
<td>$p(X = x_1, Y = y_3)$</td>
<td>$p(X = x_2, Y = y_3)$</td>
<td>$p(X = x_3, Y = y_3)$</td>
</tr>
</tbody>
</table>

How should we define statistical independence and correlations for this joint distribution?

Statistical independence means that knowing the value of one random variable does not affect the conditional probability of the other variable, in other words

$$p(Y = y_j) = p(Y = y_j | X = x_i)$$

and because

$$p(Y = y_j | X = x_i) = \frac{p(Y = y_j, X = x_i)}{p(X = x_i)}$$

$$p(Y = y_j, X = x_i) = p(X = x_i) p(Y = y_j)$$

for all $i$ and $j$. Therefore joint probability distribution of this form has statistically independent pairs of outcomes

<table>
<thead>
<tr>
<th>$Y \setminus X$</th>
<th>$p(X = x_1)$</th>
<th>$p(X = x_2)$</th>
<th>$p(X = x_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(Y = y_1)$</td>
<td>$p(X = x_1) p(Y = y_1)$</td>
<td>$p(X = x_2) p(Y = y_1)$</td>
<td>$p(X = x_3) p(Y = y_1)$</td>
</tr>
<tr>
<td>$p(Y = y_2)$</td>
<td>$p(X = x_1) p(Y = y_2)$</td>
<td>$p(X = x_2) p(Y = y_2)$</td>
<td>$p(X = x_3) p(Y = y_2)$</td>
</tr>
<tr>
<td>$p(Y = y_3)$</td>
<td>$p(X = x_1) p(Y = y_3)$</td>
<td>$p(X = x_2) p(Y = y_3)$</td>
<td>$p(X = x_3) p(Y = y_3)$</td>
</tr>
</tbody>
</table>

The probability distribution of $Y$ is

$$\mathcal{P}_Y = \{p(Y = y_1), p(Y = y_2), p(Y = y_3)\}.$$
When expressed as the sums of elements of the joint distribution $P_{XY}$

$$
P_Y = \{ +p(X = x_1, Y = y_1) + p(X = x_2, Y = y_1) + p(X = x_3, Y = y_1), \\
+ p(X = x_1, Y = y_2) + p(X = x_2, Y = y_2) + p(X = x_3, Y = y_2), \\
+ p(X = x_1, Y = y_3) + p(X = x_2, Y = y_3) + p(X = x_3, Y = y_3) \}$$

where $j$-th element is the sum of the $j$-th row. The condition of statistical independence makes these joint probabilities decomposed into products of individual probabilities. This has a strong implication. The ratios of the elements of each column are the same. (And similarly the ratios of elements of each row are the same).

Call $r_j = p(Y = y_j)$ and $c_i = p(X = x_i)$. The distribution becomes

<table>
<thead>
<tr>
<th>$Y \setminus X$</th>
<th>$p(X = x_1)$</th>
<th>$p(X = x_2)$</th>
<th>$p(X = x_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(Y = y_1)$</td>
<td>$r_1c_1$</td>
<td>$r_1c_2$</td>
<td>$r_1c_3$</td>
</tr>
<tr>
<td>$p(Y = y_2)$</td>
<td>$r_2c_1$</td>
<td>$r_2c_2$</td>
<td>$r_2c_3$</td>
</tr>
<tr>
<td>$p(Y = y_3)$</td>
<td>$r_3c_1$</td>
<td>$r_3c_2$</td>
<td>$r_3c_3$</td>
</tr>
</tbody>
</table>

Look at the ratios of elements of $P_Y$ and compare it to the ratios of the elements of the each column.

$$p(Y = y_1) : p(Y = y_2) : p(Y = y_3) = (r_1c_1 + r_1c_2 + r_1c_3) : (r_2c_1 + r_2c_2 + r_2c_3) : (r_3c_1 + r_3c_2 + r_3c_3)$$

$$= r_1(c_1 + c_2 + c_3) : r_2(c_1 + c_2 + c_3) : r_3(c_1 + c_2 + c_3)$$

$$= r_1 : r_2 : r_3$$

The ratios of elements of the first column

$$p(X = x_1, Y = y_1) : p(X = x_1, Y = y_2) : p(X = x_1, Y = y_3) = r_1c_1 : r_2c_1 : r_3c_1$$

$$= r_1 : r_2 : r_3$$

which is true for the second and third column.
These make \( p(Y = y_j) = p(Y = y_j | X = x_i) \) hold for all \( i \).

Similar relations also hold for the ratios of element of \( P \) and the ratios of elements of each row.

\[
\begin{align*}
p(X = x_1) : p(X = x_2) : p(X = x_3) &= \\
p(X = x_1, Y = y_1) : p(X = x_1, Y = y_2) : p(X = x_1, Y = y_3) &= \\
p(X = x_2, Y = y_1) : p(X = x_2, Y = y_2) : p(X = x_2, Y = y_3) &= \\
p(X = x_3, Y = y_1) : p(X = x_3, Y = y_2) : p(X = x_3, Y = y_3) &= c_1 : c_2 : c_3
\end{align*}
\]

These make \( p(X = x_i) = p(X = x_i | Y = y_j) \) hold for all \( j \).

An important conclusion of this realization implies that all columns and all rows of the joint distribution \( P_{XY} \) are linearly dependent among themselves. Which also means that the determinant of the \( P_{XY} \) is 0. Therefore the generalization of statistical independence to higher number of outcomes is that the determinant of the matrix that expresses the joint probability distribution is 0

\[
\det (P_{XY}) = 0.
\]

Note that \( p_{12}p_{12} - p_{12}p_{12} = 0 \) was the special case of this for the 2 events per variable case.

Also note that statistical independence implies of all outcome pairs implies zero determinant but not the vice versa. Two of the columns can be linearly dependent but not the third one. This case will give a zero determinant but not statistical independence between all outcome pairs.

In the case of correlation we should determine in which joint occurrences we are interested. Note that we did not give exact values for \( x_i \) and \( y_j \). We can use a convention and say we are interested in joint occurrence of \( X = x_i \) and \( Y = y_i \), the elements on the diagonal.

\[
C = \sum_i p(X = x_i, Y = y_i) - \sum_{j \neq k} p(X = x_j, Y = y_j).
\]

With this choice of a correlation function over the joint probability distribution \( C(P_{XY}) \), the more the
distribution is accumulated on the diagonal elements the closer \( C \) will be to perfect correlation. Note that there will not be a well-defined perfect anti-correlation. Because the opposite of \( X = x_i \) is not defined.

The most common correlation function between two random variables (with finite second moment) that takes the outcome values into account is the covariance. It is the average of the products of the deviations of two random variables around their mean values. Let us define the mean (or average, or expectation) value of a random variable as the sum of the products of each possible outcome and its occurrence probability

\[
\langle X \rangle = \sum_{x_i \in X} x_i p(X = x_i).
\]

Note that, when the distribution is sampled and we do not know the \( p(X = x_i) \), the mean value corresponds to \( \frac{1}{N} \sum_{n=1}^{N} X^{(n)} \) where \( X^{(n)} \in X \) is the \( n \)-th variable sampled from the distribution and \( N \) is the number of samples collected.

The deviation of an outcome from the mean value is

\[
\Delta x_i = x_i - \langle X \rangle.
\]

The product of the deviations of two random variables is

\[
\Delta x_i \Delta y_j = (x_i - \langle X \rangle)(y_j - \langle Y \rangle)
\]

Hence the covariance, which is the mean of the products of deviation is

\[
\text{cov} (X, Y) = \langle \Delta x_i \Delta y_j \rangle
\]

\[
= \sum_{x_i \in X} \sum_{y_j \in Y} (x_i - \langle X \rangle)(y_j - \langle Y \rangle)p(X = x_i, Y = y_j)
\]

\[
= \sum_{x_i \in X} \sum_{y_j \in Y} [x_i y_j - x_i \langle Y \rangle - y_j \langle X \rangle + \langle X \rangle \langle Y \rangle]p(X = x_i, Y = y_j)
\]

\[
= \langle XY \rangle - \langle X \rangle \sum_{x_i \in X} x_i p(X = x_i) - \langle Y \rangle \sum_{y_j \in Y} y_j p(Y = y_j) + \langle X \rangle \langle Y \rangle \sum_{x_i \in X} \sum_{y_j \in Y} p(X = x_i, Y = y_j)
\]

\[
= \langle XY \rangle - \langle X \rangle \langle Y \rangle - \langle X \rangle \langle Y \rangle + \langle X \rangle \langle Y \rangle
\]

\[
= \langle XY \rangle - \langle X \rangle \langle Y \rangle
\]

Note that, when the joint distribution is sampled and we do not know the \( p(X = x_i, Y = y_j) \), the mean value of \( XY \) corresponds to \( \langle XY \rangle = \frac{1}{N} \sum_{n=1}^{N} X^{(n)}Y^{(n)} \) where \( X^{(n)} \in X \) and \( Y^{(n)} \in Y \) are the \( n \)-th pair
sampled from the joint distribution.

In this light, the correlation we defined previously as

\[C = (p_{12} + p_{1\bar{2}}) - (p_{\bar{1}2} + p_{1\bar{2}})\]
\[= (P(X = x_1, Y = y_1) + P(X = x_2, Y = y_2)) - (P(X = x_2, Y = y_1) + P(X = x_1, Y = y_2))\]

can be thought as \(C = \langle XY \rangle\) for the chosen values for the outcomes \(x_1 = y_1 = +1\) and \(x_2 = y_2 = -1\).

### 1.2.8 Correlation and Statistical Dependence

Note that \(S\) and \(C\) are different measures of statistical relations. For the two event per random variable case, the joint distribution can be chosen so that while \(S\) is in one of these three cases 1) \(S > 0\), 2) \(S = 0\), 3) \(S < 0\), \(C\) can be in any of these three cases 1) \(C < 0\), 2) \(C = 0\), 3) \(C > 0\). (Except, for example, if \(C = 1\), namely the events are perfectly correlated, then, because \(p_1 = p_2 = p_{12}, S = p_{12} - p_1 p_2 = p_{12} - p_{1\bar{2}}^2 = p_{12} (1 - p_{12}) = p_{12} p_{\bar{1}\bar{2}} > 0\). And similarly for the perfect anti-correlation case \(C = -1 \Rightarrow S < 0\)).

In correlations we are interested in likelihood of some joint outcomes happening together, and some other joint outcomes not happening together.

For example, the covariance is defined so that in measures the likelihood of values of two random variables that are higher than average (and the values that are lower than average) occur at the same time. For a pair of outcomes \((x_i, y_j)\) the cases 1) \(x_i > \langle X \rangle\) and \(y_j > \langle Y \rangle\) and 2) \(x_i < \langle X \rangle\) and \(y_j < \langle Y \rangle\) have positive contribution and the cases 3) \(x_i < \langle X \rangle\) and \(y_j > \langle Y \rangle\) and 4) \(x_i > \langle X \rangle\) and \(y_j < \langle Y \rangle\) have negative contribution to covariance.

Note that the way covariance defined makes bigger deviations to have higher impact on the result. Whether this can be a feature we desire or not depends on the aspect of the probability distribution that we are studying.

In general, we use the term correlation as any linear functions of joint outcomes of random variables. We can have more than 2 random variables. If the outcomes are used as coefficients in the linear combinations then the correlations will depend on them. If constants are used as coefficients than the correlations will be independent of the outcome values.

Say we have 4 random variables \(a_1, a_2, b_1, b_2 \in \{0, 1, 2\}\), then, as an example, this expression is a correlation

\[B = +P(a_1 = 0, b_2 = 0) + 3P(a_2 = 2, b_1 = 0) - P(a_2 = 1, b_2 = 0) - 4P(a_1 = 1, b_1 = 2)\]
Here, the joint event \( a_1 = 0, b_2 = 0 \) has a positive contribution to \( B \), the joint event \( a_2 = 2, b_1 = 0 \) has even bigger contribution to \( B \), and \( a_2 = 1, b_2 = 0 \) has a negative contribution and \( a_1 = 1, b_1 = 2 \) has a bigger negative contribution.

**Bell's theorem** is about finding classical bounds for such correlations in terms of **Bell inequalities**. Whereas non-local game type Bell's theorems are about the classical bounds for statistical dependence relations, such as mutual information, where the actual values of outcomes do not matter [14].

We made this extensive description of the concept of correlation because in so many texts the phrase “quantum mechanics allows stronger correlations then classical mechanics (or local realism)” is used without pinpointing what that phrase exactly implies by leaving the meanings of correlation and strongness vague.

### 1.2.9 Boole's Conditions of Probable Experience

Now that we introduced the concepts of correlation as linear combination of joint probabilities and statistical independence, let us investigate possible values that a correlation function can have.

The first investigation of this kind was done by George Boole (who also invented the boolean algebra and logic) in his book that was published in 1862 [9]. It is about a set of **relative frequencies** of events and conditions imposed on them.

A relative frequency of an event is the measured number of occurrences of that event divided by the total number of events. It is also called the empirical probability, whereas “regular” probability means the tendency of occurrence. For example, we measure the quantity \( X \ N \) times, the event of outcome \( X = x_i \) \( E_i \), occurred \( N_i \) times. Then the relative frequency of \( E_i \) is

\[
 f(E_i) = N_i/N.
\]

The more samples are collected the better the relative frequency represents the probability of the underlying model, as \( N \rightarrow \infty \) the \( f(E_i) \rightarrow p(E_i) \).

By the way, if there is no underlying model, then the limit does not exist. As \( N \rightarrow \infty \), \( f(E_i) \) do not converge to any value and just fluctuates.

\( f(E) \) is what is measured in the laboratories. For example, in experiments done with photons, the particle production rate in time is so high that \( f \) is good representative of \( p \).

The problem Boole was working on can be expressed as: “given a set of rational numbers as relative frequencies of a set of events \( \{E_i|i = 1..n\} \) (where these \( E_i \) can correspond to single measurement outcomes such as \( X = x_i \) or joint measurement outcomes such as \( X = x_i, Y = y_j \). Also these events can be statistically independent, or not.), whether these numbers can be produced by a probability distribution.”
Boole investigates the conditions imposed on relative frequencies by being generated from a probability distribution. He calls these conditions as “probable conditions of experience”. The question is whether a set of probabilities \( \{p_i | i = 1..n\} \) is probable allowed by the conditions. His conditions are expressed in terms of inequalities formed by linear combinations of \( \{p_i\} \).

Analyzing only the relative frequencies allows us to have an abstraction from the underlying physical mechanism of the experiments conducted.

Boole says besides the obvious limit

\[
0 \leq p_i \leq 1
\]

for all \( i \) (the limits of being non-negative and being equal or smaller than certainty), if these probabilities are logically connected (i.e. if some of them are probabilities associated with joint events (intersections) or unions of events etc.) there should be other conditions between \( p_i \)'s that are expressed in the form of

\[
b_1 p_1 + b_2 p_2 + \cdots + b_n p_n + b \geq 0.
\]

Here are some of the conditions Boole mentioned [9][34]

- if \( p_1 = p(E_1) \) and \( p_2 = p(E_1 \cap E_2) \), then

\[
p_1 - p_2 \geq 0.
\]

- if \( p_1 = p(E_1), p_2 = p(E_2), p_3 = p(E_1 \cap E_2) \) and \( p_4 = p(E_1 \cup E_2) \) (here \( \cap \) means “and” and \( \cup \) means “or”), then

\[
p_1 + p_2 - p_3 - p_4 = 0.
\]

- if \( p_1 = p(E_1) \) and \( p_2 = p(\bar{E}_1) \) (here \( \bar{E} \) means “not” \( E \)), then

\[
p_1 + p_2 - 1 = 0.
\]

- if \( p_1 = p(E_1), p_2 = p(E_2), p_3 = p(E_1 \cap E_2) \), then

\[
-p_1 - p_2 + p_3 + 1 \geq 0.
\]

There are two important properties of this framework. One is that the conditions are expressed in linear forms of probabilities, there are no higher order polynomials or exponentials of probabilities in the conditions.
Second, for a finite set of probabilities and logical relations among them, there are finite number of conditions (finite inequalities).

Therefore it is possible to check whether a set of probabilities $P = \{p_i\}$ are probable by putting them into each inequality imposed by the conditions and check whether the inequalities hold. If any of them does not hold then the experience described by $P$ is not probable.

Another name for the conditions expressed in terms of inequalities linear in probabilities is Bell inequalities! Physicist reinvented them a century later than the mathematician. But, of course, by the time Boole’s conditions were invented there was no expectation of them to be violated.

Let us derive the Boole’s conditions for the simplest logically connected probabilities. Then analyze how logical relations lead to these conditions using classical probability theory. Then describe the algorithm for deriving all inequalities related to a scenario.

### 1.2.10 Probable Conditions for the Simplest Case and Their Geometric Interpretation

Say we have two events $E_1$ and $E_2$ and the simplest logical relation between them is the AND operation, $E_1 \& E_2$. In probability theory the probabilities are measures on some probability spaces. For now let’s express the probabilities as ratios of areas in a Venn diagram.

In Figure 8 $U$ is the probability space which includes all events $E_1$, $E_2$, $\bar{E}_1$, $\bar{E}_2$ and their logical relations such as $E_1 \cap E_2$ and $E_1 \cup E_2$ etc. For example, any point inside $U$ will correspond either $E_1$ or $\bar{E}_1$. The probability of an event is equal to “the area associated with it” divided by “the area of the probability space”.

$$p(E_i) = \frac{\text{Area} (E_i)}{\text{Area} (U)}.$$ 

![Venn Diagram](image)

**Figure 8:** Probability space $U$ for two events $E_1$ and $E_2$ and the logical relation AND operation. Red area is proportional to $p(E_1)$, green area to $p(E_2)$ and their intersection yellow area to $p(E_1 \& E_2)$. 
In Figure 9 we see some extreme cases of how the events can be distributed in the probability space, so that we can figure out the relations between the probabilities.

For the case of the logical relation AND, in the first one the area of $E_2$ is totally included by the area of $E_1$ hence their intersection is $\text{Area}(E_2)$. Here we see the case where the intersection is equal to one of the parties, $p(E_1 \cap E_2) = p(E_2)$. In the second one the areas of $E_1$ and $E_2$ do not intersect hence there is no joint occurrences in this probability space, $p(E_1 \cap E_2) = 0$. In the third one the areas of $E_1$ and $E_2$ exactly overlap and hence the area of the intersection is equal to both of the areas $p(E_1 \cap E_2) = p(E_1) = p(E_2)$.

![Figure 9: Different extreme cases of how two events can be distributed on the probability space.](image)

For the case of the logical relation OR, if we add the areas of two event we count the intersection twice. We know that $A(E_1) = A(E_1 \setminus E_2) + A(E_1 \cap E_2)$ and $A(E_2) = A(E_2 \setminus E_1) + A(E_1 \cap E_2)$, and the area of union is $A(E_1 \cup E_2) = A(E_1 \setminus E_2) + A(E_1 \cap E_2) + A(E_2 \setminus E_1)$. Therefore $A(E_1) + A(E_2) = A(E_1 \cup E_2) + A(E_1 \cap E_2)$.

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2).$$

These rules tell give us the conditions for two events and their joint occurrence

- any probability as a ratio of the area of the universal set $U$, should be between 0 and 1.

$$0 \leq p(E_1), p(E_2), p(E_1 \cap E_2), p(E_1 \cup E_2) \leq 1.$$

- The area of intersection can be as small as zero and as big as the smaller area, hence

$$p(E_1 \cap E_2) \leq \min(p(E_1), p(E_2))$$

- Because the “$E_1 \text{ OR } E_2$” is also a probability it has to be less than 1 too,

$$p(E_1) + p(E_2) - p(E_1 \cap E_2) \leq 1.$$
Let us define $p_1 = p(E_1)$, $p_2 = p(E_2)$ and $p_3 = (E_1 \cap E_2)$. $P = \{p_1, p_2, p_3\}$ is a point in $\mathbb{R}^3$. Let us draw the set of all points $P$ that satisfy these conditions. $\mathcal{P} = \{P|p_1, p_2, p_3\text{ obeying inequalities}\}$.

Each inequality is a plane that divides $\mathbb{R}^3$ into two. The condition of probabilities being between 0 and 1, $0 \leq p_1, p_2, p_3 \leq 1$, constraints us in the unit cube $[0,1]^3$. $p_3 \leq p_1$ and $p_3 \leq p_2$ gives us two more planes that cut the unit cube. And finally $p_1 + p_2 - p_3 \leq 1$ is the last plane.

Figure 10: The volume of the simplex represents all triplets $P = \{p_1, p_2, p_{12}\}$ that satisfy Boole's conditions of probable experience, $\mathcal{P}_C$. $x$-coordinate is $p_1$, $y$-coordinate is $p_2$ and $z$-coordinate is $p_{12}$.

Figure 10 shows $\mathcal{P}_C$, all triplets that satisfy the inequalities. The type of the geometric object in which all points of the set are a convex combinations of the vertices is called a convex polytope. This polytope has 4 vertices $P_i = (p_1, p_2, p_3)$:

1. $P_1 = (0, 0, 0)$ where none of the events happen,
2. $P_2 = (1, 0, 0)$ where $E_1$ occurs but not $E_2$
3. $P_3 = (0, 1, 0)$ where $E_2$ occurs but not $E_1$
4. $P_4 = (1, 1, 1)$ where both events happen, hence the joint event happens too.
Any point inside $P \in \mathcal{P}_C$ is a convex combination of these 4 points.

$$P = \lambda_1 P_1 + \lambda_2 P_2 + \lambda_3 P_3 + \lambda_4 P_4$$

where all coefficients are positive and add up to 1, $\lambda_i \geq 0 \forall i$, $\sum_i \lambda_i = 1$. $P = (\lambda_2 + \lambda_4, \lambda_3 + \lambda_4, \lambda_4)$ is a parametrization of all $P \in \mathcal{P}_C$.

Remember that Bell’s theorem is about the fact that quantum mechanics can generate a point outside of the polytope, $P_Q \notin \mathcal{P}_C$. For this scenario where $p_1, p_2$ and $p_3$ are defined the way they are, $\mathcal{P}_Q = \mathcal{P}_C$. Hence, even though we can call these inequalities Bell inequalities in the sense that they express classical bounds, they cannot be violated by quantum mechanics.

1.2.11 Convex Polytopes as Sets of Probable Conditions, Computational Complexity of Finding All Bell Inequalities for a Given Scenario

Last section was about the simplest analysis of probable experience in the case of two elementary events $E_1, E_2$ related with the logical operation AND, $E = \{E_1, E_2; E_{1\&E_2}\}$. We saw that the probable conditions are expressed by a list inequalities. The probabilities of elementary and joint events

$$P = \{p(E_1), p(E_2), p(E_{1\&E_2})\}$$

that satisfy the probable conditions (that satisfy the inequalities) live in a subset of $\mathcal{P}_C \subset \mathbb{R}^3$. More precisely the set $\mathcal{P}_C$ is a convex polytope.

Here we extend this framework to arbitrary number of events and joint events. This scheme was invented by Pitowsky [33]. Say we have $N_E$ elementary events (corresponding to measurement outcomes in the form $X \in \mathcal{X}_i$ where $\mathcal{X}_i \subseteq \mathcal{X}$ with $\mathcal{X}$ being all possible values of the random variable $X$.) and $N_J$ joint events constructed as joint occurrences of some pairs of elementary events

$$E = \{E_1, E_2, \ldots, E_{N_E}; E_{ij}, E_{i'j'}, \ldots\}.$$ 

Prepare a table of where the elements of $E$ are at the header. Each row will be a different possible combination of the values that they can take (1 for happening and 0 for not happening). For example for
the previous scenario the table becomes

<table>
<thead>
<tr>
<th>$E_1$</th>
<th>$E_2$</th>
<th>$E_{1&amp;E_2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

where each row corresponds to a vertex of the polytope, the entries of row $i$ are the coordinates of the $P_i$.

In the general case

<table>
<thead>
<tr>
<th>$E_1$</th>
<th>$E_2$</th>
<th>$\cdots$</th>
<th>$E_{N_E-1}$</th>
<th>$E_{N_E}$</th>
<th>$E_{1&amp;E_2}$</th>
<th>$E_{3&amp;E_4}$</th>
<th>$\cdots$</th>
<th>$E_{N_E-1&amp;E_{N_E}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>$2^{N_E}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

there are $2^{N_E}$ vertices. The choice of the joint events is arbitrary, at most $N_J$ can be $\binom{N_E}{2}$. The polytope lives in a $N_E + N_J$ dimensional space. $\mathcal{P} \subset \mathbb{R}^{N_E+N_J}$.

In the previous section we figured out the Bell’s inequalities from basic probability relations using Venn diagrams, and interpreted them as planes that separate a high dimensional space, and all intersections of half spaces turned out to be a convex polytope. Which allowed us to find the vertices of the convex polytope. It was a translation from the inequality description of the polytope to the vertex description. The inequality description is called the **H-Representation** and the vertex description is called the **V-representation**.

In this section, by creating the table, we calculate the V-representation of the polytope. And it is possible to translate this V-representation to H-representation using linear programming. The Weyl-Minkowski theorem [31, 42] says convex polytopes have equivalent representations that one can translate between each other, in other words these two statements are equivalent

- $\mathcal{P}$ is a polyhedron (a superset of convex polytopes), namely, for some real matrix $A$ and vector $b$,

$$\mathcal{P} = \{P|AP \leq b\}.$$
There are finite vectors $P_1, P_2, \ldots, P_n$ and $Q_1, Q_2, \ldots, Q_m$ in $\mathbb{R}^d$ such that

$$\mathcal{P} = \text{conv}(P_1, P_2, \ldots, P_n) + \text{nonneg}(Q_1, Q_2, \ldots, Q_m).$$

Here each row in $AP \leq b$ is another inequality that determines the probable conditions. The $j$-th row is $A_{j1}p_1 + A_{j2}p_2 + \cdots + A_{jn}p_n + b \geq 0$. If $A$ is $m \times n$ dimensional than $AP \leq b$ is a compact expression of $m$ inequalities. And conv is the convex combination, $\text{conv}(P_1, P_2, \ldots, P_n) = \sum_{i}^{n} \lambda_i P_i$ with $\sum_{i=1}^{n} \lambda_i = 1$ and $\lambda_i \geq 0 \forall i$. nonneg part creates cones, in our case the contribution from that part is zero, so we can ignore it.

Translation between the two types of representations is an NP-hard problem. As the polytope gets bigger (as we include more elementary events, and joint events to our table) the number of calculations to do the task increases exponentially for the best known algorithms.

The preparation of the truth table is the easier step that can be done in no time on a computer, which corresponds to calculating the V-representation, finding the vertices of the polytope. Translation to H-representation means finding $A, b$ in $AP \leq b$, the inequalities of probable conditions, the necessary conditions that have to be satisfied by classical systems. They correspond to the faces of the polytope.

It is the biggest unknown in theoretical computer science that whether NP-complete problems can be solved efficiently in polynomial time. The consensus is they cannot, but not proven yet. (The allowance of NP-complete problems to be solved in polynomial time leads to situations in computational science that are similar to grandfather-paradox in special relativity due to violation of the speed of light.)

Therefore the problem of finding all Bell inequalities for a given scenario is practically impossible problem to do it via brute force with this polytope scheme. We need clever solutions for special cases.

1.2.12 Wigner Bell Inequality from $N_E = 3, N_J = 3$ Polytope

Let us analyze the next simplest case with three elementary events, and all joint events among them.
<table>
<thead>
<tr>
<th></th>
<th>$E_1$</th>
<th>$E_2$</th>
<th>$E_3$</th>
<th>$E_1 &amp; E_2$</th>
<th>$E_1 &amp; E_3$</th>
<th>$E_2 &amp; E_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The table defines 8 vertices $P_i$ of the polytope in a 6 dimensional space where $P = \{ p_1, p_2, p_3, p_{12}, p_{13}, p_{23} \}$.

When H-representation is calculated from the V-representation, we get all inequalities from the previous smaller polytope: $0 \leq p_i, p_{ij} \leq 1$, $p_{ij} \leq \min (p_i, p_j)$, and $p_i + p_j - p_{ij} \leq 1$.

However, these conditions are not sufficient. All points $P \in \mathcal{P}_C$ satisfy them, but they are not enough to determine the set $\mathcal{P}_C$. The case with $p(E_1 \cup E_2 \cup E_3)$ being a probability makes it less than 1.

$$p(E_1 \cup E_2 \cup E_3) = p_1 + p_2 + p_3 - p_{12} - p_{13} - p_{23} \leq 1.$$  

By replacing $E_1$ with $\bar{E}_1$, and hence replacing $p_1$ with $1 - p_1$, $p_{12}$ with $p_2 - p_{12}$ and $p_{13}$ with $p_3 - p_{13}$ we get

$$p_1 - p_{12} - p_{13} + p_{23} \geq 0.$$  

Similarly replacing $E_2$ with $\bar{E}_2$ and $E_3$ with $\bar{E}_3$ gives us

$$p_2 - p_{23} - p_{12} + p_{13} \geq 0$$  
$$p_3 - p_{13} - p_{23} + p_{12} \geq 0.$$  

These inequalities can be produced by looking at the Venn diagrams of probability spaces. Figure 11 is the Venn diagram of 3 events, $E_1$, $E_2$ and $E_3$. Figure 12 explains the Wigner-Bell inequality $p_2 - p_{23} - p_{12} + p_{13}$ where the contribution from each term is represented in the diagram.

Inequalities from the previous case with the addition of the new four inequalities generates the necessary and sufficient conditions of probable experience.
Figure 11: Venn diagram depicting the probability space of 3 events $E_1$, $E_2$ and $E_3$.

Figure 12: The description of $p_2 - p_{23} - p_{12} + p_{13} \geq 0$ Wigner Bell inequality in a Venn diagram. Red areas are positive contributions and blue areas are negative contributions to the Bell expression. The total area has to be equal or bigger than 0.

These last four inequalities are the first (in terms of polytope complexity) Bell inequalities that can be violated by quantum mechanics. They were introduced by Wigner in 1970 [44].

Actually, Wigner invented a Bell-like framework ~ 40 years before this inequality. The Wigner distribution [43], which is a distribution on phase space of quantum mechanics, is intended as to be the quantum analogue of a distribution function of classical statistical mechanics. But for some quantum states, unlike a classical distribution function, the Wigner distribution has inevitable negative values. This was an indication of non-classical states and an example of the distinction between quantum and classical probabilities. But their importance related to the fundamentals of quantum mechanics research was not recognized by then.

1.2.13 Generic Bell Scenario

In general, a combined system can be made of $N$ parties, and we can choose among $M$ different possible quantities at each party, and each quantity can have $K$ different value. See Figure 13.


1.3 A selection of known Bell Inequalities

We showed that Pitowsky’s polytope method allows us to calculate all Bell inequalities for a given scenario, but unfortunately, it is computationally too complex to make an extensive search for larger scenario parameters $N, M, K$. In this section we mention a small selection of Bell inequalities from the literature.

1.3.1 Classes of Inequalities

For a given scenario, most of the facets of a the local polytope correspond to redundant inequalities, which can be transformed to each other by permuting observables or outcome labels. These type of permutations will give an inequality with the same form, but different variable or outcome names. Local polytopes are highly symmetric geometric objects. Their symmetry group has the cardinality $n!2^n$, where $n = NMK$ [33].

1.3.2 Clauser, Horne, Shimony, Holt (CHSH)

CHSH inequality that is found in 1969 [13] is the only class of inequality for $(2, 2, 2)$ scenario, a fact proven by Fine [21]. Four variables are $a_1, a_2, b_1, b_2 \in \{-1, +1\}$.

$$-2 \leq \langle a_1 b_1 \rangle + \langle a_1 b_2 \rangle - \langle a_2 b_1 \rangle + \langle a_2 b_2 \rangle \leq 2$$

It was the first Bell inequality generated with the intention of experimental validation. Its maximum quantum violation is $2\sqrt{2}$ was first calculated by Tsirelson [12].

Later CH version was introduced to overcome experimental difficulties. It was expressed as a linear combination of joint probabilities, instead of expectation values, and is in this form:
\[-1 \leq +P(a_1 = i, b_1 = i) + P(a_1 = i, b_2 = i) - P(a_2 = i, b_1 = i) + P(a_2 = i, b_2 = i) - P(a_1 = i) - P(b_2 = i) \leq 0\]

CH version is independent of the values that the variables can take. The observables’ only important property is that they can only take two possible values.

### 1.3.3 Greenberger, Horne, Zeilinger (GHZ)

GHZ’s Bell’s Theorem [18], which was discovered in 1989, is not in an inequality form. All other Bell’s theorems mentioned in this study are in the form of inequalities because they are about constraints on random events. GHZ theorem is about a case where the measurement outcomes can be deterministically predicted.

GHZ proposes a physical setup in which according to quantum mechanics the measurement outcomes are deterministic, but there is no single classical state that can produce those outcomes. Here we use Mermin’s explanation of GHZ theorem [30].

In a $(3, 2, 2)$ scenario, consider four quantities

$$a_1b_2c_2, \quad a_2b_1c_2, \quad a_2b_2c_1, \quad a_1b_1c_1$$

where $a_1, a_2, b_1, b_2, c_1, c_2 \in \{-1, +1\}$.

According to classical mechanics it is impossible to prepare a state $s = \{a_1, a_2, b_1, b_2, c_1, c_2\}$ for which the first three quantities have the value 1 and the fourth quantity has the value $-1$. The simple proof is that the product of the values of four quantities is $-1$, whereas the product of the quantities is $a_1^2 a_2^2 b_1^2 b_2^2 c_1^2 c_2^2$ which is $+1$.

However, it is possible to acquire these values from these four quantities in quantum mechanics if $a_1 = b_1 = c_1 = \sigma_x$, $a_2 = b_2 = c_2 = \sigma_y$ and $|\psi\rangle = (|000\rangle - |111\rangle) / \sqrt{2}$.

This example reflects the contextuality of quantum mechanics. A state cannot assign real numbers to variables, the measurement outcome of a quantity depends on its context, in order words, on which other quantities are measured simultaneously.
1.3.4 Werner, Wolf (N,2,2)

Werner and Wolf fully characterize the \((N, 2, 2)\) scenario by constructing \(2^{2N}\) independent inequalities [41]. In order words, their system of inequalities defines all facets of the local polytope, hence they provide necessary and sufficient conditions for local realism.

The observables are \(a_0, a_1, b_0, b_1, c_0, c_1, \ldots\).

\(c\), a configuration, is a list of assignments of one of the possible \(K\) outcomes to all \(NM\) observables. \(c \in \{-1, +1\}^{NM}\), hence \(c\) is an ordered list of \(NM - 1\) or 1s. Say if \(c = + + - + \ldots\) then the measured correlation is \(a_0 = 1, a_1 = 1, b_0 = -1, b_1 = 1, \ldots\).

\(s_n \in \{0, 1\}\) is the choice of measurement at the party \(n\). \(a_n (s_n) \in \{-1, +1\}\) is the observable measured by party \(n\). \(s\) is a setup, the list of choices by all parties \(s = (s_1, \cdots, s_N)\).

WW inequalities are full correlation inequalities. Each term in the inequality is an expectation of a product \(\prod_n a_n (s_n)\). Each expectation is the component of the vector \(\xi (s)\) which lie in \(2^N\)-dimensional space.

Bell inequalities are in the form

\[ \sum_s \beta (s) \cdot \xi (s) \leq 1 \]

with the normalization of coefficients \(\beta\) so that maximum classical value is 1. The Bell expression is

\[ B = \sum_s \beta (s) \left\langle \prod_{n=1}^{N} a_n (s_n) \right\rangle. \]

For example a choice of \(\beta = \left( \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2} \right)\) gives the CHSH inequality.

A configuration \(c\) is also the choice of \(c_n (s_n) \in \{-1, +1\}\) for all \(n\) and \(s_n\). There are \(2^{2N}\) of them.

The correlation vector has the components

\[ \epsilon_c (s) = \prod_{n=1}^{N} c_n (s_n). \]

Different classical configurations can give the same \(\epsilon_c (s)\), for example changing the signs of two different parties simultaneously will not change the result. Thus it is possible to rephrase as such

\[ \epsilon_c (s) = \pm (-1)^r s \]

where \(r \in \{0, 1\}^N\) goes over all bit strings of length \(N\). And the form of Bell inequalities become
\[-1 \leq \sum_{s} \beta(s) (-1)^{r_s} \leq 1\]

Define \(f(r)\) as \(\sum_{s} \beta(s) (-1)^{r_s}\). They are basically signs, which are in \((-1, 1\). Therefore facet inequalities can be found as

\[\beta(s) = 2^{-N} \sum_{r} f(r) (-1)^{r_s}\]

where \(f \in \{-1, +1\}^{2N}\). Also thanks to the index \(r\) going from 0 to \(2^{2N} - 1\) each Bell inequality has a natural numbering where each of them is assigned to an integer.

### 1.3.5 Collins, Gisin, Linden, Massar, Popescu (CLMPG)

Collins et al. found a class of Bell inequalities that works for \((2, 2, K)\) scenario [16]. The observables are \(a_1, a_2, b_1, b_2 \in \{0, \ldots, K - 1\}\). They use the following relation as their basis of defining local realistic systems. Given

\[B_K \equiv \sum_{k=0}^{[K/2]-1} \left(1 - \frac{2k}{K-1}\right) \{P (a_1 = b_1 + k) + P (b_1 = a_2 + k + 1) + P (a_2 = b_2 + k) + P (b_2 = a_1 + k)
- [P (a_1 = b_1 - k - 1) + P (b_1 = a_2 - k) + P (a_2 = b_2 - k - 1) + P (b_2 = a_1 - k - 1)]\}\]

the classical bound is

\[B_{K,c} \leq 2,\]

while quantum values for \(B_K\) can be as large as \(B_{3,q} \approx 2.87, \lim_{K \to \infty} B_{K,q} \approx 2.70\).

According to Kaszlikowski et al. higher dimensional inequalities are more resistant to the noise due to experimental conditions [29]. The introduction of ambiguities to measurement outcomes prevents the quantum violations less when the system at hand has higher number of possible outcomes.

CLMPG inequalities are later generalized by Son et al. to a bigger class of multipartite inequalities for \((N, 2, K)\) scenario [39].

### 1.3.6 Collins, Gisin (2,3,2)

Collins and Gisin worked on finding the complete set of inequalities for \(N = 2\) and small \(M\) and \(K\) scenarios [15]. They observed that even though the number of inequalities increases exponentially with increasing \(M\)
and $K$, for small values most of them are not relevant. For example for the $(2,3,2)$ there is only one new relevant inequality.

They define relevancy as such: If there is no quantum state that violates the newly found inequality that does not violate the old inequalities, then the new inequality is relevant. A member of this class is:

$$B = +P(A_1 = 0, B_1 = 0) + P(A_1 = 0, B_2 = 0) + P(A_1 = 0, B_3 = 0) + P(A_2 = 0, B_1 = 0) + P(A_2 = 0, B_2 = 0) - P(A_1 = 0, B_3 = 0) + P(A_1 = 0, B_3 = 0) - P(A_2 = 0) - P(B_1 = 0)$$

where

$$B_c \leq 0$$

The quantum state that violates this inequality but not the CHSH inequality is a mixed state

$$\rho = 0.85 |\phi_1\rangle \langle \phi_1| + 0.15 |\phi_2\rangle \langle \phi_2|$$

where $|\phi_1\rangle = (|00\rangle + |11\rangle)/\sqrt{5}$ and $|\phi_2\rangle = |01\rangle$.

1.3.7 Reviews and Online Database

For further extensive reviews see [10] and [37].

There is also a website with the goal of collecting all inequalities that appeared in the literature: www.faacets.com [36].

1.4 Non-local games as Bell Inequalities

Usually Bell inequalities are expressed as limits on certain correlations between remote random events (measurement outcomes) due to local realism. Nonlocal games, which is a recently introduced class of Bell inequalities [14], use the language of information and communication among remote parties who are investigated to tell whether the underlying physical system can be explained with classical mechanics only or whether quantum mechanics is necessary.

In a quantum nonlocal collaborative game there is a referee or verifier which randomly chooses questions
(or inputs) from a finite set for each player according to a prior probability distribution. The questions are sent to the players who are in different locations. The players give their answers (outputs) to their question without knowing which question is asked to other players or what their answers are. The answers are collected by the verifier. The verifier has the truth table in which correct answers to a set of question is listed. If their set of answers is in the list the players win the turn, otherwise they lose. The goal is to find a strategy that achieves the highest ratio of winning.

The players are not allowed to communicate while playing the game, but they can decide on a strategy beforehand. They are also allowed to share resources that do not allow direct communication. If these resources include entangled quantum systems, then their strategy is called a quantum strategy. Otherwise it is a classical strategy.

A mathematical description of a two player (Alice and Bob) game is as following [14]:

\[ S \] is the finite set of questions that are asked to Alice, and \( T \) to Bob. \( A \) is the finite set of answers that Alice can give, and \( B \) that Bob can give.

\( \pi \) is a probability distribution on \( S \times T \) and \( V \) (the truth table) is a predicate on \( S \times T \times A \times B \). \( V \) and \( \pi \) define a nonlocal game \( G = G(V, \pi) \). A pair of questions \((s, t) \in S \times T\) is chosen from the distribution \( \pi \). \( s \) is sent to Alice, \( t \) is sent to Bob. Alice gives the answer \( a \in A \), Bob gives the answer \( b \in B \). They win if \( V(s, t, a, b) = 1 \) and lose if \( V(s, t, a, b) = 0 \). The conditionality of the correctness of the pair of answers given the questions can be expressed as \( V(a, b | s, t) \).

The classical value \( \omega_c(G) \) of a game \( G \) is the maximum winning probability of the game using a classical strategy. Even though the players are allowed to share classical random resources, the classical value can always be reached by a deterministic strategy, because any probabilistic strategy can be expressed as a convex combination of classical strategies. And the maximum value that can be obtained will be the classical deterministic strategy with the highest winning rate.

A deterministic strategy means that the answer given by a party is a function of the question that is asked to that party. Namely, \( a = a(s) \) and \( b = b(t) \). Because of the lack of communication \( a \) cannot be a function of \( b \) or \( t \), similarly \( b \) cannot be a function of \( a \) or \( s \).

\[
\omega_c(G(V, \pi)) = \max_{a, b} \sum_{s, t} \pi(s, t) V(a(s), b(t) | s, t)
\]

where the maximum is over all functions \( a : S \rightarrow A \) and \( b : T \rightarrow B \). Because these four sets, \( A, B, S, T \) are finite, there are finite number of functions.

A quantum strategy allows sharing of an entangled two party system, say expressed by the state \( |\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \). Then, \( s \in S \) and \( t \in T \) correspond to the choice of the observable to measure by the parties. \( a \) is
the outcome of the measurement of observable $s$ done one $\mathcal{H}_a$, similarly $b$ on $\mathcal{H}_b$.

A quantum strategy can be expressed mathematically as following [14]:

\[
\{X^a_s | s \in S, a \in A\} \text{ and } \{Y^b_t | t \in T, b \in B\} \text{ are two sets of matrices that satisfy}
\]

\[
\sum_{a \in A} X^a_s = \sum_{b \in B} Y^b_t = I \quad \forall s, t
\]

Note that the dimensionality of these matrices $n$, can be equal or greater than the number of possible answers $|S| = |T| = K, n \geq K$. $n = K$ will correspond to a non-degenerate system, whereas $n > K$ introduces degeneracy and the quantum systems can have arbitrarily high dimensions.

The winning probability of a quantum strategy is then

\[
\sum_{s,t} \pi(s,t) \sum_{a,b} \langle \psi | X^a_s \otimes Y^b_t | \psi \rangle V(a,b|s,t)
\]

The quantum value, $\omega_q$, is the strategy that maximizes the winning probability.

The nonlocal game equivalent of the CHSH inequality can be expressed the following way [14]:

\[
S = T = A = B = \{0, 1\}. \pi \text{ is a uniform distribution.}
\]

\[
V(a,b|s,t) = \begin{cases} 
1 & \text{if } a \oplus b = s \land t \\
0 & \text{otherwise} 
\end{cases}
\]

where $\oplus$ is binary addition and $\land$ is the logical and operation.

The classical value is $\omega_c(G(V, \pi)) = 3/4$ which can be found by a search over all possible classical strategies.

The quantum value is $\omega_q(G(V, \pi)) = \cos^2(\pi/8) \approx 0.85$, which can be achieved by sharing the quantum state

\[
|\psi\rangle = (|00\rangle + |11\rangle) / \sqrt{2}
\]

and measuring following observables

\[
|\phi_0(\theta)\rangle \equiv \cos(\theta) |0\rangle + \sin(\theta) |1\rangle
\]

\[
|\phi_1(\theta)\rangle \equiv -\sin(\theta) |0\rangle + \cos(\theta) |1\rangle
\]
\[ X_0^a = |\phi_0(0)\rangle \langle \phi_0(0)| \]
\[ X_1^a = |\phi_0(\pi/4)\rangle \langle \phi_0(\pi/4)| \]
\[ Y_0^b = |\phi_1(\pi/8)\rangle \langle \phi_1(\pi/8)| \]
\[ Y_1^b = |\phi_1(-\pi/8)\rangle \langle \phi_1(-\pi/8)|. \]

There is a one-to-one translation between non-local games and full correlation Bell inequalities [38].
2 Jordan Basis Methods

2.1 Jordan Angles and Jordan Bases

In this chapter we are going to introduce a method to find the maximum quantum violations for a class of Bell inequalities. For that purpose we use the concept of Jordan principal angles.

The generalization of the idea of angles between geometric structures such as lines and planes to higher dimensional subspaces is first done by Jordan [28]. In a given inner product space, such as the Hilbert space we use in quantum mechanics, the relative orientation of two linear subspaces, which correspond to measurement outcomes in quantum mechanics, can be expressed by complex valued principle angles.

A recursive definition of principle angles is as such:

Given two subspaces \( U \) and \( V \) of a linear space \( H \) with \( \dim U = k \leq \dim V = l \), there exists \( k \) angles \( 0 \leq \theta_1 \leq \cdots \leq \theta_k \leq \pi/2 \). First one is defined as

\[
\theta_1 = \min \left\{ \arccos \left( \frac{\langle u \mid v \rangle}{\sqrt{\langle u \mid u \rangle} \sqrt{\langle v \mid v \rangle}} \right) \mid |u\rangle \in U, |v\rangle \in V \right\} = \arccos (\langle u_1 \mid v_1 \rangle).
\]

The angle \( \theta_1 \) that has the minimum value for the expression is called the first Jordan angle, or principal angle. The rest of the angles are defined such that \( \{|u_i\}\) and \( \{|v_i\}\) generate orthonormal bases, which we call Jordan bases.

\[
\theta_i = \min \left\{ \arccos \left( \frac{\langle u \mid v \rangle}{\sqrt{\langle u \mid u \rangle} \sqrt{\langle v \mid v \rangle}} \right) \mid |u\rangle \in U, |v\rangle \in V, \langle u \mid u_j \rangle = \langle v \mid v_j \rangle = 0, j = 1..i - 1 \right\} = \arccos (\langle u_i \mid v_i \rangle).
\]

With this ordering the Jordan angles \( \{\theta_i\} \) are defined uniquely for the given pair of subspaces whereas Jordan vectors are not. See figure 14 for a small dimensional geometric interpretation of Jordan angles.

The number of angles that are equal to zero indicates the dimension of the intersection of two subspaces, \( \dim (U \cap V) \). If all angles are zero, \( \theta_i = 0 \ \forall i \), then one subspace is subset of the other one, \( U \subset V \). If none of the Jordan angles is equal to zero, then the subspaces intersect only at the origin (and one of them has to be 1-dimensional) etc.

In quantum mechanics observables, represented by Hermitian matrices, split the Hilbert space \( \mathcal{H} \), into
Figure 14: Jordan angles related to the relative orientation of 2 planes, blue $U$ and red $V$, in 3D space $\mathbb{R}^3$. Blue arrows $\{|u_i\}$ is the orthonormal basis that spans $U$ and $\{|v_i\}$ spans $V$.

$\theta_1 = (|u_1\rangle, |v_1\rangle)$ which is the angle between $|u_1\rangle$ and $|v_1\rangle$ that lie on the intersecting lines between the planes, is zero. First vectors are parallel.

The angle between the other Jordan vectors is the non-zero Jordan angle, $\theta_2 = (|u_2\rangle, |v_2\rangle)$.

As can be seen, the choice of the vectors has some arbitrariness in them, but as long as minimization constraint is applied and $\theta_i = \arccos (\langle u_i | v_i \rangle)$ is chosen in $[0, \pi/2]$ the angles are unique.

Disjoint subspaces. Each possible outcome $i \in \{0, \cdots, K - 1\}$ of the observable $a_1$ corresponds a different subspace $S_i \equiv \mathcal{S}(a_1 = i)$, with

$$\bigcup_i S_i = \mathcal{H}$$

and

$$S_i \cap S_j = \begin{cases} \vec{0} & i \neq j \\ S_i & i = j \end{cases}$$

where $\vec{0}$ means the origin.

When all $\dim S_i = 1 \ \forall i$ then we say that the system is non-degenerate. Knowing the measurement outcome of that observable makes us certainly know the state of the system after the measurement. Whereas, if $\dim S_i > 1$, then the outcome $j$ do not tell us a unique measurement outcome state, though we'll know that the final state is in $S_j$.

To calculate the maximum quantum violation one cannot assume a certain dimensionality for the matrix representation of the observables. Even though $K = 2$ it does not mean that $\dim \mathcal{H} = 2$. It can be a degenerate system with $\dim \mathcal{H} > 2$. To deal with the arbitrariness in the dimensionality of the system we use Jordan bases.

Any two subspaces of $\mathcal{H}$, say $U$ and $V$, which correspond to different measurement outcomes of two
different observables can be expressed in orthonormal bases \{ |u_i\rangle \} and \{ |v_i\rangle \} with very nice overlap properties between basis elements corresponding to different spaces. These bases can be constructed for any two subspaces and their properties allow us to place bounds on the eigenvalues of Bell operators, which do not depend on the dimension of the overall Hilbert space or of the subspaces [23]. Jordan bases satisfy this property

\[ \langle u_i | v_j \rangle = \delta_{ij} \cos \theta_i. \]

Here \{ |u_i\rangle \} and \{ |v_i\rangle \} are called the Jordan bases and \{ \theta_i \} are the Jordan angles.

### 2.2 Finding Maximum Quantum Violations of a Class of Bell Inequalities

It is useful to know the largest violation of a given Bell inequality that quantum mechanics makes possible, and, in addition, which quantum state will produce this violation. The original work on this subject was done by B. Cirel’son, who showed that while the bound for the correlations in the CHSH inequality dictated by local realism is 2, the bound allowed by quantum mechanics is \(2\sqrt{2}\) [12].

Quantum bounds for other Bell inequalities have been found. Wehner found bounds for Bell inequalities of the CHSH type for \(N = 2, K = 2\), and general \(M\) [40]. Pal and Vertesi combined numerical methods with analytic upper bounds to find quantum bounds for a large number of Bell inequalities with \(N = 2,\) general \(M,\) and \(K = 2\) [32].

Here we will be considering inequalities belonging to the case \(M = K = 2\) with the object of determining their maximum quantum violation.

A Bell inequality is a full correlation Bell inequality if only expectation values of products one observable from each party appears as its terms, i.e. \(\langle a_{m_1} b_{m_2} \cdots z_{m_N} \rangle\) where \(m_\nu \in \{1, \ldots, M\}\). CHSH inequality is one of them. Full correlation Bell inequalities of this type were fully characterized by Werner and Wolf, and, in addition, they were able to show how to derive maximum quantum violations of these inequalities [41].

The inequalities we study are not full correlation inequalities, and we will use a different method to find their maximum violation. The class of inequalities we shall study can be considered generalizations of the CH inequality. The technique we shall apply is that of Jordan bases [23].

### 2.3 Inequalities

We begin by deriving the inequalities we wish to consider. Let us do it first for 2 parties, and then generalize to the case of \(N\) parties. Suppose that Alice can measure one of two observables, \(a_1\) and \(a_2\), and Bob can measure \(b_1\) and \(b_2\). Each of these variables has two measurement outcomes, \(\pm 1\).

Alice and Bob are presumed to be sufficiently far apart that their measurements are independent, i.e.
Bob’s choice of measurement will not influence the result of Alice’s, and vice versa. A source produces two particles, and sends one to Alice and one to Bob, and then Alice and Bob each perform a measurement on their respective particle. Local realism implies that the probabilities describing the measurement results satisfy

\[ P(a_1 = 1, b_1 = 1) - P(a_2 = 1, b_2 = 1) \leq P(a_2 = -1, b_1 = 1) + P(a_1 = 1, b_2 = -1). \]  

(1)

Here, \( P(a_j = m, b_k = m') \) is just the probability that if Alice measures \( a_j \) she gets \( m \), and if Bob measures \( b_k \) he gets \( m' \). We shall prove a more general version of this inequality shortly. This inequality is just a version of the CH Bell inequality. A three party version, where we add another participant, Charlie, with observables \( c_1 \) and \( c_2 \) is

\[ P(a_1 = 1, b_1 = 1, c_1 = 1) - P(a_2 = 1, b_2 = 1, c_2 = 1) \leq P(a_2 = -1, b_1 = 1, c_1 = 1) + P(a_1 = 1, b_2 = -1, c_1 = 1) + P(a_1 = 1, b_1 = 1, c_2 = -1). \]  

(2)

This can be extended to \( N \) parties. First term is the probability where first observables of all parties are 1, \( P(a_1 = 1, \ldots, z_1 = 1) \). Second term is minus the probability where second observables of all parties are 1, \( -P(a_2 = 1, \ldots, z_2 = 1) \). Then we have \( N \) terms. Probability of all but one party having their first observables set to 1 and only one party having the second observable set to \(-1\), \( P(a_1 = 1, \ldots, x_2 = -1, \ldots, z_1 = 1) \), where \( x \) goes from the first party to the last party.

Now let us prove the \( N \)-party version. In that case we label the observables as \( a_{lj} \), where \( l = 1, 2 \ldots n \) and \( j = 1, 2 \), so each of the \( n \) parties has two measurement choices, and each observable can take the values \pm 1.
The inequality can be expressed as

\[
P(a_{11} = 1, a_{21} = 1, \ldots a_{N1} = 1) \leq \]
\[
P(a_{12} = 1, a_{22} = 1, \ldots a_{N2} = 1) + P(a_{12} = -1, a_{21} = 1, \ldots a_{N1} = 1) + \ldots + P(a_{11} = 1, \ldots, a_{l-1,1} = 1, a_{l2} = -1, a_{l+1,1} = 1, \ldots, a_{N1} = 1) + \ldots + P(a_{11} = 1, a_{21} = 1, \ldots a_{N-1,1} = 1, a_{N2} = -1).
\]

(3)

Now, if our experiment can be described by a local realistic theory, then there is a joint distribution for all of the observables, \(P(a_{11}, a_{21} \ldots a_{N1}; a_{12}, a_{22}, \ldots a_{N2})\), and all of the probabilities in the above inequality can be expressed in terms of this joint distribution. For example for the \(N = 2\) case

\[
P(a_1 = 1, b_1 = 1) = P(a_{11} = 1; a_{21} = 1)
\]
\[
= +P(a_{11} = 1, a_{12} = 1; a_{21} = 1, a_{22} = 1)
\]
\[
+P(a_{11} = 1, a_{12} = -1; a_{21} = 1, a_{22} = 1)
\]
\[
+P(a_{11} = 1, a_{12} = 1; a_{21} = 1, a_{22} = -1)
\]
\[
+P(a_{11} = 1, a_{12} = -1; a_{21} = 1, a_{22} = -1)
\]
\[
\equiv P(1, 1; 1, 1) + P(1, -1; 1, 1) + P(1, 1; 1, -1) + P(1, -1; 1, -1).
\]

The joint distribution gives the probability of sequences of +1s and −1s of length \(2N\). In order to show that this inequality is true, all we need to show is that every sequence of length \(2N\) that appears in the probability on the left-hand side also appears in one of the probabilities on the right-hand side.

All of the sequences that appear on the left-hand side have their first \(N\) elements equal to 1, so these are the only sequences we need to consider.

\[
P(a_{11} = 1, a_{21} = 1, \ldots a_{N1} = 1) = \sum_{o_{21}, o_{22}, \ldots, o_{N2} \in \{-1, 1\}^N} P\left(1, 1, \ldots, 1; o_{21}, o_{22}, \ldots, o_{N2}\right)
\]

If the sequence has all of its elements equal to 1, \((o_{12}, o_{22}, \ldots, o_{N2}) = (1, 1, \ldots, 1)\), then it contributes to the first probability on the right-hand side, \(P(a_{12} = 1, a_{22} = 1, \ldots a_{N2} = 1)\), because
\[ P(a_{12} = 1, a_{22} = 1, \ldots, a_{N2} = 1) = \sum_{o_{21}, o_{22}, \ldots, o_{N2} \in \{-1, 1\}^N} P_{o_{11}, o_{21}, \ldots, o_{N1} = 1} \]

and the contributing term is \[ P_{1, 1, \ldots, 1; 1, 1, \ldots, 1} \]

Now suppose that some of its elements in its second half (in the second set of \( N \) elements) are equal to \(-1\), and suppose that one of the \(-1\)'s corresponds to the observable \( a_{12} \). This sequence will contribute to the probability \( P(a_{11} = 1, \ldots, a_{l-1,1} = 1, a_{l2} = -1, a_{l+1,1} = 1, \ldots, a_{N1} = 1) \) on the right-hand side.

Therefore, all sequences that contribute to the probability on the left-hand side do contribute to a probability on the right-hand side, so the inequality is proved.

### 2.4 Maximum quantum violation for 2 parties

In order to illustrate our approach of finding the maximum quantum violation, we will start with the simplest case of two parties. This will reproduce known results, but it gives a simple example of how the method based on Jordan bases works.

We have four observables, \( a_j \) and \( b_k \), \( j, k = 1, 2 \), which are now operators on the Hilbert space \( \mathcal{H}_a \otimes \mathcal{H}_b \), where \( \mathcal{H}_a \) is the Hilbert space in which Alice’s quantum states lie, and \( \mathcal{H}_b \) is the Hilbert space in which Bob’s states lie.

Our first problem in finding the maximum violation of Eq. (1), is that we do not know how we should choose the operators corresponding to Alice’s and Bob’s observables. For the moment let us specify them by their spectral projections.

Let \( Q_{a_j}, j = 1, 2 \), be the projection operator onto the subspace of \( \mathcal{H}_a \) on which \( a_j \) has the eigenvalue 1, and, similarly, \( Q_{b_j} \) is the projection operator onto the subspace of \( \mathcal{H}_b \) on which \( b_j \) has the eigenvalue 1. Then, the projection corresponding to the subspace on which \( a_j \) has the eigenvalue \(-1\) is \( I_a - Q_{a_j} \), where \( I_a \) is the identity on \( \mathcal{H}_a \), the projection onto the subspace of \( \mathcal{H}_b \) on which \( b_j \) has eigenvalue \(-1\) is \( I_b - Q_{b_j} \).

Defining the operator

\[
B_2 = Q_{a_1} \otimes Q_{b_1} - Q_{a_2} \otimes Q_{b_2} - (I_a - Q_{a_2}) \otimes Q_{b_1} - Q_{a_1} \otimes (I_b - Q_{b_2})
\]

the condition in Eq. (1) can be rephrased as

\[
\langle B_2 \rangle \leq 0
\]
which is

\[
\langle \Psi | B_2 | \Psi \rangle \leq \\
(\Psi | Q_{a1} \otimes Q_{b1} | \Psi \rangle - (\Psi | Q_{a2} \otimes Q_{b2} | \Psi \rangle - (\Psi | (I_a - Q_{a2}) \otimes Q_{b1} | \Psi \rangle \\
- (\Psi | Q_{a1} \otimes (I_b - Q_{b2}) | \Psi \rangle \leq \\
P(a_1 = 1, b_1 = 1) - P(a_2 = 1, b_2 = 1) - P(a_1 \neq 1, b_1 = 1) - P(a_1 = 1, b_2 \neq 1) \leq 0
\]

Our task is to find the largest positive eigenvalue of \( B_2 \), which will be the largest quantum violation of Eq. (1).

Let us do this operation for \( N = 2 \) party, non-degenerate case explicitly. For party-\( A \), \( a_1 = 1 \) corresponds to \( S_{a1} = \text{span} |a_1+\rangle \), \( a_2 = 1 \) corresponds to \( S_{a2} = \text{span} |a_2+\rangle \), and similarly for party-\( B \), \( b_1 = 1 \) corresponds to \( S_{b1} = \text{span} |b_1+\rangle \), \( b_2 = 1 \) corresponds to \( S_{b2} = \text{span} |b_2+\rangle \).

The Jordan angles are

\[
\langle a_1+ | a_2+ \rangle = \cos \theta_a \\
\langle b_1+ | b_2+ \rangle = \cos \theta_b
\]

And the projection operators are

\[
Q(S_{1a}) = Q(a_1+) = |a_1+\rangle \langle a_1+| \\
Q(S_{1b}) = Q(b_1+) = |b_1+\rangle \langle b_1+| \\
Q(S_{2a}) = Q(a_2+) = |a_2+\rangle \langle a_2+| \\
Q(S_{2b}) = Q(b_2+) = |b_2+\rangle \langle b_2+|
\]

Then the \( B \) operator in terms of the projection operators is

\[
B = Q_{1a} \otimes Q_{1b} + Q_{1a} \otimes Q_{2b} - Q_{2a} \otimes Q_{1b} + Q_{2a} \otimes Q_{2b} - Q_{1a} \otimes I - I \otimes Q_{2b}
\]

which can be expressed in terms of measurement outcome vectors
\[ B = + |a_1+\rangle \langle a_1+| \otimes |b_1+\rangle \langle b_1+| + |a_1+\rangle \langle a_1+| \otimes |b_2+\rangle \langle b_2+| \\
- |a_2+\rangle \langle a_2+| \otimes |b_1+\rangle \langle b_1+| + |a_2+\rangle \langle a_2+| \otimes |b_2+\rangle \langle b_2+| \\
- |a_1+\rangle \langle a_1+| \otimes I - I \otimes |b_2+\rangle \langle b_2+| \\
= + |a_1+, b_1+\rangle \langle a_1+, b_1+| + |a_1+, b_2+\rangle \langle a_1+, b_2+| \\
- |a_2+, b_1+\rangle \langle a_2+, b_1+| + |a_2+, b_2+\rangle \langle a_2+, b_2+| \\
- |a_1+\rangle \langle a_1+| \otimes I - I \otimes |b_2+\rangle \langle b_2+|. \]

We want to express \( B \) in the non-orthonormal basis set

\[ s_B = \{|a_1+, b_1+\rangle, |a_1+, b_2+\rangle, |a_2+, b_1+\rangle, |a_2+, b_2+\rangle\}. \]

In order to do that, apply \( B \) to each basis vector in \( s_B \) (omit the plus signs)

\[ B |a_1 b_1\rangle = |a_1 b_1\rangle + \cos \theta_b |a_1 b_2\rangle - \cos \theta_a |a_2 b_1\rangle + \cos \theta_a \cos \theta_b |a_2 b_2\rangle - |a_1 b_1\rangle - \cos \theta_b |a_1 b_2\rangle = 0 |a_1 b_1\rangle + 0 |a_1 b_2\rangle - \cos \theta_a |a_2 b_1\rangle + \cos \theta_a \cos \theta_b |a_2 b_2\rangle \]

\[ B |a_1 b_2\rangle = \cos \theta_b |a_1 b_1\rangle + |a_1 b_2\rangle - \cos \theta_a \cos \theta_b |a_2 b_1\rangle + \cos \theta_a |a_2 b_2\rangle - |a_1 b_2\rangle - |a_1 b_2\rangle = \cos \theta_b |a_1 b_1\rangle - 1 |a_1 b_2\rangle - \cos \theta_a \cos \theta_b |a_2 b_1\rangle + \cos \theta_a |a_2 b_2\rangle \]

\[ B |a_2 b_1\rangle = \cos \theta_a |a_1 b_1\rangle + \cos \theta_a \cos \theta_b |a_1 b_2\rangle - |a_2 b_1\rangle + \cos \theta_b |a_2 b_2\rangle - \cos \theta_a |a_1 b_1\rangle - \cos \theta_b |a_2 b_2\rangle = 0 |a_1 b_1\rangle + \cos \theta_a \cos \theta_b |a_1 b_2\rangle - 1 |a_2 b_1\rangle + 0 |a_2 b_2\rangle \]

\[ B |a_2 b_2\rangle = \cos \theta_a \cos \theta_b |a_1 b_1\rangle + \cos \theta_a |a_1 b_2\rangle - \cos \theta_b |a_2 b_1\rangle + |a_2 b_2\rangle - \cos \theta_a |a_1 b_2\rangle - |a_2 b_2\rangle = \cos \theta_a \cos \theta_b |a_1 b_1\rangle + 0 |a_1 b_2\rangle - \cos \theta_b |a_2 b_1\rangle + 0 |a_2 b_2\rangle \]
Let us see how $B$ acts on the subspace span $(s_B)$. Say $|\psi\rangle \in \text{span} (s_B)$

$$|\psi\rangle = \alpha^{11} |a_2 b_1\rangle + \alpha^{12} |a_1 b_2\rangle + \alpha^{21} |a_2 b_1\rangle + \alpha^{22} |a_2 b_2\rangle$$

$$\equiv (\alpha^{11}, \alpha^{12}, \alpha^{21}, \alpha^{22})_{s_B}.$$

If we investigate how $B$ applies on $|\psi\rangle$, we see that

$$B |\psi\rangle = \sum_{m,n=1}^2 \alpha^{mn}_{++} B |a_m + b_n\rangle$$

$$B |\psi\rangle = +\alpha^{11} B |a_1 b_1\rangle + \alpha^{12} B |a_1 b_2\rangle + \alpha^{21} B |a_2 b_1\rangle + \alpha^{22} B |a_2 b_2\rangle$$

$$B |\psi\rangle = -\cos \theta_a \alpha^{11} |a_2 b_1\rangle + \cos \theta_a \cos \theta_b \alpha^{11} |a_2 b_2\rangle$$

$$B |\psi\rangle = + \cos \theta_b \alpha^{12} |a_1 b_1\rangle - \alpha^{12} |a_1 b_2\rangle - \cos \theta_b \alpha^{12} |a_2 b_1\rangle + \cos \theta_a \alpha^{12} |a_2 b_2\rangle$$

$$B |\psi\rangle = + \cos \theta_b \alpha^{21} |a_1 b_2\rangle - \alpha^{21} |a_2 b_1\rangle + \cos \theta_b \alpha^{21} |a_1 b_1\rangle - \cos \theta_b \alpha^{22} |a_2 b_1\rangle$$

$$B |\psi\rangle = + |a_1 b_1\rangle (+ \cos \theta_b \alpha^{12} + \cos \theta_a \cos \theta_b \alpha^{22})$$

$$+ |a_1 b_2\rangle (-\alpha^{12} + \cos \theta_a \cos \theta_b \alpha^{21})$$

$$+ |a_2 b_1\rangle (-\cos \theta_a \alpha^{11} - \cos \theta_a \cos \theta_b \alpha^{12} - \alpha^{21} - \cos \theta_b \alpha^{22})$$

$$+ |a_2 b_2\rangle (+ \cos \theta_a \cos \theta_b \alpha^{11} + \cos \theta_a \alpha^{12})$$

Therefore the matrix representation of the $B$ operator in $s_B$ basis is
\[ B |\psi\rangle = B \begin{pmatrix} \alpha^{11} |a_1 b_1\rangle \\ \alpha^{12} |a_1 b_2\rangle \\ \alpha^{21} |a_2 b_1\rangle \\ \alpha^{22} |a_2 b_2\rangle \end{pmatrix} = \begin{pmatrix} 0 & \cos \theta_b & 0 & \cos \theta_a \cos \theta_b \\ 0 & -1 & \cos \theta_a \cos \theta_b & 0 \\ -\cos \theta_a & \cos \theta_a \cos \theta_b & -1 & -\cos \theta_b \\ \cos \theta_a \cos \theta_b & \cos \theta_a & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha^{11} |a_1 b_1\rangle \\ \alpha^{12} |a_1 b_2\rangle \\ \alpha^{21} |a_2 b_1\rangle \\ \alpha^{22} |a_2 b_2\rangle \end{pmatrix} \]

From here, the goal is to choose the parameters of the \( B \) operator, \( \theta_a \) and \( \theta_b \), namely the Jordan angles, so that we maximize the largest eigenvalue of \( B \).

Now, let us do the general case, which allows degeneracy, in order words a dimensionality higher than 1 for subspaces. Let \( S_{a1} \) and \( S_{a2} \) be the subspaces of \( \mathcal{H}_a \) onto which \( Q_{a1} \) and \( Q_{a2} \) project, respectively. Let \( \{|u_{1j}\}\} \) be an orthonormal basis for \( S_{a1} \) and \( \{|u_{2k}\}\} \) be an orthonormal basis for \( S_{a2} \). Note that we do not know the dimensions of \( S_{a1} \) and \( S_{a2} \), so we cannot specify how many vectors are in each basis. These bases can be chosen so that

\[ \langle u_{1j}|u_{2k}\rangle = \delta_{jk} \cos \theta_{aj} \]

where the \( \{\theta_{aj}\} \) are known as the Jordan angles and lie between 0 and \( \pi/2 \).

Similarly, one can define subspaces \( S_{b1} \) and \( S_{b2} \) corresponding to the ranges of \( Q_{b1} \) and \( Q_{b2} \), respectively, and Jordan bases \( \{|v_{1j}\}\} \) and \( \{|v_{2k}\}\) satisfying

\[ \langle v_{1j}|v_{2k}\rangle = \delta_{jk} \cos \theta_{bj} \]

where \( \{|v_{1j}\}\} \) is an orthonormal basis for \( S_{b1} \), and \( \{|v_{2k}\}\} \) is an orthonormal basis for \( S_{b2} \).

Note that the projection operators can be expressed as

\[ Q_{al} = \sum_j |u_{lj}\rangle \langle u_{lj}| \]
\[ Q_{bm} = \sum_k |v_{mk}\rangle \langle v_{mk}| \]

where \( l, m = 1, 2 \).

We now want to consider the behavior of the operator \( B_2 \), \( B \) operator for 2 parties, acting on states of
the form $|u_j⟩ \otimes |v_{mk}\rangle$, where $l, m = 1, 2$, in order to find its maximum eigenvalue eventually.

The terms of $B$ operator are in the form

$$Q_{ta} \otimes Q_{mb} = \sum_j |u_j⟩ \langle u_j| \otimes \sum_k |v_{mk}\rangle \langle v_{mk}|$$

$$= \sum_{j,k} |u_j, v_{mk}\rangle \langle u_j, v_{mk}|.$$ 

$$B_2 |u_{ij}⟩ |v_{1k}⟩ = (Q_{ta} \otimes Q_{1b} + Q_{1a} \otimes Q_{2b} - Q_{2a} \otimes Q_{1b} + Q_{2a} \otimes Q_{2b} - Q_{1a} \otimes I - I \otimes Q_{2b}) |u_{ij}, v_{1k}\rangle$$

$$= + \left( \sum_{j',k'} |u_{ij'}, v_{1k'}⟩ \langle u_{ij'}, v_{1k'}| u_{ij}, v_{1k}\rangle + \sum_{j',k'} |u_{ij'}, v_{2k'}⟩ \langle u_{ij'}, v_{2k'}| u_{ij}, v_{1k}\rangle \right)$$

$$+ \left( \sum_{j',k'} |u_{2j'}, v_{1k'}⟩ \langle u_{2j'}, v_{1k'}| u_{ij}, v_{1k}\rangle + \sum_{j',k'} |u_{2j'}, v_{2k'}⟩ \langle u_{2j'}, v_{2k'}| u_{ij}, v_{1k}\rangle \right)$$

$$- \left( \sum_{j'} |u_{1j'}⟩ \langle u_{1j'}| \otimes |v_{1k}\rangle - (I \otimes \sum_{k'} |v_{2k'}⟩ \langle v_{2k'}| u_{ij}, v_{1k}\rangle \right)$$

$$= + \left( \sum_{j',k'} |u_{1j'}, v_{1k'}⟩ \langle u_{1j'}, v_{1k'}| u_{ij}, v_{1k}\rangle + \sum_{j',k'} |u_{1j'}, v_{2k'}⟩ \langle u_{1j'}, v_{2k'}| u_{ij}, v_{1k}\rangle \right)$$

$$+ \left( \sum_{j',k'} |u_{2j'}, v_{1k'}⟩ \langle u_{2j'}, v_{1k'}| u_{ij}, v_{1k}\rangle + \sum_{j',k'} |u_{2j'}, v_{2k'}⟩ \langle u_{2j'}, v_{2k'}| u_{ij}, v_{1k}\rangle \right)$$

$$- \sum_{j'} |u_{1j'}⟩ \langle u_{1j'}| \otimes |v_{1k}\rangle - \sum_{k'} |v_{2k'}⟩ \langle v_{2k'}| u_{ij}, v_{1k}\rangle \right)$$

$$= + \left( \sum_{j',k'} |u_{1j'}, v_{1k'}⟩ \delta_{j',j} \delta_{k',k} + \sum_{j',k'} |u_{1j'}, v_{2k'}⟩ \delta_{j',j} \cos \theta_{bk'} \delta_{k',k} \right)$$

$$+ \left( \sum_{j',k'} |u_{2j'}, v_{1k'}⟩ \delta_{j',j} \cos \theta_{aj'} \delta_{k',k} + \sum_{j',k'} |u_{2j'}, v_{2k'}⟩ \delta_{j',j} \cos \theta_{aj'} \delta_{k',k} \cos \theta_{bk'} \right)$$

$$- \sum_{j'} |u_{1j'}⟩ \delta_{j',j} \otimes |v_{1k}\rangle - \sum_{k'} |v_{2k'}⟩ \delta_{k',k} \cos \theta_{bk'} \right)$$
\begin{align*}
&= + |u_{1j}, v_{1k}⟩ + |u_{1j}, v_{2k}⟩ \cos θ_{bk} \\
&+ |u_{2j}, v_{1k}⟩ \cos θ_{aj} + |u_{2j}, v_{2k}⟩ \cos θ_{aj} \cos θ_{bk} \\
&- |u_{1j}, v_{1k}⟩ - |u_{1j}, v_{2k}⟩ \cos θ_{bk}
\end{align*}

\[B_2 |u_{1j}, v_{1k}⟩ = 0 |u_{1j}, v_{1k}⟩ + 0 |u_{1j}, v_{2k}⟩ + \cos θ_{aj} |u_{2j}, v_{1k}⟩ + \cos θ_{aj} \cos θ_{bk} |u_{2j}, v_{2k}⟩\]

If we do this calculation for the other three vectors we will get

\begin{align*}
B_2 |u_{1j}, v_{2k}⟩ &= \cos θ_{bk} |u_{1j}, v_{1k}⟩ - 1 |u_{1j}, v_{2k}⟩ - \cos θ_{aj} \cos θ_{bk} |u_{2j}, v_{1k}⟩ - \cos θ_{aj} |u_{2j}, v_{2k}⟩ \\
B_2 |u_{2j}, v_{1k}⟩ &= 0 |u_{1j}, v_{1k}⟩ + \cos θ_{aj} \cos θ_{bk} |u_{1j}, v_{2k}⟩ - 1 |u_{2j}, v_{1k}⟩ + 0 |u_{2j}, v_{2k}⟩ \\
B_2 |u_{2j}, v_{2k}⟩ &= \cos θ_{aj} \cos θ_{bk} |u_{1j}, v_{1k}⟩ + 0 |u_{1j}, v_{2k}⟩ - \cos θ_{bk} |u_{2j}, v_{1k}⟩ + 0 |u_{2j}, v_{2k}⟩
\end{align*}

This shows an interesting feature of \(B_2\). Let us call the space spanned by the vectors

\[s_{j,k}^B = \{ |u_{1j}, v_{1k}⟩, |u_{1j}, v_{2k}⟩, |u_{2j}, v_{1k}⟩, |u_{2j}, v_{2k}⟩ \}\]

\[\text{span} s_{j,k}^B \equiv \mathcal{W}_{jk}.\] If \(B_2\) is applied on a vector in \(\mathcal{W}_{jk}\) the result will be in \(\mathcal{W}_{jk}\) too.

\[|ψ⟩ \in \mathcal{W}_{jk} \Rightarrow B_2 |ψ⟩ \in \mathcal{W}_{jk}.\]

What we find is that only states with the same values of \(j\) and \(k\) are coupled, i.e. \(B_2\) acting one one of these states yields a linear combination of states with the same values of \(j\) and \(k\). It is this feature that help us to solve the problem independent of the dimensionality of the subspaces corresponding to measurement outcome. Let us see why.
Let \( \dim S_{1a} = d_a \leq \dim S_{2a} \) and \( \dim S_{1b} = d_b \leq \dim S_{2b} \) without loss of generality, define

\[
s_B = \bigcup_{j,k} s_{j,k}^B
\]

\[
= \{ |u_{lj}, v_{mk}\rangle | l,m = 1,2, j = 1,\ldots,d_a, k = 1,\ldots,d_B \}
\]

\[
= \{ |u_{11}, v_{11}\rangle, |u_{11}, v_{21}\rangle, |u_{21}, v_{11}\rangle, |u_{21}, v_{21}\rangle, \ldots, |u_{1d_a}, v_{1d_b}\rangle, |u_{1d_a}, v_{2d_b}\rangle, |u_{2d_a}, v_{1d_b}\rangle, |u_{2d_a}, v_{2d_b}\rangle \}
\]

Say \( |\psi\rangle \in s_B \)

\[
|\psi\rangle = \sum_{j,k} \sum_{l,m} c_{lm}^{jk} |u_{lj}, v_{mk}\rangle
\]

\[
= \sum_{j,k} c_{11}^{jk} |u_{1j}, v_{1k}\rangle + c_{12}^{jk} |u_{1j}, v_{2k}\rangle + c_{21}^{jk} |u_{2j}, v_{1k}\rangle + c_{22}^{jk} |u_{2j}, v_{2k}\rangle
\]

Each term in this sum lives in a different \( \mathcal{W}_{jk} \).

Express \( |\psi\rangle \) as a linear combination of vectors in the set \( s_B \)

\[
|\psi\rangle = (c_{11}^{11}, c_{12}^{11}, c_{21}^{11}, c_{22}^{11}, \ldots, c_{11}^{jk}, c_{12}^{jk}, c_{21}^{jk}, c_{22}^{jk}, \ldots, c_{11}^{d_a d_b}, c_{12}^{d_a d_b}, c_{21}^{d_a d_b}, c_{22}^{d_a d_b})^T
\]

\[
= (\vec{c}_{11}, \ldots, \vec{c}_{jk}, \ldots, \vec{c}_{d_ad_b})^T
\]

When \( B_2 \) acts on \( |\psi\rangle \), define \( x_{aj} = \cos \theta_{aj} \) and \( x_{bk} = \cos \theta_{bk} \), then \( B_2 \) is
As can be seen \( B_2 \) is split into 4 dimensional parts. Define each of them as \( B_{2k}^{jk} \) where

\[
B_{2k}^{jk} = \begin{pmatrix}
-1 & 0 & 0 & \cos \theta_{aj} \cos \theta_{bk} \\
\cos \theta_{bk} & 0 & \cos \theta_{aj} \cos \theta_{bk} & 0 \\
\cos \theta_{aj} & \cos \theta_{aj} \cos \theta_{bk} & 0 & 0 \\
-\cos \theta_{aj} \cos \theta_{bk} & -\cos \theta_{aj} & -\cos \theta_{bk} & -1
\end{pmatrix}
\]

Then \( B \) acting on |\( \psi \rangle \) can be shown in a more concise notation as

\[
B_2 |\psi\rangle = \begin{pmatrix}
B_1^{11} & 0 & 0 \\
0 & \ddots & \ddots \\
0 & 0 & B_{KK}
\end{pmatrix}
\begin{pmatrix}
\vec{c}_1^{11} \\
\ddots \\
\vec{c}_d^{\alpha \beta}
\end{pmatrix}
\]

To find the maximum possible eigenvalue of \( B_2 \) we need to consider only one of these 4 dimensional blocks and assume that all blocks are equal.
\[ B_2|\psi\rangle = \begin{pmatrix} -1 & 0 & 0 & x_a x_b \\ x_b & 0 & x_a x_b & 0 \\ x_a & x_a x_b & 0 & 0 \\ -x_a x_b & -x_a & -x_b & -1 \end{pmatrix} \begin{pmatrix} c_{11} \\ c_{12} \\ c_{21} \\ c_{22} \end{pmatrix}, \tag{4} \]

What this means is that \( \mathcal{H}_a \otimes \mathcal{H}_b \) splits up into four-dimensional invariant subspaces under the action of \( B_2 \), and if we want to find the eigenvalues of \( B_2 \), we can examine each four-dimensional subspace individually. The characteristic equation of the above matrix is

\[ \lambda^2(\lambda^2 + 1) - x_a^2(1 - x_a^2)x_b^2(1 - x_b^2) = 0, \]

yielding a maximum eigenvalue of

\[ \lambda_{\text{max}} = \frac{1}{2} \left\{ \left[ 1 + 4x_a(1 - x_a^2)^{1/2}x_b(1 - x_b^2)^{1/2} \right]^{1/2} - 1 \right\}. \]

We can now maximize \( \lambda_{\text{max}} \) with respect to \( x_a \) and \( x_b \). The maximum occurs when \( x_a = x_b = 1/\sqrt{2} \) giving a maximum value for \( \lambda_{\text{max}} \) of \((\sqrt{2} - 1)/2\). This is the maximum quantum violation of the inequality in Eq. (1).

What we are actually doing is a search in parameter space \( \{x_a, x_b\} \) to choose an observable. In Bell inequality investigations we can start by choosing an arbitrary observable for one measurement without losing generality. For example, for the \( M = 2 \) case, the absolute orientations of the observations are not important, what is important is the relative orientation of the second observables (in the sense of Jordan angles) to the first observable. This is reflected by the fact that in polarization or spin direction measurements, the orientation of the first measurement is not essential, one can set it to any angle. The important choice in the experiment is the orientation of the second measurement direction relative to the first one.

Here, scanning the parameter space \( \{x_a, x_b\} \) allows us to try all possible second observables relative to the first one and calculating the maximum eigenvalue prevents us dealing with quantum states.

We can also find observables and a quantum state that attain this violation. Note that in the above calculation all of the subspaces, \( S_{al} \) and \( S_{bl} \), for \( l = 1, 2 \) are one dimensional, and the overlap between the vectors in the 1 and 2 subspaces is \( 1/\sqrt{2} \). We can, therefore, choose \( a_1 = \sigma_{za}, a_2 = \sigma_{za}, b_1 = \sigma_{zb}, \) and \( b_2 = \sigma_{zb}, \) where \( \sigma_{za} \) is just the sigma \( z \) Pauli matrix acting on the two-dimensional space \( \mathcal{H}_a, \sigma_{zb} \) is the
sigma \_z acting in the two-dimensional space \( \mathcal{H}_b \), and similarly for \( \sigma_{xa} \) and \( \sigma_{xb} \). We then have that

\[
|u_{11}\rangle = |0\rangle_a & |v_{11}\rangle = |0\rangle_b \\
|u_{21}\rangle = |\pm x\rangle_a & |v_{21}\rangle = |\pm x\rangle_b,
\]

where \( \sigma_z|0\rangle = |0\rangle, \sigma_z|1\rangle = -|1\rangle, \) and \( |\pm x\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2} \). The state that produces the maximum violation, which is just the eigenstate of the matrix in Eq. (4) with \( x_a = x_b = 1/\sqrt{2} \), corresponding to \( \lambda_{\text{max}} \), is, in the \( \{|0\rangle, |1\rangle\} \) basis,

\[
|\psi_2\rangle = \frac{1}{2} \left( \frac{1}{2 + \sqrt{2}} \right)^{1/2} [(1 + \sqrt{2})(|0\rangle_a|0\rangle_b - |1\rangle_a|1\rangle_b)
+ |0\rangle_a|1\rangle_b + |1\rangle_a|0\rangle_b].
\]

This state has interesting properties. First, it is a maximally entangled state. Next, suppose Alice and Bob independently decide to measure their part of the state in either the \( z \) or \( x \) basis. They then announce which basis they used. We want to maximize the chance that, no matter which basis choice they make, that once the basis choices are announced, each party can predict the other’s measurement result.

For example, suppose both parties decide to measure the state in the \( z \) basis, and suppose Alice got 1, corresponding to the state \( |0\rangle \) for her result. Then with a probability of \( (2 + \sqrt{2})/4 \approx 0.85 \), Bob will also have gotten the result 1. Similarly, if the both measured in the \( x \) basis, and Alice got 1 corresponding to \( |+x\rangle \), then the probability that Bob got \(-1\), corresponding to \(|-x\rangle \), is \( (2 + \sqrt{2})/4 \).

For all of the basis choices the correspondences are

\[
\begin{align*}
zz & \quad |0\rangle_a \leftrightarrow |0\rangle_b & |1\rangle_a \leftrightarrow |1\rangle_b \\
xx & \quad |0\rangle_a \leftrightarrow |+x\rangle_b & |1\rangle_a \leftrightarrow |-x\rangle_b \\
xx & \quad |+x\rangle_a \leftrightarrow |0\rangle_b & |1\rangle_a \leftrightarrow |1\rangle_b \\
xx & \quad |+x\rangle_a \leftrightarrow |0\rangle_b & |1\rangle_a \leftrightarrow |+x\rangle_b.
\end{align*}
\]

The first column gives the basis choices, with Alice’s choice first, and the next two columns give which of Alice’s and Bob’s measurement results correspond to each other. In all cases, if either Alice or Bob gets the measurement result shown above, then the probability that the other party obtains the corresponding state is \( (2 + \sqrt{2})/4 \). Therefore, in this state, Alice’s and Bob’s measurement results are highly correlated independent of whether they measure in the \( z \) or the \( x \) basis. It is correlations of this type that allow a quantum strategy of the CHSH nonlocal game to be better than any classical strategy [14].
2.5 Three parties

Now let us look at the three party inequality, Eq. (2). Before finding the maximum violation, which will ultimately require some numerical work, we present some simpler cases of quantum states that violate the inequality.

One possibility is to find a state that makes the probability \( P(a_1 = 1, b_1 = 1, c_1 = 1) \) nonzero and all of the others zero. In order to do this, we have to specify a quantum mechanical system and the observables.

We shall suppose that the system consists of three qubits, and that the observables labeled by 1 correspond to \( \sigma_z \) and those labeled by 2 correspond to \( \sigma_x \). That is, \( a_1 = \sigma_{za}, a_2 = \sigma_{xa}, \) etc.

Define the subspace \( S \) to be the span of the vectors \{ \( |+x, +x, +x\rangle, |−x, 0, 0\rangle, |0, −x, 0\rangle, |0, 0, −x\rangle \}, \) where in these states, the first slot is the state of qubit \( a \), the second the state of qubit \( b \), and the third of qubit \( c \). What we want is a vector that is orthogonal to \( S \) and has a nonzero overlap with the state \( |0, 0, 0\rangle \). Such a state is

\[
|\psi'_3\rangle = \frac{1}{2\sqrt{2}}(|0, 0, 0\rangle + |0, 0, 1\rangle + |0, 1, 0\rangle \\
+ |1, 0, 0\rangle - |0, 1, 1\rangle - |1, 0, 1\rangle - |1, 1, 0\rangle \\
- |1, 1, 1\rangle.
\]

With this state, the left-hand side of the inequality in Eq. (2) becomes \( 1/8 \), and the right-hand side is zero. As we shall see, this is by no means the largest violation we can obtain.

This state has the following correlation properties. If all of the parties measure in the \( z \) basis, or two of them measure in the \( x \) basis and the remaining party measures in the \( z \) basis, then the measurement results are uncorrelated. That is, one party does not have any information from his measurement, what the results of the other two measurements were.

The situation is different, however, if all of the parties measure in the \( x \) basis or one measures in the \( x \) basis and two measure in the \( z \) basis. In order to see what happens when everyone measures in the \( x \) basis, we can express the state as

\[
|\psi'_3\rangle = \frac{1}{2}(|+x, +x, −x\rangle + |+x, −x, +x\rangle \\
+ |−x, +x, +x\rangle − |−x, −x, −x\rangle).
\]

From this expression, we see that if one party gets \( +x \) for his measurement, he can be assured that the remaining parties got opposite results for theirs, that is one got \( +x \) and the other got \( −x \). If one party gets
−x, however, then he knows that the remaining parties got the same result for their measurements, either both got +x, or both got −x.

Now suppose that one party, say the first, measures in the $x$ basis while the other two measure in the $z$ basis. It is now convenient to express the state as

$$|\psi_3'\rangle = \frac{1}{2} [ |+x\rangle (|0,0\rangle - |1,1\rangle) + |−x\rangle (|0,1\rangle + |1,0\rangle)].$$

From this we see the following. If the party measuring in the $x$ basis gets $+x$, then the other two parties will get the same measurement result, and if he gets $−x$, then the other two parties will get opposite results. Now, if one of the parties measuring in the $z$ basis gets 0, then the other two parties got either $+x$ and 0 or $−x$ and 1, while if the $z$ result was 1, then the other two parties got either $+x$ and 1 or $−x$ and 0.

One can obtain a larger violation of the inequality in Eq. (2) with the same quantum system, three qubits, and the same assignment of observables, but choosing a different quantum state. Under these assumptions the state that produces the largest violation is the eigenstate of the operator

$$B_3' = |0,0,0\rangle\langle 0,0,0| - |+x,+x,+x\rangle\langle +x,+x,+x|$$

$$- (|−x,0,0\rangle\langle −x,0,0| - |0,−x,0\rangle\langle 0,−x,0|$$

$$- |0,0,−x\rangle\langle 0,0,−x|$$

with the largest eigenvalue.

Setting the characteristic polynomial of the operator equal to zero, we find that there are three eigenvalues of zero, two of $−1/2$, and the remaining three eigenvalues of roots of the cubic equation, $8\lambda^3 + 16\lambda^2 + 5\lambda − 2 = 0$. Solving this we find the one positive root is given, to three places, by 0.223. This is the largest violation of the inequality in Eq. (2) with this choice of variables, and it represents an improvement over our previous value of 1/8. As we shall see, we can do better.

Let us now apply the method developed in the previous section. We now make no assumptions as to what the observables are. With the notation as before, the operator corresponding to the inequality in Eq. (2) is

$$B_3 = Q_{a1} \otimes Q_{b1} \otimes Q_{c1} - Q_{a2} \otimes Q_{b2} \otimes Q_{c2} - (I_a - Q_{a2}) \otimes Q_{b1} \otimes Q_{c1}$$

$$- Q_{a1} \otimes (I_b - Q_{b2}) \otimes Q_{c1} - Q_{a1} \otimes Q_{b1} \otimes (I_c - Q_{c2})$$

(5)

where the $Q_{ij}$ operators, $l \in \{a, b, c\}$ and $j = 1, 2$, are projections operators on one of the Hilbert spaces $\mathcal{H}_l$,
We have the subspaces, $S_{ij}$, which are the ranges of the corresponding projections $Q_{ij}$ each with their Jordan bases. In particular, $\{|u_{1k}\}$ and $\{|w_{2k}\}$ are the Jordan bases for $S_{a1}$ and $S_{a2}$, respectively, $\{|v_{1k}\}$ and $\{|v_{2k}\}$ are the Jordan bases for $S_{b1}$ and $S_{b2}$, respectively, and $\{|w_{1k}\}$ and $\{|w_{2k}\}$ are the Jordan bases for $S_{a1}$ and $S_{a2}$, respectively.

We now consider the action of $B_3$ on vectors of the form $|u_{r\ell}\rangle_a|v_{ms}\rangle_b|w_{nt}\rangle_c$, where $l, m, n = 1, 2$, and $r$, $s$, and $t$ are fixed. We find that the subspace spanned by these eight vectors, which we shall call $T_{8,rst}$, is mapped into itself by $B_3$. In fact, denoting

$$|\phi_1\rangle = |u_{1r}\rangle_a|v_{1s}\rangle_b|w_{1t}\rangle_c \quad |\phi_4\rangle = |u_{2r}\rangle_a|v_{1s}\rangle_b|w_{1t}\rangle_c$$
$$|\phi_2\rangle = |u_{1r}\rangle_a|v_{2s}\rangle_b|w_{1t}\rangle_c \quad |\phi_5\rangle = |u_{2r}\rangle_a|v_{2s}\rangle_b|w_{2t}\rangle_c$$
$$|\phi_3\rangle = |u_{1r}\rangle_a|v_{2s}\rangle_b|w_{1t}\rangle_c,$$

and defining $T_{5,rst}$ to be the five-dimensional subspace spanned by these vectors, we find that $B_3$ maps $T_{8,rst}$ into $T_{5,rst}$. That means that if we wish to find nonzero eigenvalues of $B_3$, we only need to consider vectors in $T_{5,rst}$, which reduces our problem from an eight dimensional one to a five dimensional one.

In the subspace $T_{5,rst}$ and in the basis $\{|\phi_j\rangle|j = 1, 2, \ldots 5\}$, $B_3$ can be represented by the matrix

$$
\begin{pmatrix}
-2 & -x_c & -x_b & -x_a & x_a x_b x_c \\
x_c & 0 & x_b x_c & x_a x_c & 0 \\
x_b & x_b x_c & 0 & x_a x_b & 0 \\
x_a & x_a x_c & x_a x_b & 0 & 0 \\
x_a x_b x_c & -x_a x_b & x_a x_c & -x_b x_c & -1 \\
\end{pmatrix}
$$

The characteristic equation of this matrix is

$$
\lambda^5 + 3\lambda^4 + (2 + \gamma - \alpha + \beta)\lambda^3 + (\gamma - \alpha + \beta)\lambda^2
+ (2\gamma - \alpha - 3)\lambda + \beta(\alpha - 2\beta - 1) = 0,
$$

where

$$
\alpha = (x_a x_b)^2 + (x_a x_c)^2 + (x_b x_c)^2 \quad \beta = (x_a x_b x_c)^2
\quad \gamma = x_a^2 + x_b^2 + x_c^2
$$

We find numerically that the largest root achieved when $x_a = x_b = x_c$. An analytic argument for this condition is as following.
We first note that there is a positive root of this equation, because it is not too hard to show that when
\( \lambda = 0 \), the value of the characteristic polynomial is negative, but its value is clearly positive for sufficiently
large \( \lambda \).

We begin by defining

\[
F(\lambda, x_a, x_b, x_c) = \lambda^5 + 3\lambda^4 + (2 + \gamma - \alpha + \beta)\lambda^3 \\
+ \gamma - \alpha + \beta)\lambda^2 + (2\gamma - \alpha - 3)\lambda \\
+ \beta(\alpha - 2\beta - 1).
\]

The characteristic equation, \( F(\lambda, x_a, x_b, x_c) = 0 \), now defines \( \lambda \) as a function of \( x_a, x_b, \) and \( x_c \). We are
interested in points where \( \lambda(x_a, x_b, x_c) \) is a maximum, which means that we want

\[
\frac{\partial \lambda}{\partial x_a} = \frac{\partial \lambda}{\partial x_b} = \frac{\partial \lambda}{\partial x_c} = 0. \tag{6}
\]

Now we have that

\[
\frac{\partial F}{\partial \lambda} \frac{\partial \lambda}{\partial x_a} + \left( \frac{\partial F}{\partial x_a} \right)_\lambda = 0,
\]

and similarly for the derivatives with respect to \( x_b \) and \( x_c \). The subscript on the second term indicates that
\( \lambda \) is held constant during this differentiation. Therefore, the condition in Eq. (6) becomes

\[
\left( \frac{\partial F}{\partial x_a} \right)_\lambda = \left( \frac{\partial F}{\partial x_b} \right)_\lambda = \left( \frac{\partial F}{\partial x_c} \right)_\lambda = 0.
\]

Defining the three functions of \( x_a, x_b, \) and \( x_c \),

\[
f_0 = \beta(\alpha - 2\beta - 1) \quad f_2 = \gamma - \alpha + \beta \\
f_1 = \beta(2\gamma - \alpha - 3),
\]

the above equations become

\[
\frac{\partial f_2}{\partial x_a} (\lambda^3 + \lambda^2) + \frac{\partial f_1}{\partial x_a} \lambda + \frac{\partial f_0}{\partial x_a} = 0 \\
\frac{\partial f_2}{\partial x_b} (\lambda^3 + \lambda^2) + \frac{\partial f_1}{\partial x_b} \lambda + \frac{\partial f_0}{\partial x_b} = 0 \\
\frac{\partial f_2}{\partial x_c} (\lambda^3 + \lambda^2) + \frac{\partial f_1}{\partial x_c} \lambda + \frac{\partial f_0}{\partial x_c} = 0.
\]

These equations can be rearranged to eliminate the \( \lambda^3 \) and \( \lambda^2 \) terms, leaving us with three equations
that are linear in $\lambda$. For example, from the first two equations we find

\[ x_b(x_a^2 - 1) \left( \frac{\partial f_1}{\partial x_a} + \frac{\partial f_0}{\partial x_a} \right) = x_a(x_b^2 - 1) \left( \frac{\partial f_1}{\partial x_b} + \frac{\partial f_0}{\partial x_b} \right), \]

which becomes

\[ (x_a^2 - x_b^2)[f_1 - x_a^2 x_b^2 \beta(x_c^2 - 1)] \lambda \\
+ (x_a^2 - x_b^2)[f_0 - x_a^2 x_b^2 \beta(x_c^2 - 1)] = 0, \]

The two remaining conditions can be found from the one above by exchanging $x_b$ and $x_c$ to obtain the second condition, and by exchanging $x_a$ and $x_c$ to obtain the third (note that $f_0$, $f_1$, and $f_2$ are symmetric in $x_a$, $x_b$, and $x_c$).

Now, if $x_a = x_b = x_c$, these equations will be satisfied. If we assume that any two of the $x_a$, $x_b$, and $x_c$ are all different, then we find that it is necessary that $f_0 = f_1$. This, however, implies that $\lambda = -1$, which is not a positive root. Therefore, if a positive root is to have a maximum, then we need $x_a$, $x_b$, and $x_c$.

After the analytical argument, let us proceed with by setting $x_b$ and $x_c$ equal to $x_a$. We find that the characteristic equation can be expressed as

\[ (\lambda + x_a^2)^2[\lambda^3 + (3 - 2x_a^2)\lambda^2 + (2 - 3x_a^2 + x_a^6)\lambda \\
+ 3x_a^6 - 2x_a^8 - x_a^2] = 0. \]

The cubic equation does have a real, positive root, and that is the one in which we are interested. We find that its maximum value occurs when $x_a = [(\sqrt{5} - 1)/2]^{1/2} \approx 0.786$ and $\lambda_{max} = \sqrt{5} - 2 \approx 0.236$. Note that this is a larger violation than what we obtained when the parties measured in either the $z$ or the $x$ basis.

This solution also tells us what observables we should use to obtain the maximum violation, and gives us the state that produces this violation. For $a_1$, $b_1$, and $c_1$ we choose, as before, $\sigma_z$. Next, define the orthonormal vectors

\[ |u_+\rangle = x_a|0\rangle + \sqrt{1 - x_a^2}|1\rangle \\
|u_-angle = -\sqrt{1 - x_a^2}|0\rangle + x_a|1\rangle, \]

and for the operators $a_2$, $b_2$, and $c_2$ we choose $|u_+\rangle\langle u_+| - |u_-angle\langle u_-|$. The state that achieves the maximum
violation with this choice of observables is

$$|\psi_3\rangle = \left(4 - \frac{8}{\sqrt{5}}\right)^{1/2} |000\rangle + \left(-\frac{3}{2} + \frac{7}{2\sqrt{5}}\right)^{1/2} (-|001\rangle - |010\rangle - |100\rangle + |111\rangle)
- \left(1 - \frac{2}{\sqrt{5}}\right)^{1/2} (|011\rangle + |101\rangle + |110\rangle).$$

Note that this is not a GHZ state, i.e. it cannot be transformed by local unitaries into a state of the form $(|000\rangle + |111\rangle)\sqrt{2}$. If it could, when we formed a density matrix from the state and traced out Bob and Charlie, we would obtain a reduced density matrix proportional to the identity. That does not happen with $|\psi_3\rangle$.

Werner and Wolf showed that for all full correlation Bell inequalities with two measurement settings per party and each measurement having two outcomes, the maximally violating states are $n$-party generalizations of GHZ states [41]. The fact that $|\psi_3\rangle$ is not a GHZ state is a result of the fact that the inequality in Eq. (2) is not a full correlation inequality.

Finally, we note that the situation described by the inequality in Eq. (2) and its violation can be described in terms of a nonlocal game [14].

Each of the three parties, is sent an instruction bit by a referee, and they then send a bit back to the referee, who determines whether the parties have won the game or not. The parties are not allowed to communicate once the game has started. There are only five possible sets of instructions, and they are equally probable. They are either all zero, all one, or one of them is one and the other two are zero. The conditions for winning are

1. If all instruction bits are 0, then each party must return a 0.
2. If all instruction bits are 1, then not all parties return a 0, i.e. they only lose if all of them return a 0.
3. If two of the instruction bits are 0 and the remaining one is 1, then they only lose if the party who received a 1 returns 1 and the other two return 0.

Let us first consider a classical strategy. The optimal classical strategy is a deterministic one in which the bit each party sends is a function of the instruction bit they receive [14]. We shall show that any classical strategy must fail for at least one of the sets of instruction bits, which means that the maximum probability of winning is $4/5$. We shall then present a strategy that does succeed with this probability, which proves that this is the optimal classical probability of winning.

Now, suppose one of the parties receives an instruction bit of 0. Then in order to win in the case all of the instruction bits are 0, each party must return a bit of 0. Now consider what happens when one of
the parties receives a 1. If the other two parties receive a 0, then they will return a 0, so in order to win, the party we are considering should return a 0. So that means in all cases, each party should return a 0. However, if the instruction set consists of all 1’s, then they will all return 0’s and lose. Consequently, they cannot will all of the time, so the maximum winning probability is $4/5$ and the strategy where all parties always return 0 achieves this probability.

Now let us consider a quantum strategy. The parties share a quantum state, and if they receive an instruction bit 0 they measure observable 1 (that is, $a_1$, $b_1$, and $c_1$), and if they receive an instruction bit 1, they measure observable 2. They then send a bit corresponding to their measurement result, if their result is 1 they send 0, and if their result is −1 they send 1. Their probability of winning, $p_q$, is just

$$p_q = \frac{4}{5} + \frac{1}{5}\Delta,$$

where $\Delta$ is given by the left-hand side of the inequality in Eq. (2) minus the right-hand side.

If Alice, Bob, and Charlie share $|\psi_3\rangle$ and make the measurements that maximally violate the inequality in Eq. (2), then they achieve a winning probability of

$$p_q = \frac{4}{5} + \frac{1}{5}(\sqrt{5} - 2) \simeq 0.8472,$$

which is better than the classical result.

### 2.6 More than three parties

The same technique can be used to find maximum quantum violations of the inequality in Eq. (3) for $n$ parties. Each new party adds an additional Jordan angle. In all cases we examined, we find, numerically, that the maximum value of the positive root if the characteristic equation is achieved when all of the Jordan angles are equal.

This further implied that the relevant eigenvalue is a root of the cubic equation (this has only been verified up to $n = 7$)

$$\lambda^3 + [n - (n - 1)x_a^2]\lambda^2 + (n - 1 - nx_a^2 + x_a^{2n})\lambda + nx_a^{2n} - (n - 1)x_a^{2n+2} - x_a^2 = 0.$$
We also give the maximum violations, $\lambda_{\text{max}}$, and the values of $x_a$ that produce them in the following table:

<table>
<thead>
<tr>
<th>$n$</th>
<th>$x_a$</th>
<th>$\lambda_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.786151</td>
<td>0.236068</td>
</tr>
<tr>
<td>4</td>
<td>0.830913</td>
<td>0.249757</td>
</tr>
<tr>
<td>5</td>
<td>0.860012</td>
<td>0.257836</td>
</tr>
<tr>
<td>6</td>
<td>0.880509</td>
<td>0.263187</td>
</tr>
<tr>
<td>7</td>
<td>0.895745</td>
<td>0.266998</td>
</tr>
</tbody>
</table>

As can be seen from both the figure and the table, the size of the maximum violation increases with the number of parties. The values of $x_a$ can be used to construct the observables that achieve the maximum violation.
3 Group Theory Methods

In this chapter we will investigate an application of group theory to Bell’s theorem. It is about using group actions to generate Bell inequalities.

For a chosen finite group, the group representation operator of one or more group orbits are applied to a chosen initial state to generate a set of measurement outcome states. A Bell expression is constructed by adding the occurrence probabilities of the events corresponding to these measurement outcomes. The classical bound of the Bell expression is compared with the highest value it can take in quantum mechanics to see whether there is a violation. A search algorithm is implemented to find the orbits that gives violations.

3.1 Group Theory

Group theory is the study of the algebraic structures called groups. Groups are the main tools of abstract algebra. Elementary algebra is about numbers and the binary operations such as addition, subtraction, multiplication and division on the numbers. Abstract algebra is about the essence of elementary algebra, namely a set and a binary operation defined on the set.

The abstraction helps us to generalize the essential concepts of basic algebra. For example in a group the set on which the binary operations applied does not have to be made of numbers. And the operations do not have to be addition or multiplication etc. However, due to the choice of axioms to construct a group, the group will contain concepts such as associativity \((a + b) + c = a + (b + c)\) and identity element \(a + 0 = a\) etc.

There are other algebraic structures that capture more features of basic algebra, such as rings. A group has only one operator associated with it, whereas a ring has two operators and it can generalize addition and multiplication at the same time.

In our work we use groups because groups have matrix representations and closure property. When the matrices are unitary, they can be used to generate new quantum states from old ones, and the closure makes it certain that the new state is still in the set, because we want finite sequences.

3.1.1 Definition of a Group

A group is a mathematical structure that is made of a set \(G\) and an associated operation \(*\) that satisfy this four axioms:

- **associativity:** The operation \(*\) is associative, i.e. \(a * (b * c) = (a * b) * c\). The order of which operation is done first does not change the result.
- **identity**: There is an element \( e \) of \( G \) which satisfies \( e * g = g * e = g \) \( \forall g \in G \). \( e \) is called the **identity element**.

- **invertability**: \( \forall g \in G \) there is an element \( g^{-1} \in G \) with the property that \( g * g^{-1} = g^{-1} * g = e \). \( g^{-1} \) called the **inverse of** \( g \).

- **closure**: For all \( a, b \) \( \in G \), \( a * b \) is also in \( G \).

In some notations the group operation is omitted. For example the associativity axiom is expressed as \( a \,(bc) = (ab) \,c \).

Two important properties due to this axioms are that

1. the **identity element** of the group is unique and

2. the **inverse of a given element** is unique.

If there were two identities, \( e \) and \( e' \). The identity axiom says \( e * e' = e' \) and and \( e * e' = e \) hence \( e = e' \).

And if there were two inverses of \( a \), say \( a_1^{-1} \) and \( a_2^{-1} \) then \( a * a_1^{-1} = a_1^{-1} * a = e \) and \( a * a_2^{-1} = a_2^{-1} * a = e \).

Therefore \( a * a_1^{-1} = a * a_2^{-1} \). Apply \( a_1^{-1} \) from left \( a_1^{-1} * a * a_1^{-1} = a_1^{-1} * a * a_2^{-1} \). Which gives \( a_1^{-1} = a_2^{-1} \). □

### 3.1.2 Examples of Groups

Some example of groups expressed pairs of sets and operations are

- the set of integers and addition \((\mathbb{Z}, +)\), where \( e = 0 \) and \( x^{-1} = -x \)

- the set of rational numbers (except 0) and multiplication \((\mathbb{Q}^*, \times)\), where \( e = 1 \), \( x^{-1} = 1/x \)

As counter examples

- the set of natural numbers and addition \((\mathbb{N}, +)\) do not form a group, because the inverses are not in the group set. \( 5^{-1} = -5 \) but \( -5 \notin \mathbb{N} \).

- the set of integers and multiplication \((\mathbb{Z}, \times)\) do not form a group, because the inverses are not in the group \( 5^{-1} = 1/5 \notin \mathbb{Z} \).

The groups we mentioned has infinitely many elements. Groups can have finite number of elements too. The number of elements of the set of the group is called the **group order**, \(|G|\).

An example of a group of \(|G| = 2\) is \( \{+1, -1\}, \times \). \( +1 \times +1 = +1 \in G\), \( +1 \times -1 = -1 \in G\), \( -1 \times +1 = -1 \in G\) and \( -1 \times -1 = +1 \in G \). \( e = +1 \), \((-1)^{-1} = -1 \).
And the smallest possible group with one element is \( \{1\} \times \), 1 is the identity and only element of the group and it is its own inverse.

A way of displaying groups is using the **Cayley tables** which shows the outcome of the group operation on every pairs from the group set. The Cayley table of the group \( \{+1, -1\} \times \) is

\[
\begin{array}{c|cc}
\times & +1 & -1 \\
\hline
+1 & (1) \times +1 & (1) \times (-1) \\
-1 & (-1) \times (1) & (-1) \times (-1)
\end{array}
\]

A group’s set does not have to be made of numbers, they can constitute other mathematical structures or be abstract and have no structures. Let us show the Cayley table of an abstract group with two elements \( \{a, b\} \times \)

\[
\begin{array}{c|cc}
* & a & b \\
\hline
a & ? & \phantom{a} \\
b & \phantom{a} & \phantom{a}
\end{array}
\]

How to fill the table? First we have to decide on the outcome of the operation \( a \times a \). It can be 1) \( a \times a = a \) or 2) \( a \times a = b \)

\[
\begin{array}{c|cc}
* & a & b \\
\hline
a & a & ? \\
b & \phantom{a} & \phantom{a}
\end{array}
\]

\[
\begin{array}{c|cc}
* & a & b \\
\hline
a & a & ? \\
b & \phantom{a} & \phantom{a}
\end{array}
\]

To fill the top right entry we do not have any choice. If \( a \times a = a \) then \( a \times b = b \) and if \( a \times a = b \) then \( a \times b = a \), because \( a \times g_1 \) cannot be the same as \( a \times g_2 \) if \( g_1 \neq g_2 \). This is called the **rearrangement theorem**.

When all elements in the group set is applied to a single element one gets all elements of the group (in a different order in general), \( g \times G = \{g_1, g_2, \ldots, g_n\} = \{g_j g_1, g_j g_2, \ldots, g_j g_n\} = G \). In other words, in a Cayley table a group element cannot appear twice, and all group elements has to appear once, therefore each row (or column) is a permutation of the set of group elements.

Because, say, our group is a generic finite group, \( G = \{a_1, a_2, \ldots, a_n\} \). Then, the members of the \( j^{th} \) column are \( a_1 \times a_j, a_2 \times a_j, \ldots, a_n \times a_j \). Suppose any of these two are equal

\[
a_k \times a_j = a_l \times a_j = b
\]

\[
a_k \times a_j \times a_j^{-1} = a_l \times a_j \times a_j^{-1} = b \times a_j^{-1}
\]
(here we used associativity and invertibility) therefore

\[ a_k = a_1 = b \cdot a_j^{-1}. \]

This rule let us fill the rest of the tables like a Sudoku-like method

\[
\begin{array}{c|cc}
  & a & b \\
\hline
a & a & b \\
b & b & a
\end{array} \rightarrow \begin{array}{c|cc}
  & a & b \\
\hline
a & a & b \\
b & b & a
\end{array} \]

If we analyze these two tables closely, we can tell that in the \( a \ast a = a \) table \( a \) is the identity element, and in the \( a \ast a = b \) table \( b \) is the identity element, let’s call \( e \) and the other element \( c \)

\[
\begin{array}{c|cc}
  & a & b \\
\hline
a & a & b \\
b & b & a
\end{array} \rightarrow \begin{array}{c|cc}
  & e & c \\
\hline
e & e & c \\
\end{array}
\]

and if we rearrange \( e \)'s and \( c \)'s so that they have the same order in the top row and the left most column we get the same structure

\[
\begin{array}{c|cc}
  & e & c \\
\hline
 e & e & c \\
\end{array}
\]

\((G, \ast) = \begin{array}{cc|c}
  e & e & c \\
  c & c & e \\
\end{array}, \text { for } |G| = 2.\)

This means that there is only one group that has two elements. All other groups can be converted to this group by relabelling their elements (relabel the identity element to \( e \), the other element to \( c \), and the group operation \( \ast \)).

For example,

\[
\begin{array}{c|cc|c|c|c}
  & +1 & -1 & e & c & \oplus_{14} \\
\hline
+1 & +1 & -1 & e & e & 0 \\
-1 & -1 & +1 & c & c & 7
\end{array}
\]

In the table on the left \(+1 \leftrightarrow e\), \(-1 \leftrightarrow c\), and in the table on the right \(0 \leftrightarrow e\) and \(7 \leftrightarrow c\) (where \(\oplus_{14}\) means addition modulo 14. \(7 + 7 \mod 14 = 0\).)

This abstraction provides great simplification for classification of different structures in terms of how their elements are related with respect to the binary operation.

Let us construct groups of order 3 and 4.
\[ |G| = 3 \]

\[
\begin{array}{c|ccc}
* & e & a & b \\
\hline
 e & e & a & b \\
 a & a & ? & \\
b & b & & \\
\end{array}
\]

What can \( a \ast a \) be? Either \( e \) or \( b \).

\[
\begin{array}{c|ccc}
* & e & a & b \\
\hline
 e & e & a & b \\
 a & a & ? & \\
b & b & ? & \\
\end{array}
\rightarrow

\begin{array}{c|ccc}
* & e & a & b \\
\hline
 e & e & a & b \\
 a & a & b & e \\
b & b & e & \\
\end{array}
\]

In the second version the \( b \)'s are repeated in the third column and row, hence that is not accepted. Hence

\[
\begin{array}{c|ccc}
* & e & a & b \\
\hline
 e & e & a & b \\
 a & a & b & e \\
b & b & e & a \\
\end{array}
\]

is the only group that has order 3.

A simple group of order 3 is \((\mathbb{Z}, \oplus_3)\).

\[
\begin{array}{c|ccc}
\oplus_3 & 0 & 1 & 2 \\
\hline
 0 & 0 & 1 & 2 \\
 1 & 1 & 2 & 0 \\
 2 & 2 & 0 & 1 \\
\end{array}
\]

Or, using a different notation \((\mathbb{Z}_4, +) = (\{0, 1, 2, 3\}, +)\).

\[ |G| = 4 \]

A similar approach will lead two different types of groups of order 4. First let us define the **order** (or **period**) of a group element. The order of an element \( a \) is the smallest integer \( m = |a| \) that satisfies \( a^m = e \), where \( a^m \) means \( a \ast a \ast \cdots \ast a \). When \( m = 1 \), \( a = e \). For example in the previous group \((\mathbb{Z}, \oplus_3)\)

\[ |0| = 1, \text{ because } 0^1 = 0 = e, \text{ and } |1| = 3 \text{ because } 1^3 = 1 \oplus_3 1 \oplus_3 1 = 0, \text{ and } |2| = 3 \text{ because } 2^3 = 2 \oplus_3 2 \oplus_3 2 = 0. \]
Now, for a group of order four, assume that one element has order 3, say \(|a| = 3\), which means \(a^2 = a \ast a \neq e\). Call that element \(b\). And the last element is \(c\).

<table>
<thead>
<tr>
<th>*</th>
<th>e</th>
<th>a</th>
<th>b = a^2</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>e</td>
<td>a</td>
<td>a^2</td>
<td>c</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>a^2</td>
<td>a^3 = e</td>
<td>?</td>
</tr>
<tr>
<td>(b = a^2)</td>
<td>(a^2)</td>
<td>(a^3 = e)</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The only element that is not used in row and column two is \(c\) but we cannot use it because otherwise it will be repeated at the fourth row and column, therefore there is no element of order 3.

Other options for the order are \(|a| = 1, 2 \text{ and } 4\). 1 means \(a = e\), so that can not happen too. We need distinct elements.

The choice of \(|a| = 4\) will create a group that is equivalent to \((\mathbb{Z}_4, +)\).

The last choice \(|a| = 2\) will create a different type of group (this is the smallest group order where we have more than one type of group).

<table>
<thead>
<tr>
<th>*</th>
<th>e</th>
<th>a</th>
<th>b</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>e</td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>e</td>
<td>→</td>
<td>a</td>
</tr>
<tr>
<td>(a^2)</td>
<td>(b)</td>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td>c</td>
<td>c</td>
<td>c</td>
<td></td>
<td>c</td>
</tr>
</tbody>
</table>

Here we observe that \(c = ab\) and \(c = ba\) hence \(c = ab = ba\). Fill the rest accordingly

<table>
<thead>
<tr>
<th>*</th>
<th>e</th>
<th>a</th>
<th>b</th>
<th>ab</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>e</td>
<td>a</td>
<td>b</td>
<td>ab</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>e</td>
<td>ab</td>
<td>b</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>ab</td>
<td>(b^2)</td>
<td>(ab^2)</td>
</tr>
<tr>
<td>c</td>
<td>ab</td>
<td>b</td>
<td>(ab^2)</td>
<td>(a^2b^2)</td>
</tr>
</tbody>
</table>

Now \(b^2 = a\) or \(b^2 = e\). In the beginning of this case we assumed that there are no elements of order 4. \(b^2 = a\) means \(b^4 = a^2 = e\) hence \(|b| = 4\). We also proved that group elements can’t have order 3. Therefore \(|b|^2 = |c|^2 = 2\) too.
This group has a special name, it is called the Klein-4 group, $V_4$. And it is the second and the last group that has order $4$.

The groups of the form $(\mathbb{Z}_n, +)$ (namely, $(\mathbb{Z}_2, +), (\mathbb{Z}_3, +)$) have a special name. They are called **cyclic groups**. A group is cyclic if all of its elements can be expressed as some power of a single element. For example in $(\mathbb{Z}_4, +)$ the group set $\mathbb{Z}_4$ is $\{0, 1, 2, 3\}$. $1 = 1^1$, $2 = 1^2$, $3 = 1^3$, $0 = 1^4$.

Klein group is the smallest group that is not-cyclic.

Another categorization of groups are about the commutation of its elements. A group in which all elements commute, namely $g_1 \ast g_2 = g_2 \ast g_1 \ \forall g_1, g_2 \in G$, is called **Abelian**. If some group element do not commute, then that group is not abelian.

For example, Klein four group is abelian. This can be seen by the symmetric matrix structure of the Cayley table.

The problem of the number of different type of groups that have a given order is a complex problem. No explicit expression found yet, and the value is bounded from above by $n^{n^2}$,\(^1\)

### 3.1.3 Subgroups and Conjugacy Classes

A **subgroup** of a group is a subset of the group set of which elements form a group by themselves. For example in the Klein-four group $\langle \{e, a\} \rangle$, $\ast$ is a subgroup

$$
\begin{array}{c|cccc}
\ast & e & a & b & ab \\
e & e & a & b & ab \\
a & a & e & ab & b \\
b & b & ab & e & a \\
c & ab & b & a & e \\
\end{array}
$$

Similarly $\langle \{e, b\} \rangle$ and $\langle \{e, c\} \rangle$, $\ast$ too.

Similarly $\langle \{0, 2\} \rangle$, $+$ is a subgroup of $\langle \mathbb{Z}_4, + \rangle$

\(^1\)http://groupprops.subwiki.org/wiki/Number_of_groups_of_given_order
\[
\begin{array}{c|cc}
+ & 0 & 2 \\ 
0 & 0 & 2 \\ 
2 & 2 & 0 \\
\end{array}
\]

\{e\} and \(G\) itself are **improper** subgroups and the rest are **proper** subgroups.

Two elements \(a, b\) of a group are called **conjugate** if there is a group element \(g\) for which

\[a = gbg^{-1} \]

Conjugation is an equivalence relation, in other words conjugation is reflexive (group elements are conjugate to themselves), symmetric (if an element is conjugate to another element, then that another element is conjugate to the first element) and transitive (if one element is conjugate to a second element, and that second element is conjugate to a third element, then first element is conjugate to the third element).

Choose \(g = e\), then \(a = eae^{-1}\).

If \(a = gbg^{-1} \Rightarrow g^{-1}ag = g^{-1}gbg^{-1}g \Rightarrow cbe = g^{-1}ag \Rightarrow b = hah^{-1}\) where \(h = g^{-1}\).

If \(a = g_1b_1^{-1}\) and \(b = g_2c_2^{-1}\), then \(a = g_1g_2c_2^{-1}g_1^{-1} = hch^{-1}\) where \(h = g_1g_2\).

Equivalence relations leads to **classes**. Conjugation produces **conjugacy classes**. The set of all group elements that are conjugate to an element is the conjugacy class of that element.

The elements that belong to the same class have the same order. Say \(b = gag^{-1}\) and \(|a| = n\), namely, \(a^n = e\). Then

\[b^n = (gag^{-1})(gag^{-1})\cdots(gag^{-1})\quad\text{with} \; n\text{factors.}
\]

\[= ga^n g^{-1}
\]

\[= geg^{-1}
\]

\[= e\]

hence \(|b| = n\) too.

For an abelian group, all conjugacy classes have only one element. Because \(gag^{-1} = gg^{-1}a = a\) for all \(g \in G\). For more interesting cases we need to analyze non-abelian groups.

Two classes \(C_1\) and \(C_2\) are either the same or they have no common elements. Therefore conjugacy classes splits a group set into disjoint subsets.

Say \(a \in C_1\) and \(a \in C_2\). If \(c_1 \in C_1\) and \(c_2 \in C_2\) then due to the definition of a class there is a \(g_1\) such that \(a = g_1c_1g_1^{-1}\) and similarly \(a = g_2c_2g_2^{-1}\). Therefore \(g_1c_1g_1^{-1} = g_2c_2g_2^{-1} \Rightarrow c_1 = g_1^{-1}g_2c_2g_2^{-1}g_1\).
hence $c_1$ is conjugate to $c_2$. $C_1 \subseteq C_2$ and similarly $C_2 \subseteq C_1$ hence $C_1 = C_2$.

$$G = C_1 \cup C_2 \cup \cdots \cup C_L$$

where $L$ is the number of different conjugacy classes.

### 3.1.4 Generators and Generating Set of a Group

For a chosen subset of group elements, $S \subset G$, any product that uses only subset elements and their inverses is a **word** in $S$. A word has this form in general

$$g_1^{k_1} g_2^{k_2} \cdots g_n^{k_n}$$

where $g_i \in S$ and $k_i \in \{-1, +1\}$, $n$ is the length of the word. $g_i$’s in a word do not have to be unique, a group element can appear at several different locations in a word. Say $S = \{a, b, c\}$ then $abca^{-1}a, aa^{-1}a^{-1}cba$ are words.

Two words can be equivalent, such as $aa^{-1}b = ba^{-1}a$, and successively repeated elements can be expressed in power notation $bbca^{-1}a^{-1}a^{-1}c^{-1} = b^2ca^{-3}c^{-1}$.

According to the closure axiom each word made by the elements of $S \subset G$ has to be a group element of $G$ too.

If $S \subset G$, then $\langle S \rangle$, the **subgroup generated by** $S$, is the smallest subgroup of $G$ of which elements are equivalent to the words made of elements of $S$. (For finite $G$) $\langle S \rangle$ is not infinitely big, because even though the number of words that can be generated from a finite alphabet is infinite, most of them will be equivalent because due to the closure axiom elements of $\langle S \rangle$ has to be elements $G$. Therefore $\langle S \rangle$ can at most have $|G|$ elements.

A **generating set** of a group is a subset of the group from which all group elements can be produced as finite-length words. In other words, when $\langle S \rangle = G$ then $S$ is a generating set of $G$. The elements of the generating set are called **generators**.

The cyclic groups can be generated from a single element only. For $(\mathbb{Z}_n, +)$, $\{1\}$ is a generating set where $2 = 1^2, \ldots, n = 1^n$ and $0 = 1^{n+1}$. $(\mathbb{Z}_n, +) = \langle 1 \rangle$.

Among the groups that we saw until now, only the Klein-four group, $V_4$, is not cyclic. $V_4 = \langle \{a, b\} \rangle$ (for $a, b \in G$) because the rest of the $V_4$’s elements are words made of them: $c = ab$ and $e = (ab)^2 = abab$. 

3.1.5 Defining Relations and Group Presentations

A subset of group elements are called **independent** when each of them cannot be expressed as a combination of the rest. For example in $V_4$ $a$ and $b$ are independent. $a^k \neq b$ and $b^l \neq a$ for any $k, l \in \mathbb{Z}$, whereas in $(\mathbb{Z}_n, +)$ 1 and 2 are dependent, $1^2 = 2$.

The expressions that show how independent elements are related are called the **defining relations**. For example $(ab)^2 = 1$ is a defining relation for $V_4$.

Just knowing the elements of $S$ is not enough to reconstruct the group $G$, $S$ itself does not involve information on how the group operation acts and there is no way of telling which elements of $\langle S \rangle$ are equivalent when $G$ is not known.

A way to reconstruct a group $G = \langle S \rangle$ using the set $S$ is to know all necessary defining relations. The most compact definition of a group is expressing it by a generating set $S$ and the defining relations among the elements of $S$. This way of defining a group is called the **group presentation**. For example $V_4 = \{a, b | a^2 = b^2 = (ab)^2 = e\}$, $(\mathbb{Z}_n, +) = \{1 | 1^n = 0\}$.

A common convention is using only the defining relations that show which combinations of independent generators are equivalent to identity element. In that notation the equality to the identity element is omitted:

$$V_4 = \{a, b | a^2, b^2, (ab)^2\}$$

$$(\mathbb{Z}_n, +) = \{1 | 1^n\}$$

<table>
<thead>
<tr>
<th>*</th>
<th>e</th>
<th>a</th>
<th>b</th>
<th>ab</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>e</td>
<td>a</td>
<td>b</td>
<td>ab</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>e</td>
<td>ab</td>
<td>b</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>ab</td>
<td>e</td>
<td>a</td>
</tr>
<tr>
<td>c</td>
<td>ab</td>
<td>b</td>
<td>a</td>
<td>e</td>
</tr>
</tbody>
</table>

$$(\mathbb{Z}_n, +) = \{1 | 1^n\}$$

The presentation of a group is not unique because there can be different generating generating sets that generate the same group. For example $V_4 = \{a, c | a^2, c^2, caca^{-1}\}$.

3.1.6 Mapping Between Groups

A map between two algebraic structures that preserves the structure (keeps the relationship between its elements) is called a **homomorphism** in abstract algebra.
Say \((G, \ast)\) and \((H, \ast)\) are two groups and \(\phi\) is a map from \(G\) to \(H\),

\[
\phi : (G, \ast) \to (H, \ast).
\]

If

\[
\phi(g_1 \ast g_2) = \phi(g_1) \ast \phi(g_2)
\]

for all \(g_1, g_2 \in G\) then \(\phi\) is a homomorphism between \(G\) and \(H\). In other words, \(g_1 \in G\) is mapped to \(h_1 \in H\), \(\phi(g_1) = h_1\) and similarly \(\phi(g_2) = h_1\). Say \(g_1 \ast g_2 = g_3\) and \(h_1 \ast h_2 = h_3\). If \(\phi(g_3) = h_3\) then the group structure is preserved. The result of the group operation on two elements of the domain group is analogous to the result of the group operation on the image group’s analogous elements.

A map that is one-to-one (each element in the image is mapped by a unique element from the domain set, injective map) and onto (no element in the image is left without being mapped, surjective map) is called a bijection. A bijective homomorphism is called isomorphism. In other words if there is a unique \(\phi^{-1}, \phi^{-1} : (H, \ast) \to (G, \ast)\) given \(\phi, \phi : (G, \ast) \to (H, \ast)\) then \(\phi\) is an isomorphism and \(G\) and \(H\) are isomorphic, \(G \cong H\).

Homomorphism is a more general relationship than isomorphism. For example, all groups are homomorphic to the trivial group, the group that has only one element, which is identity, \(T = \{e\}, \ast\). Examples of trivial groups are \((\{0\}, +)\) and \((\{1\}, \times)\). If all elements of a group is mapped to the identity element of the trivial group, \(\phi(g) = e\) for \(g \in G\) and \(e \in T\), then the group structure will be preserved. \(\phi(g_1 \ast g_2) = e, \phi(g_1) \ast \phi(g_2) = e \ast e = e\).

As can be seen from this example we loose information on \(G\) when the image group \(H\) of the map is homomorphic but not isomorphic to \(G\).

From the perspective of abstract algebra isomorphic groups are equivalent, or the same. For example, \((\{+1, -1\}, \times) \cong (\{0, 1\}, +)\) because of the existence of the bijective map \(\phi(+1) = 0\) and \(\phi(-1) = 1\). This is what we meant when we said that there is only one group of order 2. The only difference between these groups is the names (labels) of their elements.

\[
\begin{array}{c|cc}
\ast & e & a \\
\hline
e & e & a & 0 & 2 & + & 0 & 1 \\
a & a & e & 2 & 0 & \times & +1 & -1 & 0 & 1
\end{array}
\]

Some properties due to isomorphism are

- Identity element of the domain group is always mapped to the identity element of the image group
• A cyclic group of order \( n \), \(|G| = n\), is isomorphic to \((\mathbb{Z}_n, +)\). \( G \cong (\mathbb{Z}_n, +)\).

• Every group that has a prime number order, \(|G| = p\), is isomorphic to the cyclic group of order \( p \), \((\mathbb{Z}_p, +)\). This means that there is only one group of order \( p \). We can say that groups of prime order are boring.

3.1.7 Direct Product of Groups

The **cartesian product** of two sets \( S \) and \( T \) is \( \{(s, t) | s \in S, \ t \in T\} \). \((s, t)\) is called an **ordered pair**, which means that the order of elements is important, \((0, 1)\) is different than \((1, 0)\). For example \( A = \{a_1, a_2\}\) and \( B = \{b_1, b_2, b_3\}\), then \( A \times B \) is \( \{(a_1, b_1), (a_1, b_2), (a_1, b_3), (a_2, b_1), (a_2, b_2), (a_2, b_3)\}\).

As a generalization, direct product of more than two sets, \( S_1 \times S_2 \times \cdots \times S_n \), is \( \{(s_1, s_2, \ldots, s_n) | s_i \in S_i, i = 1 \ldots n\} \), where \((s_1, s_2, \ldots, s_n)\) are ordered \( n \)-tuples.

Cartesian product is a way to create bigger sets from smaller sets. In a similar manner we can construct bigger groups from smaller groups. One way of doing that is using **direct product** of groups.

The direct product of two groups \((G, \ast)\) and \((H, \star)\) with \( g_i \in G \) and \( h_i \in H \) is a group, \((D, \odot)\), of which elements are \( d_k = (g_i, h_j) \). The group operation \( \odot \) is applied component-wise. For \( d_l = (g_m, h_n) \)

\[
d_k \odot d_l = (g_i, h_j) \odot (g_m, h_n)
= (g_i \ast g_m, h_j \star h_n)
\]

As an example let us show that Klein-four group is isomorphic to \((\mathbb{Z}_2, +) \times (\mathbb{Z}_2, +)\).

\[ V_4 \cong (\mathbb{Z}_2, +) \times (\mathbb{Z}_2, +) \]

\( \mathbb{Z}_2 = \{0, 1\} \) and

\[
\mathbb{Z}_2 \times \mathbb{Z}_2 = \{0, 1\} \times \{0, 1\}
= \{(0, 0), (0, 1), (1, 0), (1, 1)\}
\]

Let us calculate \((\mathbb{Z}_2 \times \mathbb{Z}_2, \odot)\) where \( \odot = (+, +) \) using the rule \((i, j) + (k, l) = (i + k, j + l)\). \((0, 0) \odot (0, 0) = (0 + 0, 0 + 0) = (0, 0), (1, 0) \odot (1, 1) = (1 + 1, 0 + 1) = (0, 1)\) etc. The Cayley table for the direct product group is
A map $e \leftrightarrow (0, 0) \leftrightarrow 0$, $a \leftrightarrow (0, 1) \leftrightarrow 1$, $b \leftrightarrow (1, 0) \leftrightarrow 2$, $c \leftrightarrow (1, 1) \leftrightarrow 3$ is an isomorphism between $V_4$ and $(\mathbb{Z}_2 \times \mathbb{Z}_2, (+, +))$. The positions of elements of $V_4$ on the Cayley table are the same as their counterparts’ positions of $(\mathbb{Z}_2 \times \mathbb{Z}_2, (+, +))$.

Call $U = e$ and $W = b$. Both of them are isomorphic to $Z_2 = (\mathbb{Z}_2, +)$. We can see that $V_4$ is made of blocks of $U$ and $W$.

This is due to the tensor-product-like nature of direct product. $V_4 = \mathbb{Z}_2 \times \mathbb{Z}_2$ which implies to take the direct product of each element in the Cayley table of the first $\mathbb{Z}_2$ with the Cayley table of the second $\mathbb{Z}_2$.
where direct product of elements \( a_i \in A, b_i \in B \) mean ordered pairs, \( a_i \times b_j \equiv (a_i, b_j) \).

This is an example of the direct product of two cyclic groups being not cyclic. Direct products can create more complex structures than their factors.

### 3.1.8 Group Theory and Symmetries

Group theory is used in studying symmetries of various systems such as physical systems or geometrical objects. A symmetry is a feature of a system that remains the same under a transformation.

As an example, time symmetries are about quantities that remain constant in time. When a time translation (which represents the passage of time) or reversal operation is applied we get the values of quantities at different time points. When a quantity remains the same after the time translation operation that quantity is a constant of motion.

Time translation operations form a group. Say \( T(t, t') \) operator maps the system at \( t' \), \( \psi (t') \), to the system at \( t \), \( \psi (t) \). \( T(t, t') \psi (t') = \psi (t) \). \( T(t, t') T(t', t'') = T(t, t'') \) is another time translation that maps the system as \( t'' \) to the system at \( t \). There are infinitely many time translations, therefore the group of time translation is continuous, infinite group.

Another example can be the symmetries of geometric objects. An object is symmetric under operations (such as rotation, reflection, translation) that leaves them the same. A square is symmetric under 0°, 90°, 180° and 270° rotations around their center, or reflections around the diagonals or lines that connect opposite sides’ centers.

These 8 operations form a group because any combination of them will be equal to another operation in the same group and each of them has their inverses. For example a 90° rotation followed by a 180° rotation is equal to a 270°rotation etc.

### 3.1.9 Permutations and Permutation Notations for Representing Group Elements

Imagine group elements acting on sets by permuting their elements. Say the set is the label of vertices of an \( n \) regular sided polygon. The vertices are labelled from 1 to \( n \) in clockwise order. Then a \( 2\pi/n \) rotation

| \( Z_2 \times Z_2 \rightarrow \) | \( \begin{array}{ccc}
0 & 1 \\
0 & 1
\end{array} \) | \( \begin{array}{ccc}
0 & 0 \times 0 & 0 \times 1 \\
0 \times 1 & 0 \times 0 & 1 \times 1
\end{array} \) | \( \begin{array}{ccc}
1 \times 0 & 1 \times 1 \\
1 \times 0 & 1 \times 0
\end{array} \) |
|-----------------|-----------------|-----------------|-----------------|
| \( \begin{array}{ccc}
0 \times 0 & 0 \times 1 \\
0 \times 1 & 0 \times 0
\end{array} \) | \( \begin{array}{ccc}
0 \times 0 & 0 \times 1 \\
0 \times 0 & 0 \times 1
\end{array} \) | \( \begin{array}{ccc}
0 \times 1 & 1 \times 0 \\
1 \times 1 & 1 \times 0
\end{array} \) | \( \begin{array}{ccc}
0 \times 1 & 1 \times 0 \\
1 \times 1 & 1 \times 0
\end{array} \) |
corresponds to a permutation where $1 \rightarrow 2$, $2 \rightarrow 3$, \ldots, $n-1 \rightarrow n$, $n \rightarrow 1$.

At the end of the permutation the vertex labels changes but the geometrical shape does not. The transformed version exactly overlaps the original version. Therefore some permutations on vertices are symmetry operations, and group theory is the link between the symmetries and permutations. This link allows us to represent group elements as permutation operations. In general, Cayley's Theorem says that every group is isomorphic to a group of permutations.

A bijection function from a set $A$ to itself is called a permutation. Say, for $A = \{1, 2, 3, 4\}$, and $\sigma : A \to A$ where $\sigma(1) = 2$, $\sigma(2) = 3$, $\sigma(3) = 4$, $\sigma(4) = 1$, $\sigma$ is a permutation. It can be shown as

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$$

The general notation is

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ \sigma(1) & \sigma(2) & \sigma(3) & \sigma(4) \end{pmatrix}$$

Functional compositions of permutation functions are also permutations. These compositions are called products of permutations. Say, there is another permutation $\rho : A \to A$. Then $\tau = \rho \sigma$, their product, defined by $\tau(a_j) = \rho(\sigma(a_j))$, $a_j \in A$. For a $\rho = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}$. Then, say $\tau(2) = \rho(\sigma(2)) = \rho(3) = 1$.

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$$

Note that here we used the notation where the rightmost permutation is applied on the set first. This is a convention. There is also the other convention where the leftmost permutation is applied on the set first.

Because the first line of the permutation expression is redundant there is a "one-line permutation notation" in which $\sigma = \begin{pmatrix} 2 & 3 & 4 & 1 \end{pmatrix}$.

Some mathematicians thought that even one-line notation is redundant and they use (disjoint) cycle notation. A cycle is a permutation in which every element goes to the element on its right, and the rightmost element goes to the first element.

$$(a \ b \ \cdots \ z) = (a \ b \ \cdots \ z) = (b \ c \ \cdots \ a) = (z \ a \ \cdots \ y)$$

Orbits are subsequences of the set on which a permutation is applied such that when one starts from
an element of the orbit an applies the permutation successively one passes through all elements in the orbit
returns to the first element. Every permutation can be decomposed into products disjoint cycles. Say
\[ \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 1 & 6 & 4 & 5 \end{pmatrix} \]

\( \sigma \) is applied on the set \( A = \{1, 2, 3, 4, 5, 6\} \). Starting from \( \{1, 3\} \) 1 \( \rightarrow \) 3 \( \rightarrow \) 1. Starting from \( \{2\} \) 2 \( \rightarrow \) 2.
Starting from \( \{4, 5, 6\} \) 4 \( \rightarrow \) 6 \( \rightarrow \) 5 \( \rightarrow \) 4. \( A \) is decomposed into three subsets \( A = \{1, 3\} \cup \{2\} \cup \{4, 5, 6\} \).
Under the permutation operation \( \sigma \) these subsets are orbits. As can be seen different orbits do not have
common elements. In the cycle notation each disjoint operation is shown between parentheses.

\[ \sigma = \begin{pmatrix} 1 & 3 \\ 2 \end{pmatrix} \begin{pmatrix} 4 & 6 & 5 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 6 & 5 & 4 \end{pmatrix} = \begin{pmatrix} 5 & 4 & 6 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 2 \end{pmatrix} \]

Because the orbits are disjoint the order in which they are written is not important. Finally, the orbits
that only have a single element are omitted.

\[ \sigma = \begin{pmatrix} 1 & 3 \\ 4 & 6 & 5 \end{pmatrix} \]

When they are omitted, there is an ambiguity on which set the permutation is applied. This cycle
notation is how group elements are stored in the computer algebra system SAGE. When the user creates a
group, its elements are calculated and stored in the form of permutation operators on the computer memory.
That is how the computer applies the abstract group operation on group elements.

Let us go back to Klein-four group, \( V_4 \), and see how its elements can be seen as permutations. First,
label the elements via numbers

<table>
<thead>
<tr>
<th>*</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>*</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1*1</td>
<td>1*2</td>
<td>1*3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2*1</td>
<td>2*2</td>
<td>2*3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3*1</td>
<td>3*2</td>
<td>3*3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4*1</td>
<td>4*2</td>
<td>4*3</td>
</tr>
</tbody>
</table>

Now, we want to express the elements as permutations. Imagine that each element of the Cayley table
is a permutation, \( \sigma_g \), which is a product of two permutations, \( \sigma_{g_1}, \sigma_{g_2} \), on the set \( G, g, g_1, g_2 \in G \).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_1 )</td>
<td>( \sigma_1(1) )</td>
<td>( \sigma_1(2) )</td>
<td>( \sigma_1(3) )</td>
<td>( \sigma_1(4) )</td>
</tr>
<tr>
<td>( \sigma_2 )</td>
<td>( \sigma_2(1) )</td>
<td>( \sigma_2(2) )</td>
<td>( \sigma_2(3) )</td>
<td>( \sigma_2(4) )</td>
</tr>
<tr>
<td>( \sigma_3 )</td>
<td>( \sigma_3(1) )</td>
<td>( \sigma_3(2) )</td>
<td>( \sigma_3(3) )</td>
<td>( \sigma_3(4) )</td>
</tr>
<tr>
<td>( \sigma_4 )</td>
<td>( \sigma_4(1) )</td>
<td>( \sigma_4(2) )</td>
<td>( \sigma_4(3) )</td>
<td>( \sigma_4(4) )</td>
</tr>
</tbody>
</table>

From which we can deduce

\[
\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix},
\]

\[
\sigma_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}, \quad \sigma_4 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}.
\]

These four permutations are a way to represent elements of \( V_4 \). \( V_4 \cong (\{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}, \times) \) (where \( \times \) is the permutation product).

### 3.1.10 Group Representations

A **representation** of a group is a homomorphism from the group elements to the general linear operators of dimension \( N \). \( \phi : G \to GL(V) \) (first \( G \) is the abstract group, \( GL \) means general linear group and \( V \) is a vector space on which the linear operator acts, it is called the **representation space** of which dimension is the **dimension of representation**). A linear operator is a linear transformation on the vector space.

When a basis \( \{v_i\} \) is chosen in \( V \) the linear operators can be expressed as matrices. Basically, a representation of a group is an actualization of an abstract group using matrices, instead of numbers or labels.

Representations are useful to investigate group theoretic problem using linear algebra, or the other way around, to investigate some linear algebra problems using group theory, which is what we are doing in our work.

If the matrix corresponding to the group element \( g \) is \( \Gamma(g) \), homomorphism implies

\[
\Gamma(g_1) \Gamma(g_2) = \Gamma(g_1 \ast g_2)
\]

where the product on the left hand side (LHS) is a matrix multiplication.
Some simple properties of representations are

- \( \Gamma (g^k) = \Gamma^k (g) \) which also implies \( \Gamma (g^{-1}) = \Gamma^{-1} (g) \)

- \( \Gamma (e) = I \)

Just like the trivial group is being a homomorphism of all groups \( \Gamma (g) = I_{N \times N} \ \forall g \in G \) is a **trivial representation** for all groups, where \( I_{N \times N} \) is the \( N \times N \) dimensional identity matrix. A special case of trivial representation is choosing all \( \Gamma (g) = 1 \), which is called **identical representation**. Trivial and identical representations contain absolutely no information about the group structure.

Another simple representation is the **parity representation**. Group elements can be represented as permutation operations. And every permutation operation on a set can be expressed as a combination of permutations on only two elements of the set (swaps). For example \( \sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} \) from the previous section can be expressed as \( \sigma_2 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \), which has even number of swaps, and hence has an even parity.

Consider a map from \( \Gamma : G \rightarrow \{-1, +1\} \), where \( \sigma (g) \) with even parity goes to \(+1\) and with odd parity goes to \(-1\). This simple representation will carry non-zero information about the group.

Group representations can be used to study symmetries. Think of the set of points \( P \in \mathbb{R}^2 \) in cartesian coordinate system, that are on a square centered around the coordinate center. The representation of the group element “90° rotation” is the matrix that rotates a point 90° around the coordinate center which is \( \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \).

A representation of a group can be found by figuring out the matrices corresponding the linear symmetry operations which the group elements represent.

Let us investigate some possible representations of cyclic groups, \( \mathbb{Z}_n \). \( \mathbb{Z}_n = (\mathbb{Z}_n, +) \) where \( 1^n = 0 \), \( \mathbb{Z}_n = \{1|1^n\} \). Which means that when the generator of the group, 1, is applied to itself \( n \) times one gets the identity element. If we interpret the group generator as a transformation under which a geometrical object is symmetric, that object can be a regular \( n \)-sided 2D shape at the coordinate center, and the transformation corresponding to the group generator is a rotation by \( 2\pi/n \). The 2D rotation corresponding to this angle is

\[
\Gamma (1) = \begin{pmatrix} \cos \left( \frac{2\pi}{n} \right) & -\sin \left( \frac{2\pi}{n} \right) \\ \sin \left( \frac{2\pi}{n} \right) & \cos \left( \frac{2\pi}{n} \right) \end{pmatrix}
\]

\( k \)-th element of the group is \( \Gamma (k) = \Gamma (1^k) = \Gamma (1)^k \), therefore
\[ \Gamma (k) = \begin{pmatrix} \cos \left( \frac{2\pi}{n} k \right) & -\sin \left( \frac{2\pi}{n} k \right) \\ \sin \left( \frac{2\pi}{n} k \right) & \cos \left( \frac{2\pi}{n} k \right) \end{pmatrix} \]

This defines a map from \( Z_n \) to \( GL(V) \) where \( V = \mathbb{R}^2 \), the two dimensional cartesian system. The exact shape of the geometric object is irrelevant. We are dealing with its abstraction as “an object that is symmetric under a \( \frac{2\pi}{n} \) rotation”.

Group representations are not unique. A group can have many representations, with different dimensions.

Just like \( Z_n \cong (Z_n, +) \), \( Z_n \) is also isomorphic to \( \{ z^\nu \} \) where \( z \in \mathbb{C} \) and multiplication as the group operation. According to the \( z^n = 1 \) condition a possible value for \( z \) is \( z = e^{\frac{2\pi}{n}} \). Actually \( z \) can be any of \( e^{\frac{2\pi}{n} j} \) where \( j = 0 \ldots n - 1 \). For each \( z_j = e^{\frac{2\pi}{n} j} \) \( z_j^n = 1 \).

Now, we can construct a diagonal matrix and fill its diagonal elements with \( z_j \)'s:

\[ \Gamma (1) = \begin{pmatrix} e^{\frac{2\pi}{n} j_1} & 0 & 0 \\ 0 & e^{\frac{2\pi}{n} j_2} & 0 \\ 0 & 0 & \cdots \end{pmatrix} \]

using this generator the rest of the representation matrices will be

\[ \Gamma (k) = \begin{pmatrix} e^{\frac{2\pi}{n} j_1} & 0 & 0 \\ 0 & e^{\frac{2\pi}{n} j_2} & 0 \\ 0 & 0 & \cdots \end{pmatrix} \]

Note that for \( k = n \) \( e^{\frac{2\pi}{n} j} = 1 \). This expression for \( \Gamma \) can produce many different representations of arbitrary dimensions for the group \( Z_n \) according to the choices for \( j \), and the dimensions of the matrices.

There are even more ways of producing representations. Two different representations \( \Gamma' \) and \( \Gamma \) are called equivalent when they are related with a similarity transformation, \( \Gamma' (g) = A^{-1} \Gamma (g) A \) for all \( g \in G \), where
A is an invertible matrix with the same dimensions as $\Gamma$. Equivalent representations are representations of the same group.

Because equivalent representations are related with similarity transformations their traces are equal.

$$\text{Tr} (\Gamma' (g)) = \text{Tr} (A^{-1} \Gamma (g) A) = \text{Tr} (\Gamma (g) A A^{-1}) = \text{Tr} (\Gamma (g)).$$

The rotation matrix representation and complex diagonal representation are equivalent via the similarity operation $T = \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}$ for which $T^{-1} = \frac{1}{2i} \begin{pmatrix} 1 & i \\ -1 & i \end{pmatrix}$.

$$\begin{pmatrix} \cos \left( \frac{2\pi}{n} k \right) & -\sin \left( \frac{2\pi}{n} k \right) \\ \sin \left( \frac{2\pi}{n} k \right) & \cos \left( \frac{2\pi}{n} k \right) \end{pmatrix} = \frac{1}{2i} \begin{pmatrix} 1 & i \\ -1 & i \end{pmatrix} \begin{pmatrix} e^{i \frac{2\pi}{n} k} & 0 \\ 0 & e^{-i \frac{2\pi}{n} k} \end{pmatrix} \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}.$$

We can apply a similarity transformation to the matrices of known representations to produce new ones.

$$\Gamma (k) = S^{-1} \begin{pmatrix} \cos \left( k \frac{2\pi}{n} \right) & -\sin \left( k \frac{2\pi}{n} \right) \\ \sin \left( k \frac{2\pi}{n} \right) & \cos \left( k \frac{2\pi}{n} \right) \end{pmatrix} S$$

and

$$\Gamma (k) = T^{-1} \begin{pmatrix} e^{i \frac{2\pi}{n} k} & 0 & 0 \\ 0 & e^{i \frac{2\pi}{n} k} & 0 \\ 0 & 0 & \ddots \end{pmatrix} T$$

are also representations of $\mathbb{Z}_n$. The similarity transformation of the representation matrices actually means that we are applying coordinate transformation on the basis vectors $\{ v_i \}$ in $\mathcal{V}$, $v_i \rightarrow Sv_i$.

Another way of producing new representation from known ones is creating block matrices

$$\Gamma (k) = \begin{pmatrix} S^{-1} \begin{pmatrix} \cos \left( k \frac{2\pi}{n} \right) & -\sin \left( k \frac{2\pi}{n} \right) \\ \sin \left( k \frac{2\pi}{n} \right) & \cos \left( k \frac{2\pi}{n} \right) \end{pmatrix} S & 0 & 0 \\ 0 & T^{-1} \begin{pmatrix} e^{i \frac{2\pi}{n} k} & 0 & 0 \\ 0 & e^{i \frac{2\pi}{n} k} & 0 \\ 0 & 0 & \ddots \end{pmatrix} T & 0 \\ 0 & 0 & \ddots \end{pmatrix}$$

As a more general expression, if $\Gamma_i$ are different representation of a group then
\[
\Gamma (g) = \begin{pmatrix}
\Gamma_1 (g) & 0 & 0 & 0 \\
0 & \Gamma_2 (g) & 0 & 0 \\
0 & 0 & \ddots & 0 \\
0 & 0 & 0 & \Gamma_d (g)
\end{pmatrix}
\]
is another representation. This operation of combining different representation elements on blocks of a different matrix is called the **direct sum** which is shown as

\[
\Gamma (g) = \Gamma_1 (g) \oplus \Gamma_2 (g) \oplus \cdots \oplus \Gamma_d (g).
\]

When we construct a bigger representation from smaller ones, we are not adding, generating or acquiring new information about the group that is not there in the smaller representations. We are just reproducing previously known properties.

We saw constructing bigger representations from smaller ones. The opposite way is also possible. One can decompose a representation into direct sums of smaller representations. To be able to do that we need to put the matrix into a block diagonal form using similarity transformations.

A representation \( \Gamma \), which is equivalent to any of its similarity transformations \( S T S^{-1} \), cannot be further decomposed into smaller representations is called an **irreducible representation** (or irrep for short). A representation is either an irreducible representation or it (or at least one of its similarity transformations) can be decomposed as a direct sum of irreducible representations. If some decomposition into a block diagonal form is possible for all the matrices in the representation than the representation is called **reducible**.

Finding irreducible representations of a group is useful because we can use theorems from group representation theory in our problems. Some of those theorems will be explained in the next chapters.

One dimensional representations such as identical, parity and complex number representations are always irreducible.

A representation is called a **faithful representation** if all matrices of a representation are distinct, each group element is mapped to a different matrix. Therefore a faithful representation is an isomorphism of the group, hence they are complex enough to capture the whole group structure.

A representation that is a homomorphism but not isomorphism will lose some details about the group. Because there are different group elements that are mapped to the same matrix. \( \Gamma (g_1) = \Gamma (g_2) \) where \( g_1 \neq g_2 \). Irreducible representations are not faithful in general, hence they do not carry the total information about the group. But on all irreducible representations of a group together capture all information carried by the group.
In our work we tend to use faithful representations because we want to get different quantum outcome states corresponding to different group elements.

Permutations can also be expressed as permutation matrices, matrices that are used to permute rows or columns of matrices, and that have all zero entries except one 1 in every row and column. The permutation matrix corresponding to the permutation $\sigma$

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}$$

is

$$\Sigma = \begin{pmatrix} \langle \sigma(1) \rangle \\ \langle \sigma(2) \rangle \\ \vdots \\ \langle \sigma(n) \rangle \end{pmatrix}$$

where $\langle k \rangle$ is a row vector of which components are 0 except the component at the index $k$. According to this construction $\Sigma_i$ corresponding to the elements of $V_4$ are

$$\Sigma_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$\Sigma_3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad \Sigma_4 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

A permutation matrix $\Sigma$ applied to a matrix $A$ on its right, $\Sigma A$, permutes the rows of $A$ (row $k$ goes to row $\sigma(k)$), and on its left, $A\Sigma$, permutes the columns.

The set $\{\Sigma_i\}$ is a faithful matrix representation of the group. It is the easiest faithful representation to compute.
Table 2: Symmetry operations corresponding to symmetries of equilateral triangle

<table>
<thead>
<tr>
<th>Operation</th>
<th>Inverse</th>
<th>Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e$</td>
<td>$e$</td>
<td>1</td>
</tr>
<tr>
<td>$(12)$ Fix 3. switch $1 \leftrightarrow 2$</td>
<td>$(12)$</td>
<td>2</td>
</tr>
<tr>
<td>$(23)$ Fix 1. $2 \leftrightarrow 3$</td>
<td>$(23)$</td>
<td>2</td>
</tr>
<tr>
<td>$(123)$ Rotate CW by $2\pi/3$</td>
<td>$(132)$</td>
<td>3</td>
</tr>
<tr>
<td>$(132)$ Rotate CCW</td>
<td>$(123)$</td>
<td>3</td>
</tr>
</tbody>
</table>

3.1.11 Symmetric Group

Symmetric group of degree $n$, $S_n$, is the group of all permutations on a set of $n$ elements. $n$ objects can be permuted in $n!$ different ways, therefore the order of symmetric group is $|S_n| = n!$.

Geometrically, for $n = 3$, $S_3$ is the group of all possible symmetries of an equilateral triangle on a plane, and for $n = 4$, $S_4$ is all symmetries of a tetrahedron. For higher degrees this interpretation becomes harder to visualize.

The interesting property of symmetric group is any group of order $n$ is a subgroup of the symmetric group of degree $n$. For example, the Klein-four group $V_4$ which is of order 4, $|V_4| = 4$, is a subgroup of $S_4$ which is of order $|S_4| = 4! = 24$. 

Figure 16: Symmetries of a triangle. Its vertices are labeled in clock-wise direction as $\{1, 2, 3\}$.
3.1.12 Dihedral Group

Dihedral group of degree \( n \), \( D_n \), consists of rotations and reflections in the plane that leave an \( n \)-sided regular polygon invariant.

\( D_3 \) is a small non-abelian, non-cyclic groups and has an easy to visualize geometrical meaning. It is non-abelian because a rotation is different than a rotation after a reflection. Non-cyclic because a just using rotations we cannot get a reflection (and vice versa). These are the reasons why we used it in our work.

\( D_3 = S_3 \) hence the Figure 16 for symmetric group \( S_3 \) applies to \( D_3 \) too.

One can generate all group elements of \( D_n \) using only two generators. One generator for a \( 2\pi/n \) rotation, \( r \), and one for reflection \( s \). \( D_n = \{ r, s | r^n, s^2, (sr)^2 \} \).

\( D_3 = \{ e, r, r^2, s, sr, sr^2 \} \), \( D_4 = \{ e, r, r^3, s, sr, sr^2, sr^3 \} \) etc.

A representation that comes from the geometrical meaning can be: \( \Gamma (s) = V \), \( \Gamma (r) = U \)

\[
V = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad U = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad \theta = \frac{2\pi}{n}
\]

where \( V \) is a reflection around \( y \)-axis. (If upper left element was chosen as \(-1\) then it would be a reflection around \( x \)-axis.)

Then representation elements become: \( \{ I, U, U^2, V, VU, VU^2 \} \)

As can be seen, cyclic group of order 3, \( Z_3 = \{ r | r^3 \} \), is a subgroup of \( D_3 \). But \( D_3 \) has an additional non-commuting generator (reflection) that generates a more complicated structure.

Another interpretation of \( D_3 \) can be the symmetries of complex plane. A rotation by \( 2\pi/n \), \( r \), is a multiplication with \( e^{2\pi i/n} \) and a reflection around \( x \)-axis is complex conjugation \( z \rightarrow \bar{z} \), \( c. \) \( D_n = \{ r, c | r^n, c^2, (sc)^2 \} \).

3.2 Group Representation Theory

Let us dive into deeper concepts of group representations that we used in our work. We saw that a group has infinitely many different representations. Representation theory is about classification of these representations up to isomorphisms.

A theorem shows that any representation is equivalent to a unitary representation.

Say \( \Gamma (g) \) is a given representation. Construct a hermitian matrix

\[
H = \sum_{g \in G} \Gamma (g) \Gamma ^\dagger (g)
\]

As quantum physicists we know it well that hermitian matrices can be diagonalized into real diagonal
matrices by a unitary transformation $U$. Let us call the diagonalized version $D = U^{-1}HU$.

$$D = \sum_g U^\dagger \Gamma (g) \Gamma^\dagger (g) U$$

$$= \sum_g (U^\dagger \Gamma (g) U) (U^\dagger \Gamma (g) U)^\dagger$$

$$= \sum_g \tilde{\Gamma} (g) \tilde{\Gamma}^\dagger (g)$$

Now define $\rho$ as a similarity transformation of $\tilde{\Gamma} (g)$

$$\rho (g) = D^{-1/2} \tilde{\Gamma} (g) D^{1/2}$$

$$\rho (g) \rho^\dagger (g) = \left( D^{-1/2} \tilde{\Gamma} (g) D^{1/2} \right) \left( D^{1/2} \tilde{\Gamma}^\dagger (g) D^{-1/2} \right)$$

$$= D^{-1/2} \tilde{\Gamma} (g) D \tilde{\Gamma}^\dagger (g) D^{-1/2}$$

$$= D^{-1/2} \tilde{\Gamma} (g) \left( \sum_{g'} \tilde{\Gamma} (g') \tilde{\Gamma}^\dagger (g') \right) \tilde{\Gamma}^\dagger (g) D^{-1/2}$$

$$= \sum_{g'} D^{-1/2} \tilde{\Gamma} (g) \tilde{\Gamma} (g') \left( \tilde{\Gamma} (g) \tilde{\Gamma} (g') \right)^\dagger D^{-1/2}$$

$$= \sum_{g'} D^{-1/2} \tilde{\Gamma} (g) \tilde{\Gamma} (g') \left( \tilde{\Gamma} (g) \tilde{\Gamma} (g') \right)^\dagger D^{-1/2}$$

$\tilde{\Gamma} (g) \tilde{\Gamma} (g')$ is another matrix in the representation and due to rearrangement theorem $\tilde{\Gamma} (g * g')$ goes over all matrices in the representation.

$$\rho (g) \rho^\dagger (g) = D^{-1/2} \sum_{g'} \tilde{\Gamma} (g * g') \tilde{\Gamma}^\dagger (g * g') D^{-1/2}$$

$$= D^{-1/2} DD^{-1/2}$$

$$= I$$

Therefore $\rho (g)$ is a unitary representation and the similarity transformation from $\Gamma$ to $\rho$ is

$$\rho (g) = D^{-1/2} U^{-1} \Gamma (g) UD^{1/2}$$
Thanks to this result we can always assume that we are dealing with unitary representations without losing generality.

3.2.1 Schur’s Lemma 1

Schur’s first lemma says that a matrix, $M$, that commutes with all matrices that belong to an irreducible representation, $\Gamma (g)$, is (either zero or) a constant times identity matrix. If $[M, \Gamma (g)] = 0 \ \forall g \in G$ then $M = cI$.

Start with

$$M \Gamma (g) = \Gamma (g) M$$

therefore

$$\Gamma ^\dagger (g) M ^\dagger = M ^\dagger \Gamma ^\dagger (g) \Gamma (g) \Gamma ^\dagger (g) = \Gamma (g) M ^\dagger \Gamma ^\dagger (g) \Gamma (g)$$

$$M ^\dagger \Gamma (g) = \Gamma (g) M ^\dagger$$

where we assumed that $\Gamma$ is unitary. We see that any linear combination of $M$ and $M ^\dagger$ will also commute with all representation matrices. Choose two combinations that are hermitian

$$H_1 = M + M ^\dagger, \quad H_2 = i (M - M ^\dagger)$$

Showing that $H_1$ and $H_2$ are multiples of identity also implies that $M = cI$, because

$$M = \frac{1}{2} (H_1 - iH_2).$$

$H$ is a hermitian matrix which can be diagonalized using $U$. $D = U^{-1} H U$. And $[H, \Gamma (g)] = 0$. Apply the symmetry transformation via $U$ to the commutation relation

$$U ^\dagger H \Gamma (g) U = U ^\dagger \Gamma (g) H U$$

$$U ^\dagger H U U ^\dagger \Gamma (g) U = U ^\dagger \Gamma (g) U U ^\dagger H U = \left( U \Gamma (g) U \right) ^\dagger U ^\dagger H U$$

$$D \tilde{\Gamma} (g) = \tilde{\Gamma} (g) D$$

$D$ being a diagonal matrix means that $D_{mn} = \delta_{mn} d_m$. Express the matrix multiplication on the left side
\[
\left(D\hat{\Gamma}(g)\right)_{mn} = \sum_k D_{mk}\hat{\Gamma}(g)_{kn} \\
= \sum_k \delta_{mk}d_m\hat{\Gamma}(g)_{kn} \\
= d_m\hat{\Gamma}(g)_{mn}
\]

Doing the same for the right hand side gives

\[
\left(\hat{\Gamma}(g)D\right)_{mn} = d_n\hat{\Gamma}(g)_{mn}
\]

Hence \(d_m\hat{\Gamma}(g)_{mn} = d_n\hat{\Gamma}(g)_{mn}\)

\[
\hat{\Gamma}(g)_{mn}(d_m - d_n) = 0
\]

- If \(d_m \neq d_n\) for all \(m, n\) then all off-diagonal entries of \(\hat{\Gamma}\) has to be zero, hence \(\hat{\Gamma}\) is reducible, contradicting with our assumption.

- If only \(l\) of the \(d_i\) are equal to each other, and the rest are distinct (use a reordering so that \(d_i = d_j\) for \(i = 1..l\)). Then \(\hat{\Gamma}(g)_{mn} = 0\) for \(d_m \neq d_n\). Which means all matrices are in this form

\[
\hat{\Gamma}(g) = \begin{pmatrix}
\rho(g) & 0 \\
0 & \delta(g)
\end{pmatrix}
\]

where \(\rho(g)\) is an \(l \times l\) matrix, and \(\delta(g)\) is a diagonal matrix (corresponding to \(d_m = d_n\) when \(m = n\) hence \(\hat{\Gamma}(g)\) is reducible.

- if \(d_m = d_n\) for all diagonal elements then \(\hat{\Gamma}(g)\) is irreducible (there will not be a \(\delta(g)\))

\(d_m = d_n = d\) for all \(m, n\) means that \(D = dI\), hence \(H_1 = dI\). Similarly \(H_2 = cI\) and hence \(M = cI\).

### 3.2.2 Schur’s Lemma 2

Schur’s first lemma said that, for a given irreducible representation \(\Gamma(g)\), if \(M\Gamma(g) = \Gamma(g)M\) holds for all \(g\) then \(M = cI\). **Schur’s second lemma** is about two unequivalent irreducible representations, \(\Gamma^{(i)}(g)\) and \(\Gamma^{(j)}(g)\). They can have different dimensions, \(l_i\) and \(l_j\). For an \(l_i \times l_j\) dimensional matrix \(M\), if \(M\Gamma^{(i)}(g) = \Gamma^{(j)}(g)M\) then \(M = 0\). Start by taking the dagger
\[
\Gamma^{(i)}(g) M^\dagger = M^\dagger \Gamma^{(j)}(g) \\
\Gamma^{(i)-1}(g) M^\dagger = M^\dagger \Gamma^{(j)-1}(g) \\
M \Gamma^{(i)-1}(g) M^\dagger = MM^\dagger \Gamma^{(j)-1}(g)
\]

\[
\Gamma^{(j)-1}(g) = \Gamma^{(i)}(g^{-1}) \quad \text{and} \quad \Gamma^{(j)-1}(g) = \Gamma^{(j)}(g^{-1}) \quad \text{for all} \quad g \in G \quad \text{where} \quad g^{-1} \in G \text{ too.}
\]

Therefore \(\Gamma^{(i)}(g)\) = \(\Gamma^{(j)}(g) M \Rightarrow M \Gamma^{(i)-1}(g) = \Gamma^{(j)-1}(g) M\). Which gives

\[
\Gamma^{(j)-1}(g) M M^\dagger = MM^\dagger \Gamma^{(j)-1}(g)
\]

which implies \([MM^\dagger, \Gamma^{(j)}(g)] = 0\). According to the first lemma \(MM^\dagger = cI\).

For the case \(l_i = l_j\). If \(c \neq 0\)

\[
det(MM^\dagger) = det(cI) \\
det(M)^2 = c^{l_j}
\]

because \(det M \neq 0 \quad M^{-1}\) exists. \(M \Gamma^{(i)}(g) = \Gamma^{(j)}(g) M \Rightarrow \Gamma^{(i)}(g) = M^{-1} \Gamma^{(j)}(g) M\) hence irrep-\(i\) is equivalent to irrep-\(j\).

If \(c = 0\) then

\[
MM^\dagger = 0 \\
\sum_k M_{mk} (M^\dagger)_{kn} = 0 \\
\sum_k M_{mk} M^*_{nk} = 0 \\
m = n \Rightarrow \sum_k |M_{mk}|^2 = 0 \\
\Rightarrow M_{mk} = 0
\]

for all \(m, k\) hence \(M = 0\). For the non-equivalent representations of the same dimension \(M = 0\).

For the case \(l_i \neq l_j\), assume \(l_i < l_j\), embed \(M\) (which is an \(l_j \times l_i\) dimensional matrix) in an \(l_j \times l_j\) matrix we call \(N\). In block diagrams
\[ N = \begin{pmatrix} M & 0 \end{pmatrix}, \quad N^\dagger = \begin{pmatrix} M^\dagger \\ 0 \end{pmatrix} \]

\[ MM^\dagger = NN^\dagger = cI_{l_i \times l_j}. \]

Because of the 0 in the diagonal, \( \det N = 0 \). \( \det NN^\dagger = \det MM^\dagger = (\det N)^2 = c^{l_i} \). Therefore, \( c = 0 \) and hence \( N = 0 \) and \( M = 0 \). Again, for the case of non-equivalent irreps having different dimensions, \( M = 0 \).

### 3.2.3 Great Orthogonality Theorem

**Great orthogonality theorem** is helpful in figuring out whether a given representation is irreducible or not and the number of non-equivalent representations of a group.

Say \( \Gamma^{(i)}(g) \) and \( \Gamma^{(j)}(g) \) are two unitary irreducible representations of \( G \), then

\[
\sum_{g \in G} \Gamma^{(i)\dagger}(g)_{\mu\nu} \Gamma^{(j)}(g)_{\alpha\beta} = \frac{|G|}{l_i} \delta_{ij} \delta_{\mu\alpha} \delta_{\nu\beta}
\]

where \( \Gamma^{(i)}(g)_{\mu\nu} \) is the entry of the matrix \( \Gamma^{(i)}(g) \) at the indices \( (\mu, \nu) \) and \( l_i \) is the dimension of \( \Gamma^{(i)}(g) \).

For the case when representations are different, \( i \neq j \), let \( X \) be any \( l_j \times l_i \) dimensional matrix. Define another matrix with the same dimensions

\[ M = \sum_{g \in G} \Gamma^{(j)}(g) X \Gamma^{(i)}(g^{-1}) \]

\[
\Gamma^{(j)}(g_0) M = \sum_{g} \Gamma^{(j)}(g_0) \Gamma^{(j)}(g) X \Gamma^{(i)}(g^{-1})
\]

\[
= \sum_{g} \Gamma^{(j)}(g_0) \Gamma^{(j)}(g) X \Gamma^{(i)}(g^{-1}) \Gamma^{(i)}(g_0^{-1}) \Gamma^{(i)}(g_0)
\]

\[
= \sum_{g} \Gamma^{(j)}(g_0 \ast g) X \Gamma^{(i)}(g_0^{-1}) \Gamma^{(i)}(g_0) \Gamma^{(i)}(g_0)
\]

\[
= \sum_{g} \Gamma^{(j)}(g) X \Gamma^{(i)}(g^{-1}) \Gamma^{(i)}(g_0) \quad \text{(by rearrangement theorem)}
\]

\[ = M \Gamma^{(i)}(g_0) \]

Using Schur’s second lemma \( M = 0 \). Choose \( X \) so that
\[ X_{\lambda\lambda'} = \begin{cases} 
0 & \lambda \neq \beta \text{ or } \lambda' \neq \nu \\
1 & \lambda = \beta \text{ and } \lambda' = \nu 
\end{cases} \]

which gives

\[ M = 0 = \sum_{g \in G} \Gamma^{(j)}(g) X \Gamma^{(i)}(g^{-1}) \]

\[ = \sum_{g} \sum_{\lambda} \sum_{\lambda'} \left[ \Gamma^{(j)}(g) \right]_{\alpha\lambda} \left[ X_{\lambda\lambda'} \right] \left[ \Gamma^{(i)}(g^{-1}) \right]_{\lambda'\mu} \]

\[ = \sum_{g} \left[ \Gamma^{(j)}(g) \right]_{\alpha\beta} \left[ \Gamma^{(i)}(g^{-1}) \right]_{\nu\mu} \]

\[ = \sum_{g} \left[ \Gamma^{(i)}(g^{-1}) \right]_{\nu\mu} \left[ \Gamma^{(j)}(g) \right]_{\alpha\beta} \]

For the case \( i = j \) where representations are the same

\[ M = \sum_{g \in G} \Gamma^{(i)}(g) X \Gamma^{(i)}(g^{-1}) \]

as before

\[ \Gamma^{(i)}(g_0) M = M \Gamma^{(i)}(g_0) \]

According to Schur’s Lemma for the irreducible \( \Gamma^{(i)} \) \( M = cI \). Choose an \( X \) of which all entries are 0 except \( X_{\nu\nu'} = 1 \).

\[ cI = \sum_{g \in G} \Gamma^{(i)}(g) X \Gamma^{(i)}(g^{-1}) \]

\[ c\delta_{\mu\mu'} = \sum_{g} \Gamma^{(i)}(g)_{\mu\nu} \Gamma^{(i)}(g^{-1})_{\nu'\mu'} \]

Set \( \mu = \mu' \) and sum over \( \mu \)
\[ c_{li} = \sum_{\mu} \sum_{g} \Gamma^{(i)}(g)_{\mu \nu} \Gamma^{(i)}(g^{-1})_{\nu' \mu} \]
\[ = \sum_{g} \sum_{\mu} \Gamma^{(i)}(g^{-1})_{\nu' \mu} \Gamma^{(i)}(g)_{\mu \nu} \]
\[ = \sum_{g} \left[ \Gamma^{(i)}(g^{-1}) \ast \Gamma^{(i)}(g) \right]_{\nu' \nu} \]
\[ = \sum_{g} \left[ \Gamma^{(i)}(e) \right]_{\nu' \nu} \]
\[ = \sum_{g} [I]_{\nu' \nu} \]
\[ = |G| \delta_{\nu' \nu} \]

hence

\[ c = \frac{|G|}{l_i} \delta_{\nu' \nu} \]

Therefore

\[ \frac{|G|}{l_i} \delta_{\nu' \nu} \delta_{\mu' \mu} = \sum_{g} \Gamma^{(i)}(g)_{\mu \nu} \Gamma^{(i)}(g^{-1})_{\nu' \mu'} \]
\[ = \sum_{g} \Gamma^{(i)}(g)_{\mu \nu} \Gamma^{(i)}(g)_{\mu' \nu'} \]

The great orthogonality theorem can be interpreted in terms of vectors. The representation matrices are 2 dimensional tensors (a 2D array of numbers) with indices \( \mu \) and \( \nu \). Imagine vertically stacking those matrices corresponding to different \( g \in G \) on top of each other and building a 3 dimensional tensor where the index related to height is \( g \). Call that tensor \( T_{g \mu \nu}^{(i)} \) where a choice of \( g \) corresponds to \( T_{g \mu \nu}^{(i)} = \Gamma^{(i)}(g)_{\mu \nu} \), and \( T_{g_1}^{(i)} = \Gamma^{(i)}(g_1) \). Fixing \( \mu \) and \( \nu \) will give columns from \( T_{g \mu \nu}^{(i)} \). Interpret those columns as vectors \( v \in \mathbb{C}^{|G|} \).
where the component-$g$ of the vector is $v^g_{(i;\mu,\nu)} = T^{(i)}_{g\mu\nu} = \Gamma^{(i)}(g)_{\mu\nu}$.

\[
v^g_{(i;\mu,\nu)} = \begin{pmatrix}
\Gamma^{(i)}(g_1)_{\mu\nu} \\
\Gamma^{(i)}(g_2)_{\mu\nu} \\
\vdots \\
\Gamma^{(i)}(g_{|G|})_{\mu\nu}
\end{pmatrix}
\]

Great orthogonality theorem says that $\langle v^g_{(i;\mu,\nu)} | v^g_{(j;\mu',\nu')} \rangle = \delta_{ij} \delta_{\mu\mu'} \delta_{\nu\nu'}$. See Figure 17 for a visual description.

Figure 17: Two towers represent two irreducible representations. Each floor is a representation matrix with $\mu, \nu, \alpha$ and $\beta$ being indices of matrix elements. Red and blue columns represent the vectors $v^g_{(i;\mu,\nu)}$. Great orthogonality theorem is about their dot product.

3.2.4 Characters of Representations

The character of a representation matrix is basically its trace.

\[
\chi^{(i)}(g) = \text{Tr}\left(\Gamma^{(i)}(g)\right)
\]

It is used to classify different representations. Equivalent representations have the same set of characters, because trace does not change under similarity transformations.

Say $\Gamma^{(i)}(g)$ and $\Gamma^{(j)}(g)$ are equivalent, namely $\Gamma^{(i)}(g) = S\Gamma^{(j)}(g)S^{-1}$ then
\[ \chi^{(i)}(g) = \text{Tr} \left( \Gamma^{(i)}(g) \right) \]
\[ = \text{Tr} \left( S \Gamma^{(j)}(g) S^{-1} \right) \]
\[ = \text{Tr} \left( \Gamma^{(j)}(g) S^{-1} S \right) \]
\[ = \text{Tr} \left( \Gamma^{(j)}(g) \right) \]
\[ = \chi^{(j)}(g) \]

For a chosen irreducible representation, matrices belonging to the same class have the same character. Say \( g_1, g_2 \in C \). Then there is a \( g \) for which \( g_1 = gg_2g^{-1} \).

\[ \Gamma(g_1) = \Gamma(gg_2g^{-1}) \]
\[ = \Gamma(g) \Gamma(g_2) \Gamma(g^{-1}) \]

Calculate the trace of both sides

\[ \chi(g_1) = \text{Tr} \left( \Gamma(g) \Gamma(g_2) \Gamma^{-1}(g) \right) \]
\[ = \text{Tr} \left( \Gamma(g_2) \right) \]
\[ = \chi(g_2) \]

Let us denote the character of the group elements that belong to the same class \( C_k \) as \( \chi(C_k) \).

### 3.2.5 Orthogonality Theorem for Characters

If two representations \( \Gamma^{(i)} \) and \( \Gamma^{(j)} \) are irreducible, \( N_k \) is the number of group elements in the conjugacy class \( C_k \) and \( m \) is the number of different classes, then

\[ \sum_k^{m} \chi^{(i)*}(C_k) \chi^{(j)}(C_k) N_k = |G| \delta_{ij} \]

This theorem can be derived from great orthogonality theorem
\[
\sum_{g \in G} \Gamma^{(i)}(g)_{\mu \nu} \Gamma^{(j)}(g)_{\alpha \beta} = \frac{|G|}{l_i} \delta_{ij} \delta_{\mu \alpha} \delta_{\nu \beta}
\]
to get traces from this expression let \( \nu = \mu \) and \( \beta = \alpha \) and sum over \( \mu \) and \( \alpha \)

\[
\sum_{\mu} \sum_{\alpha} \sum_{g \in G} \Gamma^{(i)}(g)_{\mu \mu} \Gamma^{(j)}(g)_{\alpha \alpha} = \sum_{\mu} \sum_{\alpha} |G| \delta_{ij} \delta_{\mu \alpha} \delta_{\nu \alpha}
\]

\[
\sum_{g} \chi^{(i)*}(g) \chi^{(j)}(g) = |G| \delta_{ij}
\]
Express the sum on the left hand side in terms of classes instead of group elements

\[
\sum_{g} \chi^{(i)*}(g) \chi^{(j)}(g) = \sum_{k} \chi^{(i)*}(C_k) \chi^{(j)}(C_k) N_k
\]
Because this is an orthogonality relation let us express it again in terms of vectors. Define a vector

\[
v^{(i)} = \begin{pmatrix}
\chi^{(i)}(C_1) \sqrt{N_1} \\
\chi^{(i)}(C_2) \sqrt{N_2} \\
\vdots \\
\chi^{(i)}(C_m) \sqrt{N_m}
\end{pmatrix}
\]
Orthogonality theorem for characters says

\[
\langle v^{(i)} | v^{(j)} \rangle = |G| \delta_{ij}
\]
Say there are \( q \) different irreducible representations. Then we have \( q \) orthogonal vectors in the \( m \) dimensional space. \( m \geq q \).

The number of times a representation \( \Gamma^{(p)} \) appears in the direct sum decomposition of a reducible representation \( \Gamma(g) \) is given by

\[
n_p = \frac{1}{|G|} \sum_{g} \chi(g) \chi^{(p)*}(g)
\]
In a block diagonal form \( \Gamma(g) \) can be shown as
\[ \Gamma (g) = \begin{pmatrix} \Gamma^{(1)} & & & & 0 \\ & \Gamma^{(1)} & & \cdots & 0 \\ & & \ddots & & \vdots \\ & & & \Gamma^{(1)} & 0 \\ 0 & & \cdots & & \Gamma^{(2)} \end{pmatrix} \]

where \( \Gamma^{(j)} \) appear \( n_j \) times. Say there are \( L \) irreducible representations that construct \( \Gamma (g) \).

\[
\Gamma (g) = \Gamma^{(1)}(g) \oplus \Gamma^{(1)}(g) \oplus \cdots \oplus \Gamma^{(1)}(g) \oplus \Gamma^{(2)}(g) \oplus \cdots \oplus \Gamma^{(L)}(g)
\]

The trace is distributed over direct sum.

\[
\chi(g) = \text{Tr}(\Gamma(g)) = \text{Tr}
\left( \bigoplus_{l} \Gamma^{(l)} (g) \right) = \sum_{l} \text{Tr}
\left( \Gamma^{(l)} (g) \right) = \sum_{j} n_j \text{Tr}
\left( \Gamma^{(j)} (g) \right) = \sum_{j} n_j \chi^{(j)}(g)
\]

Multiply both sides with \( \chi^{(p)*}(g) \) and sum over \( g \)

\[
\sum_{g} \chi^{(p)*}(g) \chi(g) = \sum_{j} n_j \sum_{g} \chi^{(p)*}(g) \chi^{(j)}(g) = \sum_{j} n_j \sum_{k} N_{k} \chi^{(p)*}(C_k) \chi^{(j)}(C_k)
\]

apply the orthogonality theorem to the right hand side

\[
\sum_{g} \chi^{(p)*}(g) \chi(g) = \sum_{j} n_j \delta_{pj} = n_p
\]
3.2.6 The Regular Representation

The regular representation of a group is constructed with a modification on the Cayley table in which the header row has the group elements in an order \( g_1, \ldots, g_n \), and the order in the column header is chosen such that identity elements lie on the diagonal.

\[
\begin{array}{c|cccc}
\ast & g_1 & g_2 & \cdots & g_n \\
\hline
\ast & e & \cdot & \cdots & \cdot \\
l_{g} & \cdot & e & \cdot & \cdot \\
\vdots & \cdot & \cdot & \ddots & \cdot \\
g_{n} & \cdot & \cdots & \cdot & e \\
\end{array}
\]

From this table, the representation matrices \( \Gamma (g) \) corresponding to the element \( g \) is constructed by putting zero everywhere except to the entries that are equal to \( g \).

For example for Klein four group which already has identities on the diagonal.

\[
\begin{array}{cccc}
\ast & 1 & 2 & 3 & 4 \\
1 & 1 & 2 & 3 & 4 \\
V_4 = & 2 & 2 & 1 & 4 & 3 \\
3 & 3 & 4 & 1 & 2 \\
4 & 4 & 3 & 2 & 1 \\
\end{array}
\]

\[
\Gamma^{(\text{reg})} (g_1) = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{pmatrix}, \quad \Gamma^{(\text{reg})} (g_2) = \begin{pmatrix}
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 \\
\end{pmatrix}
\]

\[
\Gamma^{(\text{reg})} (g_3) = \begin{pmatrix}
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
\end{pmatrix}, \quad \Gamma^{(\text{reg})} (g_4) = \begin{pmatrix}
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
\end{pmatrix}
\]

which are strikingly similar to the permutation matrices corresponding to the permutation operators corresponding to the group elements.
\[ \chi^{\text{reg}}(g) = \Tr \left( \Gamma^{\text{reg}}(g) \right) = \begin{cases} |G| & g = e \\ 0 & g \neq e \end{cases} \]

because only for \( e \)’s are on the diagonal.

Using regular representation we can prove that the sum of squares of the dimensions of the inequivalent irreducible representations is equal to the group order. Start with

\[
n_i = \frac{1}{|G|} \sum_g \chi^{\text{reg}}(g) \chi^{(i)*}(g) \\
= \frac{1}{|G|} \sum_g |G| \delta_{ge} \chi^{(i)*}(g) \\
= \chi^{(i)*}(e) \\
= l_i
\]

In regular representation the number of times a representation appears in its direct sum decomposition is equal to the dimension of the representation.

\[ |G| = \sum_i n_i l_i = \sum_i l_i^2 \]

### 3.2.7 Character Tables

Character table is a way to represent the information related to the conjugacy classes, corresponding characters and irreducible representations of a group.

<table>
<thead>
<tr>
<th></th>
<th>( C_1 = C_e )</th>
<th>( N_2C_2 )</th>
<th>( N_3C_3 )</th>
<th>( \cdots )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Gamma^{(1)} )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>( \cdots )</td>
</tr>
<tr>
<td>( \Gamma^{(2)} )</td>
<td>( l_2 )</td>
<td>( \chi^{(2)}(C_2) )</td>
<td>( \chi^{(2)}(C_3) )</td>
<td>( \cdots )</td>
</tr>
<tr>
<td>( \Gamma^{(3)} )</td>
<td>( l_3 )</td>
<td>( \chi^{(3)}(C_2) )</td>
<td>( \chi^{(3)}(C_2) )</td>
<td>( \cdots )</td>
</tr>
<tr>
<td>( \vdots )</td>
<td></td>
<td></td>
<td></td>
<td>( \cdots )</td>
</tr>
</tbody>
</table>

First column contains the dimensions of representations. \( \chi(C_e) = \Tr (\Gamma^{(i)}(e)) = l_j \). Sum of their squares is equal to \( |G| \).

First row is 1 for all classes. This is the trivial or identity representation where all group elements are mapped to 1, of which trace is 1.
According to the orthogonality theorem for characters, after weighted by $N_k$, rows are orthogonal. Columns are orthogonal too.

These properties can be used to complete a character table that is partially given.

### 3.2.8 Invariant Subspaces and Subrepresentations

In general, an **invariant subspace** $W \subset V$ of a linear operator $A : V \to V$ is a subspace all of its vectors are mapped into that subspace, $\{w | Aw \in W, \forall w \in W\}$.

$W$ is called $A$-**invariant**. If $W$ is $A$-invariant then $A|_W$ is a mapping for which $A|_W : W \to W$.

$\{0\}$ is always an invariant subspace because $A0 = 0$ for all linear maps. Similarly $V$ is an invariant subspace of itself. These are called **trivially** invariant subspaces. There can be linear maps which do not have any non-trivial subspaces.

Say a group $G$ has a representation $\Gamma$ from the group elements to the vector space $V$, $\Gamma : G \to V$, where $\Gamma (g)$ is a linear map $\Gamma (g) : V \to V$. If a subspace $W \subset V$ is $\Gamma (g)$-invariant then $\Gamma|_W : G \to W$ is a **subrepresentation**.

We can rephrase the irreducibility and Schur’s lemma using subrepresentation language.

If a subspace $W$ has only trivial subrepresentations then $\Gamma|_W$ is called irreducible and if $W$ has non-trivial proper subrepresentations then $\Gamma|_W$ is a reducible representation which can be further decomposed into subrepresentations.

Schur’s lemma says a map $\phi : U \to V$ between irreducible representations is either a zero map ($U \to \{0\} \in V$, corresponding to the case of $M = 0$), or an isomorphism (corresponding to the case where two representations are equivalent).

If $W$ is an invariant subspace $A : V \to V$ then $V$ can be decomposed as $V = W \oplus W^\perp$. So, a vector $v \in V$ is either in $W$ or in the complementary subspace $W^\perp$. This also means that a representation on $V$, $\Gamma : G \to V$ can be expressed as the direct sum of two subrepresentations on $W$ and $W^\perp$.

$$\Gamma (g) = \Gamma|_W (g) \oplus \Gamma|_{W^\perp} (g).$$

In a similar sense if $V$, on which a representation $\Gamma$ acts, can be decomposed into all $\Gamma$-invariant subspaces

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_p = \bigoplus_i W_i$$

where $W_i \cap W_j = \{0\}$ for $i \neq j$. This decomposition corresponds to the direct sum decomposition of $\Gamma$ into its irreducible subrepresentations, $\Gamma|_{W_i}$. 
\[ \Gamma = \Gamma_{W_1} \oplus \Gamma_{W_2} \oplus \cdots \oplus \Gamma_{W_p} = \bigoplus_i^p \Gamma_{W_i} \]

where we can define \( \Gamma_{W_i} = \Gamma^{(i)} \)

\[ \Gamma(g) = \bigoplus_i^p \Gamma^{(i)}(g). \]

This decomposition of subspaces and irreducible representations will help us in calculating the eigenvalues of our Bell operators.

### 3.2.9 Group Actions

In the section on groups and permutations we saw an example of group elements acting on sets by permuting the elements of the set. That was one example of group action. In general we say that \( G \) acts on the set \( X \) if there is a map \( \alpha \)

\[ \alpha : G \times X \rightarrow X \]

which satisfies for \( e, g_1, g_2 \in G, x \in X \)

- identity: \( \alpha(e, x) = x \)
- compatibility: \( \alpha(g_1, \alpha(g_2, x)) = \alpha(g_1 * g_2, x) \)

The basic idea is to choose \(|G|\) functions \( f_g : X \rightarrow X \) with \( f_g(x) = y \) that satisfy identity and compatibility axioms. But because group elements themselves are abstract and are not functions (or groups can have infinitely many elements) it is shown in as \( \alpha(g, x) = y \).

The subset of \( X \) given by \( O_x = \{ \alpha(g, x) \mid g \in G \} \) is called an orbit. The orbit we mentioned in the permutation section was an example of this general concept of an orbit. And just like the permutation orbits do not have common elements, orbits in general are either distinct or identical. \( O_x = O_y \) or \( O_x \cap O_y = \emptyset \).

Therefore orbits partition \( X \),

\[ X = \bigcup_i O_i, \quad O_i \cap O_j = \begin{cases} O_i & i = j \\ \emptyset & i \neq j \end{cases} \]

where the indices indicate different orbits.

In our work the set \( X \) is the Hilbert space \( \mathcal{H} \), and the map is the application of a unitary representation of the group on a quantum state \( |\psi\rangle \in \mathcal{H}, \alpha(g, |\psi\rangle) = \Gamma^{(i)}(g) |\psi\rangle \) where \( \alpha : G \times \mathcal{H} \rightarrow \mathcal{H} \). We are interested in orbits, sequences of quantum states that we get from starting an initial state. We will interpret these
states as measurement outcome state, corresponding to some events. We will compare probabilities of joint
events produced by these sequences with their classical counterparts.

These were the group theory related concepts that we are going to use. Let us do some work next.

3.3 Bell Inequalities from Group Actions of Single-Generator Groups

In this first part of group theoretical Bell’s theorem studies we investigate cyclic sequences of quantum
states and operators that generate those sequences. We did not analyze the underlying group but we just
constructed a representation that acts on the Hilbert space of the combined parties, so that it will serve to
our purposes and generate a sequence that will lead Bell violations. The connection with the group theory
will become more apparent when we move to the scheme of the [26] where we do not use the swap operator
and explicitly choose the group and construct its representations.

3.3.1 Reproduction of CH Using a Cyclic Group

First we want to reproduce the CH inequality which is the only tight Bell inequality for the case of 2 parties,
2 measurement per party and 2 outcomes per measurement, $N = M = K = 2$.

We want to produce different measurement outcome states by the application of the same operator
successively on an initial state. We first define the “translation operator”. Its function is to “shift” the
computational basis elements:

$$ T |k\rangle = |k + 1 \mod K\rangle $$

where $|k\rangle$ is in computational basis. It sends $|0\rangle$ to $|1\rangle$, $|1\rangle$ to $|2\rangle$, ..., $|K - 1\rangle$ to $|0\rangle$.

$$ T = \begin{pmatrix}
0 & 0 & \cdots & 1 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & \ddots & 0
\end{pmatrix} $$

For $K = 2$, $T$ becomes

$$ T = \begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix} $$

which is the $\sigma_z$ matrix. The translation matrix is specific to the computational basis. It does the translation
$|k\rangle \rightarrow |k + 1\rangle$ (sum in modulo $K$) in computational basis. We could construct $T$ for different bases too. Say
\{u_i | i = 0..K - 1\} is an orthonormal basis, then

\[ T \alpha = |u_1\rangle \langle u_0| + |u_2\rangle \langle u_1| + \cdots + |u_{K-1}\rangle \langle u_{K-1}| \]

\[ = \sum_i |u_{i+1}\rangle \langle u_i| \]

does the translation in the \( u \)-basis.

Successive application of \( T \) to \( |0\rangle \) will give as this sequence of two elements, \( T |0\rangle = |1\rangle, T |1\rangle = |0\rangle \)

\[ |0\rangle \rightarrow |1\rangle \rightarrow |0\rangle \rightarrow \cdots \]

But we need 4 outcome states instead of 2. Because we are reproducing the CH scenario, there are \( M \times K = 4 \) different events (measurement outcomes) associated with a party. Say observables \( x_1 \) and \( x_2 \) are measured at the party \( X \). We can name the events as “first quantity having the first outcome” \( x_1 = 0 \), “first quantity having the second outcome”, \( x_1 = 1 \), “second quantity having the first outcome” \( x_2 = 0 \), “second quantity having the second outcome” \( x_2 = 1 \). Therefore, for each party we need these four states \( |x_1 = 0\rangle, |x_1 = 1\rangle, |x_2 = 0\rangle, |x_2 = 1\rangle \).

\( |0\rangle \) and \( |1\rangle \) can correspond to \( |x_1 = 0\rangle \) and \( |x_1 = 1\rangle \), hence we need one more basis, two more elements in our sequence.

We can achieve that by taking the square root of \( T \) operator. Define a \( U \) such that

\[ U^2 = T \]

If \( t_i \) are the eigenvalues and \( |t_i\rangle \) the corresponding eigenvectors, the spectral decomposition of \( T = \sigma_x \) is

\[ T = \sum_i t_i |t_i\rangle \langle t_i| \]

\[ = |x+\rangle \langle x+| - |x-\rangle \langle x-| \]

where \( \{ |x+\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, |x-\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \} \) is the \( x \)-basis.

A function of an operator is

\[ f(T) = \sum_i f(t_i) |t_i\rangle \langle t_i| \]

Therefore the square root of \( T \) is calculated by summing up the projection operators on its eigenspaces.
weighted by the square roots of its eigenvalues.

\[ U = \sqrt{1} |x+\rangle \langle x+| + \sqrt{-1} |x-\rangle \langle x-| \]

which is not a unique expression. \( \sqrt{1} \in \{1, -1\} \) and \( \sqrt{-1} \in \{i, -i\} \). We have four different options. We choose

\[ U = |x+\rangle \langle x+| + i |x-\rangle \langle x-| \]

which leads to Bell violations that we will later see.

Now, successive application of \( U \) on \(|0\rangle\) will give us a 4 state sequence two of which are \( U^2 |0\rangle = T |0\rangle = 1 \) and \( U^4 |0\rangle = T^2 |0\rangle = |0\rangle \). The other two are

\[
U |0\rangle = \frac{1}{\sqrt{2}} (e^{i\pi/4} |0\rangle + e^{-i\pi/4} |1\rangle) \equiv |v_0\rangle \\
U |1\rangle = \frac{1}{\sqrt{2}} (e^{-i\pi/4} |0\rangle + e^{i\pi/4} |1\rangle) \equiv |v_1\rangle
\]

These make the \( v \)-basis \( \{|v_0\rangle, |v_1\rangle\} \). The second quantity that will be measured on the parties.

The sequence due to successive application of \( U \) on \(|0\rangle\),

\[ |\psi_i\rangle = U^i |0\rangle, \]

is

\[ |0\rangle \rightarrow |v_0\rangle \rightarrow |1\rangle \rightarrow |v_1\rangle \rightarrow |0\rangle \rightarrow \cdots. \]

Our physical system is made of \( N = 2 \) parties that are \( K = 2 \) dimensional. 2 qubits make such as system.

Let us define the swap operator \( S : \mathbb{C}^K \otimes \mathbb{C}^K \rightarrow \mathbb{C}^K \otimes \mathbb{C}^K \) which swaps the factors of a tensor product

\[ S |j\rangle \otimes |k\rangle = |k\rangle \otimes |j\rangle \]

where \( |j\rangle, |k\rangle \) are members of \( z \)-basis. Any vector in \( \mathbb{C}^4 \) can be written as

\[ |\Phi\rangle = \sum_{j,k} c_{jk} |j\rangle \otimes |k\rangle. \]
Therefore when $S$ is applied on $|\Psi\rangle$

$$S|\Psi\rangle = \sum_{j,k} c_{jk} S|j\rangle \otimes |k\rangle$$

$$= \sum_{j,k} c_{jk} S|k\rangle \otimes |j\rangle$$

and if $|\Psi\rangle$ is separable then $S$ corresponds to swapping the quantum states belonging to the parties.

$$S|\psi\rangle \otimes |\phi\rangle = |\phi\rangle \otimes |\psi\rangle$$

Note that here we are not talking about a physical process of swapping actual quantum states. This is a mathematical tool that we use in generating our sequences.

Now we define the $B$ operator that is going to be applied to $|0\rangle \otimes |0\rangle$ to generate our sequence of states on two parties.

$$B = (U \otimes I) S$$

The general form $(U_1 \otimes U_2) S$ is the most general form of operators that preserve separability. We want to keep the separability because at every iteration of the sequence we want to have measurement outcome states for the joint system in the form $|a_m = i\rangle \otimes |b_n = j\rangle$ which corresponds to the joint event of “getting the outcome-$i$ for the measurement of quantity-$m$ on the first party and getting the outcome-$j$ for the measurement of quantity-$n$ on the second party”. The associated joint probability for both classical and quantum cases is $p(a_m = i, b_n = j)$.

If we get an entangled state, a state that cannot be separated into parties in $m$- and $n$-basis, then we cannot define these events.

The application of $U$ (in $B$) on only one party generates a sequence in which the states belonging to different parties are iterated alternatingly.
where the sequence is

\[|\psi_j\rangle|\psi_k\rangle \rightarrow |\psi_{k+1}\rangle|\psi_j\rangle \rightarrow |\psi_{j+1}\rangle|\psi_{k+1}\rangle \rightarrow |\psi_{k+2}\rangle|\psi_{j+1}\rangle \rightarrow \cdots\]

Let us determine our sequence on two parties by successively applying \(B\) on \(|0\rangle|0\rangle\)

\[
|0\rangle|0\rangle \rightarrow |v_0\rangle|0\rangle \rightarrow |v_0\rangle|v_0\rangle \\
|1\rangle|v_0\rangle \rightarrow |1\rangle|1\rangle \rightarrow |v_1\rangle|1\rangle \\
|v_1\rangle|v_1\rangle \rightarrow |0\rangle|v_1\rangle \rightarrow |0\rangle|0\rangle \rightarrow \cdots
\]

This is a sequence made of 8 different states, \(\{|\Psi_i\rangle |i = 0..7\}\). Each of them is a joint measurement outcome. The measurements on parties are either a measurement on \(z\)-basis (first observable, \(a_1\) on first party, \(b_1\) on second party) or \(v\)-basis (second observable, \(a_2\) on first party, \(b_2\) on second party).

A choice of quantum state of combined system, say \(|\phi\rangle\), gives us probabilities of occurrences of corresponding joint outcomes. For example \(p(v = 0, z = 0) = p(a_2 = 0, b_1 = 0) = |\langle\phi | v_0, 0 \rangle|^2\) or \(p(z = 0, v = 1) = p(a_1 = 0, b_2 = 1) = |\langle\phi | 0, v_1 \rangle|^2\) etc.

Our Bell expression is the sum of all probabilities that are produced by the sequence. We want to maximize this value with respect to \(|\phi\rangle\).

\[
\max_{|\phi\rangle} \sum_j |\langle\phi | \Psi_j\rangle|^2 = \max_{|\phi\rangle} \sum_{j=0}^{7} |\langle\phi | B^j | 0, 0 \rangle|^2 \\
= \max_{|\phi\rangle} \sum_{j=0}^{7} \langle\phi | B^j | 0, 0 \rangle \langle 0, 0 | B^{j\dagger} | \phi \rangle \\
= \max_{|\phi\rangle} \langle\phi \left( \sum_{j=0}^{7} B^j | 0, 0 \rangle \langle 0, 0 | B^{j\dagger} \right) |\phi\rangle
\]
\[
A = \sum_{j=0}^{7} B^j |0,0\rangle \langle 0,0| B^j
\]

We want the highest possible expectation value of the \( A \) operator

\[
\max_{|\phi\rangle} \langle \phi | A | \phi \rangle = \max \langle A \rangle
\]

Highest possible expectation value of an operator is its biggest eigenvalue, because an expectation value is a convex linear combination of its eigenvalues. Therefore we calculate the spectral decomposition of the \( A \) operator

\[
A = \sum_i a_i Q(a_i)
\]

where \( a_i \) are the distinct eigenvalues and \( Q(a_i) \) are projection operators onto subspaces corresponding to the eigenvalue \( a_i \), \( \mathcal{V}(a_i) \). \( \mathbb{C}^4 = \bigoplus_i \mathcal{V}(a_i) \). Writing the spectral decomposition in terms of projection operators rather than eigenstates allows us to deal with degenerate and non-degenerate cases using the same formulation.

Given the spectral decomposition, biggest \( a_i \), call it \( a^* \), is the maximum value that \( \langle A \rangle \) can have, and any \( |\phi\rangle \in \mathcal{V}(a^*) \) is a quantum state that will give the maximum quantum value for the Bell expression.

Say the eigenvalues of \( B \) are \( \{b_k\} \) and corresponding eigenstates are \( \{|b_k\}\), \( k = 0..3 \).

\[
B^2 = (U \otimes I) S (U \otimes I) S
\]
\[
= U \otimes U.
\]

Because \( U^4 = I \)

\[
B^8 = (B^2)^4
\]
\[
= (U \otimes U)^4
\]
\[
= U^4 \otimes U^4
\]
\[
= I \otimes I.
\]

Therefore \( b_k^8 = 1 \). \( b_k \)'s are 8th roots of unity, hence they are of the form \( \exp(i m_k \pi/4) \) where \( 0 \leq m_k \leq 7 \).
Now look at the relationship between the eigenvalues of \( A \) and \( B \).

\[
A |b_k\rangle = \sum_{j=0}^{7} B^j |0,0\rangle \langle 0,0| B^{\dagger j} |b_k\rangle \\
\quad = \sum_{j=0}^{7} B^j |0,0\rangle (b^*_k)^j (0,0 | b_k) \\
\]

Remember that \( I_{4 \times 4} = \sum_i |b_i\rangle \langle b_i| \)

\[
A |b_k\rangle = \langle 0,0 | b_k\rangle \sum_{j=0}^{7} (b^*_k)^j B^j \left( \sum_{l=0}^{3} |b_l\rangle \langle b_l| \right) |0,0\rangle \\
\quad = \langle 0,0 | b_k\rangle \sum_{j=0}^{7} (b^*_k)^j \sum_{l=0}^{3} B^j |b_l\rangle \langle b_l| |0,0\rangle \\
\quad = \langle 0,0 | b_k\rangle \sum_{l=0}^{3} \sum_{j=0}^{7} (b^*_k)^j (b^*_l)^j |b_l\rangle \langle b_l| |0,0\rangle \\
\quad = \langle 0,0 | b_k\rangle \sum_{l=0}^{3} |b_l\rangle \langle b_l| |0,0\rangle \sum_{j=0}^{7} (b^*_k b_l)^j \\
\]

Where

\[
\sum_{j=0}^{7} (b^*_k b_l)^j = \sum_{j=0}^{7} \left[ \exp \left( -im_k \pi/4 \right) \exp \left( im_l \pi/4 \right) \right]^j \\
\quad = \sum_{j=0}^{7} \exp \left( i (m_l - m_k) 2\pi/8 \right)^j \\
\quad = 8\delta_{kl} \\
\]

as long as \( m_l \neq m_k \) for \( k \neq l \), namely \( B \) is not degenerate. Therefore, for non-degenerate \( B \)

\[
A |b_k\rangle = \langle 0,0 | b_k\rangle \sum_{l=0}^{3} |b_l\rangle \langle b_l| |0,0\rangle 8\delta_{kl} \\
\quad = \langle 0,0 | b_k\rangle |b_k\rangle \langle b_k| |0,0\rangle 8 \\
\quad = \left( 8 |\langle 0,0 | b_k| \rangle|^2 \right) |b_k\rangle \\
\]
Which means that $|b_k\rangle$ is an eigenstate of $A$ corresponding to the eigenvalue $8 \langle 0, 0 | b_k\rangle^2$. Hence, luckily, eigenstates of $B$ are eigenstates of $A$ too. This implies that if $|\phi\rangle$ is chosen as an eigenvector of $B$ then all terms in the summation $\sum_{j=0}^{7} |\langle \phi | B^{j} | 0, 0 \rangle|^{2}$ will be the same.

$$\sum_{j=0}^{7} |\langle \phi | B^{j} | 0, 0 \rangle|^{2} = \sum_{j=0}^{7} |\langle b_k | B^{j} | 0, 0 \rangle|^{2}$$

$$= \sum_{j=0}^{7} |b_k \langle b_k | 0, 0 \rangle|^{2}$$

$$= \sum_{j=0}^{7} |b_k|^{2} |\langle b_k | 0, 0 \rangle|^{2}$$

$$= \sum_{j=0}^{7} |\langle b_k | 0, 0 \rangle|^{2}$$

where we used $|b_k|^{2} = |\exp (im_{k}\pi/4)|^{2} = 1$. This means all probabilities generated by the sequence will be the same, $|\langle b_k | 0, 0 \rangle|^{2}$, if we use an eigenvector of $B$. Then, $\langle A \rangle$ becomes

$$\langle A \rangle = 8 |\langle b_k | 0, 0 \rangle|^{2}.$$ 

What we have to do is to choose the eigenstate of $B$ (which is also an eigenstate of $A$) that has the largest overlap with $|0, 0\rangle$, our initial state in the sequence. Therefore let us calculate the eigenstates of $B$.

Remember $U$’s eigenstates are $|x+\rangle$ and $|x−\rangle$ with corresponding eigenvalues $+1$ and $+i$. Therefore

$$B |x+\rangle |x+\rangle = (U \otimes I) S |x+\rangle |x+\rangle$$

$$= (U \otimes I) |x+\rangle |x+\rangle$$

$$= U |x+\rangle \otimes I |x+\rangle$$

$$= +1 |x+\rangle |x+\rangle$$

and similarly

$$B |x−\rangle |x−\rangle = U |x−\rangle \otimes I |x−\rangle$$

$$= +i |x−\rangle |x−\rangle$$
Hence two eigenstates of $B$ are $|b_0\rangle = |x^+\rangle |x^+\rangle$ and $|b_1\rangle = |x^-\rangle |x^-\rangle$ with corresponding eigenvalues $+1$ and $+i$.

Apply $B$ on $|x^+\rangle |x^-\rangle$ and $|x^-\rangle |x^+\rangle$ to find the other two eigenstates

$$B |x^+\rangle |x^-\rangle = U |x^-\rangle \otimes I |x^+\rangle$$
$$= +i |x^-\rangle |x^+\rangle$$

$$B |x^-\rangle |x^+\rangle = U |x^+\rangle \otimes I |x^-\rangle$$
$$= +1 |x^+\rangle |x^-\rangle$$

$$B\alpha |x^+\rangle |x^-\rangle = \alpha e^{i\pi/2} |x^-\rangle |x^+\rangle$$

$$B\beta |x^-\rangle |x^+\rangle = \beta e^0 |x^+\rangle |x^-\rangle$$

Choose $\alpha$ and $\beta$ so that we can factor out $|x^+\rangle |x^-\rangle$ and $|x^-\rangle |x^+\rangle$ into the same parenthesis at both sides of the equation. $\alpha = 1$ and $\beta = e^{i\pi/4}$ will do the job.

$$B |x^+\rangle |x^-\rangle = e^{i\pi/2} |x^-\rangle |x^+\rangle$$

$$Be^{i\pi/4} |x^-\rangle |x^+\rangle = e^{i\pi/4} |x^+\rangle |x^-\rangle .$$

Sum both sides

$$B \left(|x^+\rangle |x^-\rangle + e^{i\pi/4} |x^-\rangle |x^+\rangle\right) = e^{i\pi/4} \left(|x^+\rangle |x^-\rangle + e^{i\pi/4} |x^-\rangle |x^+\rangle\right)$$

$$B |b_2\rangle = e^{i\pi/4} |b_2\rangle$$

Similarly, subtract both sides
\[ B \left( |x+ \rangle |x- \rangle - e^{i\pi/4} |x- \rangle |x+ \rangle \right) = e^{-i\pi/4} \left( |x+ \rangle |x- \rangle + e^{i3\pi/4} |x- \rangle |x+ \rangle \right) \]
\[ = e^{-i\pi/4} \left( |x+ \rangle |x- \rangle + e^{-i\pi/4} |x- \rangle |x+ \rangle \right) \]
\[ B |b_3 \rangle = e^{-i\pi/4} |b_3 \rangle \]

The last two eigenstates are 
\[ |b_2 \rangle = \frac{1}{\sqrt{2}} \left( |x+ \rangle |x- \rangle + e^{i\pi/4} |x- \rangle |x+ \rangle \right) \] (normalized) and 
\[ |b_3 \rangle = \frac{1}{\sqrt{2}} \left( |x+ \rangle |x- \rangle + e^{-i\pi/4} |x- \rangle |x+ \rangle \right) \] (normalized) with eigenvalues \( e^{i\pi/4} \) and \( e^{-i\pi/4} \). 

\( B \) has eigenvalues \( \{ e^{-i\pi/4}, 1, e^{i\pi/4}, i \} \). Hence \( B \) is not degenerate.

Now we need \( |b_k \rangle \) that has largest overlap with \( |0,0 \rangle \).

\[ |\langle 0, 0 | b_0 \rangle| = |\langle 0, 0 | x+, x+ \rangle| \]
\[ = |\langle 0 | x+ \rangle \langle 0 | x+ \rangle| \]
\[ = |\langle 0 | x+ \rangle|^2 \]
\[ = \left| \frac{1}{\sqrt{2}} \langle 0 \rangle \left( \langle 0 \rangle + |1 \rangle \right) \right|^2 \]
\[ = \frac{1}{2} \]

\[ |\langle 0, 0 | b_1 \rangle| = |\langle 0, 0 | x-, x- \rangle| \]
\[ = \left| \frac{1}{\sqrt{2}} \langle 0 \rangle \left( \langle 0 \rangle - |1 \rangle \right) \right|^2 \]
\[ = \frac{1}{2} \]
\[
\langle 0, 0 | b_2 \rangle = \left| \langle 0, 0 | \frac{1}{\sqrt{2}} \left( |x+, x-\rangle + e^{i\pi/4} |x-, x+\rangle \right) \right|
\]
\[
= \frac{1}{\sqrt{2}} \left| \langle 0, 0 | x+, x-\rangle + e^{i\pi/4} \langle 0, 0 | x-, x+\rangle \right|
\]
\[
= \frac{1}{\sqrt{2}} \left| \langle 0 | x+ \rangle \langle 0 | x- \rangle + e^{i\pi/4} \langle 0 | x- \rangle \langle 0 | x+ \rangle \right|
\]
\[
= \frac{1}{\sqrt{2}} \left| \langle 0 | x+ \rangle \langle 0 | x- \rangle \left( 1 + e^{i\pi/4} \right) \right|
\]
\[
= \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2}} (|0\rangle (|0\rangle + |1\rangle) \frac{1}{\sqrt{2}} (|0\rangle (|0\rangle - |1\rangle)) \left( 1 + e^{i\pi/4} \right)
\]
\[
= \frac{1}{\sqrt{2}} \left[ \frac{1}{2} \left( 1 + e^{i\pi/4} \right) \right] = \frac{1}{2\sqrt{2}} \left| 1 + \frac{\sqrt{2}}{2} i + \frac{\sqrt{2}}{2} \right|
\]
\[
= \frac{1}{2\sqrt{2}} \sqrt{\left( 1 + \frac{\sqrt{2}}{2} \right)^2 + \left( \frac{\sqrt{2}}{2} \right)^2} = \frac{1}{2\sqrt{2}} \sqrt{1 + \sqrt{2} + \frac{1}{2}}
\]
\[
= \frac{1}{2} \sqrt{2 + \sqrt{2}}
\]

\[
= \frac{1}{2} \sqrt{1 + \frac{1}{\sqrt{2}}}
\]

Similarly

\[
\langle 0, 0 | b_1 \rangle = \left| \langle 0, 0 | \frac{1}{\sqrt{2}} \left( |x+, x-\rangle - e^{i\pi/4} |x-, x+\rangle \right) \right|
\]
\[
= \frac{1}{\sqrt{2}} \left| \frac{1}{2} \left( 1 - e^{i\pi/4} \right) \right| = \frac{1}{2\sqrt{2}} \left| 1 - \frac{\sqrt{2}}{2} i - \frac{\sqrt{2}}{2} \right|
\]
\[
= \frac{1}{2\sqrt{2}} \sqrt{\left( 1 - \frac{\sqrt{2}}{2} \right)^2 + \frac{1}{2}} = \frac{1}{2\sqrt{2}} \sqrt{2 - \sqrt{2}}
\]
\[
= \frac{1}{2} \sqrt{1 - \frac{1}{\sqrt{2}}}
\]

Largest overlap is with \( |b_2\rangle \)

\[
\langle 0, 0 | b_2 \rangle = \frac{1}{2} \sqrt{1 + \frac{1}{\sqrt{2}}}
\]

Therefore the largest eigenvalue of \( A \), \( 8 |\langle 0, 0 | b_2 \rangle|^2 \), is
\[ a_{\text{max}} = 8 \left( \frac{1}{2} \sqrt{1 + \frac{1}{\sqrt{2}}} \right)^2 \]
\[ = \frac{8}{4} \left( 1 + \frac{1}{\sqrt{2}} \right) \]
\[ = 2 + \sqrt{2} \]

The maximum value of the sum of the joint probabilities for the chosen observables given by quantum mechanics is \( q = 2 + \sqrt{2} \). We want to compare this with the maximum value allowed by classical mechanics.

### 3.3.2 Classical Bound of Reproduced CH

The classically maximum value of the sum of these 8 joint probabilities is independent of the quantities that are chosen (because there are no uncertainty relations between classical quantities). It is enough to assume that they have only \( K = 2 \) outcomes.

We will use the labeling from the previous section where we called \( z \)-basis the first quantity, and \( v \)-basis the second quantity. Then the sequence of joint probabilities become

\[ \{p_j\} = \{p(a_1 = 0, b_1 = 0), p(a_2 = 0, b_1 = 0), p(a_2 = 0, b_2 = 0), p(a_1 = 1, b_2 = 0), p(a_1 = 1, b_1 = 1), p(a_2 = 1, b_1 = 1), p(a_2 = 1, b_2 = 1), p(a_1 = 0, b_2 = 1)\}. \]

These probabilities being produced by a classical system means that there exists a joint probability of four variables \( p(a_1, a_2, b_1, b_2) \) of which marginals are \( p(a_m, b_n) \). Such as

\[ p(a_2 = i, b_1 = j) = \sum_{k=0}^{1} \sum_{l=0}^{1} p(a_1 = k, a_2 = i, b_1 = j, b_2 = l) \]
\[ = \sum_{k=0}^{1} \sum_{l=0}^{1} p_{kijl} \]

where we used a short hand notation for the 4-tuple joint probability where the indices correspond to the outcomes of \( a_1, a_2, b_1 \) and \( b_2 \) in the given order:

\[ \{p(a_1 = 0, b_1 = 0) = p_{0000} + p_{0101}, p(a_2 = 0, b_1 = 0) = p_{0000} + p_{1001}, \]
\[ p(a_2 = 0, b_2 = 0) = p_{0000} + p_{1010}, p(a_1 = 1, b_2 = 0) = p_{1000} + p_{1110}, \]
\[ p(a_1 = 1, b_1 = 1) = p_{1010} + p_{1111}, p(a_2 = 1, b_1 = 1) = p_{0110} + p_{1111}, \]
\[ p(a_2 = 1, b_2 = 1) = p_{0101} + p_{1111}, p(a_1 = 0, b_2 = 1) = p_{0001} + p_{0111}\} \)

The sum of 8 probabilities in terms of 4-tuple joint probabilities is
\[ 3p_{0000} + 1p_{0001} + 2p_{0101} + 1p_{0110} + 1p_{0111} + 1p_{1000} + 1p_{1001} + 2p_{1010} + 1p_{1110} + 3p_{1111} \]

where \( \sum_{ijkl} p_{ijkl} = 1 \).

The sum can be expressed as the dot product of two vectors

\[ \beta \cdot p \]

where

\[ \beta = (3, 1, 0, 0, 0, 2, 1, 1, 1, 2, 0, 0, 0, 1, 3) \]

and \( p \) is a vector of which components are joint probabilities of 4-outcomes from \( p_{0000} \) to \( p_{1111} \), and of which 1-norm is 1. This means we are calculating a convex linear combination of the components of \( \beta \). The highest value we can get from this operation by choosing values for \( p_{ijkl} \) with the constraint \( \sum p_{ijkl} = 1 \) is the biggest component of \( \beta \).

\[ \max_{p, \|p\|_1 = 1} \beta \cdot p = \beta_{\text{argmax}} \]

where \( \text{argmax} \) means the index of the biggest component.

Therefore the classical bound for this Bell inequality is \( c = 3 \).

\[ 2 + \sqrt{2} \approx 3.4 > 3 \]

\[ q > c \]

which is a Bell inequality violation.

### 3.3.3 Non-local Game Properties of the Reproduced CH System

There are 4 different joint measurements \( \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2)\} \) that can be done on 2 parties and each joint measurement has 4 possible outcomes \( \{(0,0), (0,1), (1,0), (1,1)\} \). In total there are 16 different probabilities and 8 of them appear in our sequence with the same value \( (2 + \sqrt{2})/8 \approx 0.427 \).
Let us analyze the non-local game properties of the quantum correlations produced by our choice of observables and quantum state $|\phi\rangle = |b_2\rangle$ and compare it with the classical version.

Define question that are asked to Alice and Bob as $a_1 \rightarrow s = 0$, $a_2 \rightarrow s = 1$, $b_1 \rightarrow t = 0$, $b_2 \rightarrow t = 1$. Alice and Bob are asked 1-bit questions $(s, t)$, which can be one of these 4, $(s, t) \in\{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

Similarly their 1-bit responses can be 1 of these 4 possible cases: $(a, b) \in\{(0, 0), (0, 1), (1, 0), (1, 1)\}$

Assume the joint outcomes that are produced by our sequence are the correct answers. They can be summarized as

1. If the question pair $(s, t)$ is one of $(0, 0)$, $(0, 1)$, $(1, 0)$ then Alice and Bob should give the same response, namely $(0, 0)$ or $(1, 1)$.
2. If $(s, t)$ is $(0, 1)$ Alice and Bob should give different response, namely $(0, 1)$ or $(1, 0)$.

The answer pattern can be expressed in a compact way by this expression

$$\bar{s} \land t = a + b$$

where $\bar{s}$ is the negation of $s$, and addition is binary addition which is modulo 2.

Because Alice and Bob cannot interact while giving their answers and classically they do not share any non-local resource their answers are only function of their own inputs.

$$a = a(s) \neq a(s, t), \quad b = b(t) \neq b(s, t)$$

And any probabilistic strategy will be a convex combination of deterministic strategies, we just need the deterministic strategy (in which the responses (outputs) are functions of questions (inputs)).

Each party can have 4 different deterministic strategy.

1. $a(s) = 0$
2. $a(s) = 1$
3. $a(s) = s$
4. $a(s) = \bar{s}$

and same strategies for Bob. In total they can have $4 \times 4 = 16$ strategies.

$$w_c = \frac{1}{4} \sum_{s, t} \delta_{\bar{s} \land t, a(s)+b(t)}$$
If we calculate $w_c$ for all 16 strategies, we see that its maximum value is $3/4$. For example for the strategy $a(s) = 0, b(t) = 0$

$$w_c = \frac{1}{4} \sum_{s,t} \delta_{s\land t,0}$$

$$= \frac{1}{4} (\delta_{0\land 0,0} + \delta_{0\land 1,0} + \delta_{1\land 0,0} + \delta_{1\land 1,0})$$

$$= \frac{1}{4} (\delta_{1\land 0,0} + \delta_{1\land 1,0} + \delta_{0\land 0,0} + \delta_{0\land 1,0})$$

$$= \frac{1}{4} (\delta_{0,0} + \delta_{1,0} + \delta_{0,0} + \delta_{0,0})$$

$$= \frac{1}{4} (1 + 0 + 1 + 1)$$

$$= \frac{3}{4} = 0.75$$

Let us calculate $w_q$. Alice and Bob share the quantum state $\left| b_k \right\rangle$. When a question $(s, t)$ is asked, Alice measures $z$-basis if $s = 0$ (or Bob if $t = 0$) and $v$-basis if $t = 1$ (or Bob if $t = 1$) and responds the measurement outcome.

We calculated that all the probabilities in the Bell expression are

$$p_j = \left| \langle 0, 0 | b_2 \rangle \right|^2$$

$$= \frac{2 + \sqrt{2}}{8}.$$ 

We know that for each question there are 4 responses 2 of which are correct answers. For each question the sequence produces two correct answers of which probability is $p_j$ (which are the same for different questions) hence quantum winning probability is

$$w_q = 2p_j$$

$$= \frac{2 + \sqrt{2}}{4} \approx 0.85$$

If we compare the classical an quantum strategies, we see that
\[
\frac{2 + \sqrt{2}}{4} > \frac{3}{4}
\]
\[
0.85 > 0.75
\]
\[
w_q > w_c
\]

which is a Bell violation in terms of non-local games.

### 3.3.4 Generalization to Qudits and More Than Two Measurements

We saw this method of sequence generation using translation and swap operator on the familiar setting of a CH scenario. Let us generalize this scheme to arbitrary number of outcomes and measurements.

To have \( K \) different outcomes the quantum states should live in \( \mathcal{H} = \mathbb{C}^K \). Computational basis is \( \{ |k\rangle |k = 0..K - 1 \} \). The translation operator on the computational basis is

\[
T |k\rangle = |k + 1\rangle
\]

where the summation is modulo \( K \), so that

\[
T^K = I.
\]

We want to have \( M \) measurements per party, therefore we define the \( U \) operator as

\[
U^M = T.
\]

This choice gives us \( M \) different orthogonal bases per party

\[
\left\{ |v_j^{(m)}\rangle = U^m |j\rangle |j = 0..K - 1 \right\}
\]

where \( m = 0..M - 1 \). (|\( v_j^{(0)}\rangle = |j\rangle\))

The \( B \) operator is again

\[
B = (U \otimes I) S
\]

and \( B^2 = U \otimes U \). This implies

\[
B^{2MK} = I \otimes I.
\]
We choose $|0,0\rangle$ as our initial state. A state that is not in computational basis will again generate a closed sequence under the application of the translation operator but its elements will not be orthonormal vectors anymore. If we had chosen a different initial state, then we needed a translation operator adjusted for that state. Any choice of initial state can be made useful with the a right choice of translation operator, or vice versa. In other words a similarity transformation on $T = \sum |k\rangle \langle k+1|$ could provide a new basis

$$STS^{-1} = \sum S |k\rangle \langle k+1| S^{-1}$$

$$\tilde{T} = \sum |\tilde{k}\rangle \langle \tilde{k}+1|$$

where we can define $\tilde{U}^M = T$ and apply the $B$ operator on $|0\rangle |0\rangle$.

Similarly the function of the translation operator is to generate a sequence of orthonormal basis elements. The order of $|k\rangle \rightarrow |k+1\rangle$ is not essential. Any different order can be generated by a similarity transformation too. Which means we are not loosing generality by choosing the computational basis.

Repeated application of $B$ on $|0,0\rangle$ generates a sequence of length $2MK$ of which elements are $B^j |0,0\rangle$

$$|0\rangle |0\rangle \rightarrow |v_0^{(1)}\rangle |0\rangle \rightarrow |v_0^{(1)}\rangle |v_0^{(1)}\rangle \rightarrow |v_0^{(2)}\rangle |v_0^{(1)}\rangle \rightarrow |v_0^{(2)}\rangle |v_0^{(2)}\rangle \rightarrow \cdots$$

$$|1\rangle |1\rangle \rightarrow \cdots |v_{K-1}^{(M-1)}\rangle |v_{K-1}^{(M-1)}\rangle \rightarrow |0\rangle |v_{K-1}^{(M-1)}\rangle \rightarrow |0\rangle |0\rangle \rightarrow \cdots$$

The type of states of the form $|v_{j}^{(m)}\rangle |v_{k}^{(n)}\rangle$ are

- $m = n$ and $j = k$, $|v_{j}^{(m)}\rangle |v_{k}^{(m)}\rangle$
- $m = n + 1$ and $j = k$ for $0 \leq n \leq M - 2$, $|v_{j}^{(n+1)}\rangle |v_{j}^{(n)}\rangle$
- $m = 0$, $n = M - 1$ and $j = k + 1$, $|v_{j+1}^{(0)}\rangle |v_{j}^{(M-1)}\rangle$

We want to maximize the sum of probabilities

$$\max_{|\phi\rangle} \sum_{j=0}^{2MK-1} |\langle \phi | B^j |0,0\rangle|^2$$

by finding the largest eigenvalue of

$$A = \sum_{j=0}^{2MK-1} B^j |0,0\rangle \langle 0,0| B^j.$$
We start with the fact that eigenstates of $B$ are also eigenstates of $A$ because $A$ and $B$ commute, $[A, B] = 0$ and hence they can be diagonalized simultaneously.

$$[A, B] = AB - BA$$

$$= \sum_{j=0}^{2MK-1} B^j |0, 0\rangle \langle 0, 0| B^{1j} - \sum_{j=0}^{2MK-1} BB^j |0, 0\rangle \langle 0, 0| B^{1j}$$

$$= (|0, 0\rangle \langle 0, 0| B - B |0, 0\rangle \langle 0, 0|) + (B^1 |0, 0\rangle \langle 0, 0| - B^2 |0, 0\rangle \langle 0, 0| B^1)$$

$$+ (B^2 |0, 0\rangle \langle 0, 0| B^1 - B^3 |0, 0\rangle \langle 0, 0| B^{12}) + \cdots$$

$$+ (B^{2MK-2} |0, 0\rangle \langle 0, 0| B^{12MK-3} - B^{2MK-1} |0, 0\rangle \langle 0, 0| B^{12MK-2})$$

$$+ (B^{2MK-1} |0, 0\rangle \langle 0, 0| B^{12MK-2} - B^{2MK} |0, 0\rangle \langle 0, 0| B^{12MK-1})$$

where the second term in a every parenthesis cancels the first term in the next parenthesis leaving only

$$[A, B] = |0, 0\rangle \langle 0, 0| B - B^{2MK} |0, 0\rangle \langle 0, 0| B^{12MK-1}.$$  

$$B^{2MK} = I$$ and $B^{12MK-1} = B^{12MK-1} B^1 B = B^{12MK} = IB = B$, hence

$$[A, B] = |0, 0\rangle \langle 0, 0| B - |0, 0\rangle \langle 0, 0| B$$

$$= 0.$$

Because $B^{2MK} = I$ the eigenvalues of $B$ are of the form

$$b_k = \exp \left( i \frac{2\pi}{2MK} m_k \right)$$

Like the $M = K = 2$ case

$$A |b_k\rangle = \sum_{j=0}^{2MK-1} B^j |0, 0\rangle \langle 0, 0| B^{1j} |b_k\rangle$$

$$= \sum_{j=0}^{K-1} B^j I |0, 0\rangle b_k^* \langle 0, 0| b_k\rangle, \quad I = \sum_{l=0}^{K-1} |b_l\rangle \langle b_l|$$

$$= \langle 0, 0| b^*_k \sum_{l=0}^{K-1} \langle b_l| 0, 0\rangle |b_l\rangle \sum_{j=0}^{K-1} (b_l b^*_l)^j$$
\[
\sum_{j=0}^{2MK-1} (b_l b_k^*)^j = \sum_j \exp \left( i \frac{2\pi}{2MK} (m_l - m_k) j \right)
\]

which is \(2MK\delta_{kl}\) for the non-degenerate case where \(m_k\) are all different for different \(k\), hence the only case where \(b_k = b_l\) can be is when \(m_l = m_k\) and hence \(l = k\). Therefore

\[
A|b_k\rangle = 2MK |\langle 0, 0 | b_k\rangle|^2 |b_k\rangle
\]

where the corresponding eigenvalue is \(a_{k,0} = b_k = 2MK |\langle 0, 0 | b_k\rangle|^2\).

For the case \(B\) is degenerate (which is all cases with \(M, K > 2\)) there are cases where \(m_l = m_k\) even when \(k \neq l\). Say \(k\) is \(d_k\) degenerate, there are \(d_k\) \(l\) values for which \(b_l = b_k\).

\[
\sum_{j=0}^{2MK-1} (b_l b_k^*)^j = 2MK \delta_{b_l b_k}
\]

\[
A|b_k\rangle = 2MK \left( \langle 0, 0 | b_k\rangle \sum_{\{l(b_l=b_k)\}} \langle b_l | 0, 0\rangle |b_l\rangle \right)
\]

Change the notation to indicate all eigenvectors in the degenerate space of \(b_k\) with \(|b_{k,l}\rangle\). The set of eigenstates belonging to the eigenvector \(b_k\) is \(\{|b_{k,l}\rangle | l = 0..d_k - 1\}\).

\[
A|b_k\rangle = 2MK \langle 0, 0 | b_k\rangle \sum_{l=0}^{d_k-1} \langle b_{k,l} | 0, 0\rangle |b_{k,l}\rangle
\]

Call \(\sum_{l=0}^{d_k-1} |b_{k,l}\rangle \langle b_{k,l} | 0, 0\rangle \equiv |\beta_k\rangle\). As can be seen from upper equation all \(|b_k\rangle\) belonging to the degenerate eigenvalue \(b_k\) are mapped to \(|\beta_k\rangle\) by \(A\).

Apply \(A\) on \(|\beta_k\rangle\) to see that it is an eigenstate of \(A\)

\[
A|\beta_k\rangle = \sum_l \langle b_{k,l} | 0, 0\rangle A|b_{k,l}\rangle
\]

\[
= \sum_l \langle b_{k,l} | 0, 0\rangle \left( 2MK \langle 0, 0 | b_{k,l}\rangle \sum_{l'} \langle b_{k,l'} | 00\rangle \right)
\]

\[
= 2MK \sum_l |\langle b_{k,l} | 0, 0\rangle|^2 |\beta_k\rangle
\]

This means that \(|\beta_k\rangle\) is an eigenstate of \(A\) with the eigenvalue \(\beta_k = 2MK \sum_l |\langle b_{k,l} | 0, 0\rangle|^2\). The other eigenvalues of \(A\) in the span \(\{|b_{k,l}\rangle\}\) are zero.
A and B have the same eigenstates. But their eigenvalues and degeneracy structures are different. Both $A$ and $B$ are operators from $\mathbb{C}^{KN} \to \mathbb{C}^{KN}$. And say $B$ has the eigensubspace decomposition

$$\mathbb{C}^{KN} = \bigoplus_k B_k$$

where the dimension of the eigensubspace corresponding to the eigenvalue $b_k$ is $\dim B_k = d_k$.

When $d_k$ is one, in other words $b_k$ is not degenerate, the eigenvalue of $A$, corresponding to this eigenspace is $a_{k,0} = 2MK|\langle 0,0 |b_k\rangle|^2$ where $B_k = \text{span} \{ |b_k\rangle \}$ with $\dim B_k = 1$.

When $d_k > 2$, $B_k$ is spanned by the eigenstates corresponding to the degenerate eigenvalue $b_k$, $B_k = \text{span} \{ |b_{k,l}\rangle |l = 0..d_k-1\}$. $A$ decomposes the subspace $B_k$ differently according to its eigenvalues (still any state in $B_k$ is an eigenstate of $B$). $B_k = \text{span} \{ |\beta_k\rangle \} \oplus \text{span} \{ |\beta_k\rangle \}^\perp$. $A$ maps $B_k$ to span $\{ |\beta_k\rangle \}$. The eigenvalue of $A$ corresponding to the eigenstate $|\beta_k\rangle \in B_k$ is $a_{k,0} = 2MK \sum_l |\langle b_{k,l} | 0,0 \rangle|^2$ and the rest of the eigenvalues $a_{k,l} = 0$ for $l = 1..d_k-1$. span $\{ |\beta_k\rangle \}^\perp = \text{span} \{ |a_{k,l} = 0\rangle \}$. ($A |b_{k,l}\rangle = |\beta_k\rangle$ can only happen, namely an eigenstate is sent to a state that is not parallel to itself, when the corresponding eigenvalue is 0).

Let us find the eigenstates of $B$. $B$ is made of $I$ and $U$ which is a function of $T$. Therefore, start with the eigensystem of $T$

$$T = \sum_{k=0}^{K-1} |k\rangle \langle k+1|.$$

The eigenvalues of $T$ can be calculated from the relation $T^K = I$. Because $t_j^K = 1$, $t_j = 1^{1/K}$, which are

$$t_j = e^{\frac{2\pi}{K}j}.$$

Corresponding eigenstates can be found using the eigenproblem. Assume $|t_j\rangle = \sum_k c_k |k\rangle$

$$T|t_j\rangle = t_j |t_j\rangle$$

$$\sum_k c_k^{(j)} T|k\rangle = e^{\frac{2\pi}{K}j} \sum_k c_k^{(j)} |k\rangle$$

$$\sum_{k=0}^{K-1} c_k^{(j)} |k+1\rangle = \sum_{k=0}^{K-1} c_k^{(j)} e^{\frac{2\pi}{K}j} |k\rangle$$

$$\sum_{l=1}^{K} c_l^{(j)} |l\rangle = \sum_{l=1}^{K} c_l^{(j)} e^{\frac{2\pi}{K}j} |l\rangle$$

which gives a relationship between successive coefficients.
\[ c_{l-1}^{(j)} = c_l^{(j)} e^{i \frac{2\pi}{K} j} \]

or

\[ c_l^{(j)} e^{-i \frac{2\pi}{K} j} = c_{l+1}^{(j)} \]

hence

\[ c_0^{(j)} \left( e^{-i \frac{2\pi}{K} j} \right)^l = c_l^{(j)}. \]

The normalization condition \( \langle t_k \mid t_k \rangle = 1 \), tells us that \( |c_j^{(j)}| = \frac{1}{\sqrt{K}} \)

\[ c_k^{(j)} = \frac{1}{\sqrt{K}} e^{-i \frac{2\pi}{K} j k} \]

\[ |t_j \rangle = \frac{1}{\sqrt{K}} \sum_{k=0}^{K-1} e^{i \left( \frac{2\pi}{K} j \right) k} |k\rangle \]

\( T \) is called the generalized Pauli-X operator. \( X = \sigma_X \) for \( \dim X = 2 \).

\[ U^M = T \]

therefore

\[ U = \sum_{j=0}^{K-1} \sqrt{t_j} |t_j \rangle \langle t_j| \]

We choose the eigenvalues of \( U \), \( u_j = t_j^{1/M} \) so that \( U \) is non-degenerate.

\[ B = (U \otimes I) S \]

\[ = \left( \sum_{t} u_t |t_t \rangle \langle t_t| \otimes I \right) S \]

Apply \( B \) on \( |t_j \rangle \otimes |t_j \rangle \)
$$B |t_j\rangle |t_j\rangle = (U \otimes I) S |t_j\rangle |t_j\rangle$$

$$= (U \otimes I) |t_j\rangle |t_j\rangle$$

$$= U |t_j\rangle I |t_j\rangle$$

$$= u_j |t_j\rangle |t_j\rangle$$

Therefore $K$ of the eigenstates of $B$ are $|b_j\rangle = |t_j\rangle |t_j\rangle$. To find the rest of the $K^2 - K$ eigenstates apply $B$ on $|t_j\rangle |t_k\rangle$ and $|t_k\rangle |t_j\rangle$

$$B |t_j\rangle |t_k\rangle = (U \otimes I) S |t_j\rangle |t_k\rangle$$

$$= (U \otimes I) |t_k\rangle |t_j\rangle$$

$$= U |t_k\rangle I |t_j\rangle$$

$$= u_k |t_k\rangle |t_j\rangle$$

Similarly

$$B |t_k\rangle |t_j\rangle = u_j |t_j\rangle |t_k\rangle .$$

These two equalities imply that the other eigenstates of $B$ lie in $2 \times 2$ blocks spanned by the vectors $|t_k\rangle |t_j\rangle$ and $|t_j\rangle |t_k\rangle$.

This can be seen by looking at the $B$ operator for the case $K = 3$ expressed in $\{|t_j\rangle |t_k\rangle |j, k = 0, 1, 2\}$ basis. The entries $(j\kappa, j'k')$ correspond to $B_{jk,j'k'} = \langle t_j | t_k | B |t_j'\rangle |t_k'\rangle$. 

<table>
<thead>
<tr>
<th>( j \backslash j'k' )</th>
<th>00</th>
<th>11</th>
<th>22</th>
<th>01</th>
<th>10</th>
<th>02</th>
<th>20</th>
<th>12</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>( u_0 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>( u_1 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>0</td>
<td>( u_2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( u_0 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( u_1 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>02</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( u_0 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( u_2 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( u_0 )</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( u_2 )</td>
<td>0</td>
</tr>
</tbody>
</table>

The order of \( j, k \) pairs are chosen so that the 2 dimensional subspaces become apparent. At every \( 2 \times 2 \) subspace \( B_{jk} \) we have

\[
B \in B_{jk} = \begin{pmatrix}
|t_j\rangle |t_k\rangle & |t_k\rangle |t_j\rangle \\
\langle t_j\rangle \langle t_k| & 0 & u_j \\
\langle t_k\rangle \langle t_j| & u_k & 0
\end{pmatrix}
\]

To diagonalize \( B \) in that subspace

\[
B^2 |t_j\rangle |t_k\rangle = B (u_k |t_k\rangle |t_j\rangle)
\]

\[
= u_k B |t_k\rangle |t_j\rangle
\]

\[
= u_k u_j |t_j\rangle |t_k\rangle
\]

or similarly

\[
B^2 |t_k\rangle |t_j\rangle = u_j u_k |t_k\rangle |t_j\rangle
\]

which means the square of \( B \)'s eigenvalues are

\[
b^2 = u_j u_k
\]

so that

\[
b = \pm \sqrt{u_j u_k}.
\]
The corresponding eigenstates will be

$$B |b\rangle = \pm \sqrt{u_j u_k} |b\rangle.$$ 

Because $|b\rangle \in \text{span} \{|t_j\} |t_k\rangle, |t_k\} |t_j\rangle\}$ we can assume this form

$$|b\rangle = \frac{1}{\sqrt{2}} (|t_j\} |t_k\rangle + \alpha |t_k\} |t_j\rangle)$$

for which the eigenvalue equation becomes

$$B \frac{1}{\sqrt{2}} (|t_j\} |t_k\rangle + \alpha |t_k\} |t_j\rangle) = \pm \sqrt{u_j u_k} \frac{1}{\sqrt{2}} (|t_j\} |t_k\rangle + \alpha |t_k\} |t_j\rangle)$$

$$B |t_j\} |t_k\rangle + \alpha B |t_k\} |t_j\rangle = \pm \sqrt{u_j u_k} |t_j\} |t_k\rangle \pm \sqrt{u_j u_k} \alpha |t_k\} |t_j\rangle$$

$$u_k |t_k\} |t_j\rangle + \alpha u_j |t_j\} |t_k\rangle = \pm \sqrt{u_j u_k} |t_j\} |t_k\rangle \pm \sqrt{u_j u_k} \alpha |t_k\} |t_j\rangle$$

which means

$$\alpha u_j = \pm \sqrt{u_j u_k}.$$

Hence the values of $\alpha$ in terms of the eigenvalues of $U$ are

$$\alpha = \pm \frac{\sqrt{u_j u_k}}{u_j}.$$

Then, the eigensystem of $B$ is made of $K$ eigenvalues $b_{jj} = u_j$ with corresponding eigenstates $|b_{jj}\rangle = |t_j\} |t_j\rangle$, $(K^2 - K) / 2$ eigenvalues $b_{jk} = \sqrt{u_j u_k} (j < k)$ with corresponding eigenstates $\frac{1}{\sqrt{2}} (|t_j\} |t_k\rangle + \frac{\sqrt{u_j u_k}}{u_j} |t_k\} |t_j\rangle)$ and $(K^2 - K) / 2$ eigenvalues $b_{kj} = -\sqrt{u_j u_k} (k < j)$ with corresponding eigenstates $\frac{1}{\sqrt{2}} (|t_j\} |t_k\rangle - \frac{\sqrt{u_j u_k}}{u_j} |t_k\} |t_j\rangle)$.

We developed the general case of the scheme. Let us look at some specific cases.

3.3.5 $N = 2, M = 2, K = 3$ Scenario

In this scenario we use two qutrits and measurement outcomes can have 3 different values. The states of individual parties lie in $\mathcal{H} = \mathbb{C}^3$ where the computational basis is $\{|0\}, |1\}, |2\}$. The eigenstates of the translation operator $T = \sum_{k=0}^{2} |k\rangle \langle k + 1| = X_{3 \times 3}$ are

$$|t_j\rangle = \frac{1}{\sqrt{3}} \sum_{k=0}^{2} e^{i(2\pi j)k} |k\rangle$$
with corresponding eigenvalues \( t_j = e^{i\frac{2\pi}{3}j} \), \( j = 0, 1, 2 \). Therefore the eigenvalues of \( U \) are \( u_j = t_j^{1/2} \). Our choice is

\[
U = |t_0\rangle \langle t_0| + e^{-i\pi/3}|t_1\rangle \langle t_1| + e^{i\pi/3}|t_2\rangle \langle t_2|.
\]

The \( v \)-basis generated by \( U \) is

\[
|v_0\rangle = \frac{1}{3}(2|0\rangle + 2|1\rangle - |2\rangle) \\
|v_1\rangle = \frac{1}{3}(-|0\rangle + 2|1\rangle + 2|2\rangle) \\
|v_2\rangle = \frac{1}{2}(2|0\rangle - |1\rangle + 2|2\rangle).
\]

The sequence generated by \( B \) operator is \(|0\rangle|0\rangle, |v_0\rangle|0\rangle, |v_0\rangle|v_0\rangle, |1\rangle|v_0\rangle, |1\rangle|1\rangle, |v_1\rangle|1\rangle, |v_1\rangle|v_1\rangle, |2\rangle|v_1\rangle, |2\rangle|2\rangle, |v_2\rangle|2\rangle, |v_2\rangle|v_2\rangle, |0\rangle|v_2\rangle\).

The eigenvalues of \( B \), \( \{b_k\} \), become: \( \pm 1, e^{i\pi/3}, e^{-i\pi/3}, \pm e^{i\pi/6}, \) and \( \pm e^{-i\pi/6} \). Only 1 is degenerate and the eigenstates corresponding to it, are the ones that lead to the eigenstate of \( A \) with the largest eigenvalue. Eigenstates corresponding to 1 are \( |a_{00}\rangle = |t_0\rangle|t_0\rangle \) and \( |a_{12}\rangle = (|t_1\rangle|t_2\rangle + e^{i\pi/3}|t_2\rangle|t_1\rangle) / \sqrt{2} \).

The non-zero eigenvalue of \( A \) in the subspace spanned by these vectors is

\[
12 \left(|\langle 0, 0 | a_{00}\rangle|^4 + |\langle 0, 0 | a_{12}\rangle|^4\right) = \frac{10}{3}
\]

where

\[
|\beta_1\rangle = \sqrt{\frac{2}{5}} \left[|t_{00}\rangle + \frac{1}{\sqrt{2}}(1 + e^{-i\pi/3})|t_{12}\rangle\right] \\
= \sqrt{\frac{2}{5}} \left[\frac{5}{6}(|00\rangle + |11\rangle + |22\rangle) \\
+ \frac{1}{3}(|01\rangle + |20\rangle + |12\rangle) \\
- \frac{1}{6}(|01\rangle + |20\rangle + |12\rangle)\right].
\]

If we put the two party system into the state \( |\phi\rangle = |\beta_1\rangle \) and measure the probabilities that appear in the sequence, \( p_j = |\langle \phi | B^j | 0, 0 \rangle|^2 \), their sum will be \( \frac{10}{3} \). And if we assume that all of these probabilities of the form \( p(a_m = i, b_n = j) \) are marginals of a bigger joint probability distribution \( p(a_1 = i, a_2 = j, b_1 = k, b_2 = l) \), the classical bound for their sum will be 3.
which is a Bell violation.

If we analyze the violation in terms of a non-local game the question that can be asked to Alice and Bob are \((s, t) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}\). The correct answers are returning same bits when the question is \((0, 0), (1, 0)\) or \((1, 1)\), and returning \((0, 2), (1, 0)\) or \((2, 1)\) \((s, t) = (0, 1)\). Correct answer formula can be expressed as \(\bar{s} \land t = a - b\) in a more concise way.

\[
a(s) = j, \quad b(t) = j
\]

strategy will let them win with the probability

\[
w_c = \frac{3}{4}.
\]

In quantum version each probability in the sequence is

\[
p_j = \frac{a_{\text{max}}}{N M K} = \frac{10/3}{12} = \frac{5}{18}.
\]

There are 3 correct answers for each question, hence the probability of giving a correct answer is

\[
w_q = \frac{5}{18} \cdot \frac{3}{2} = \frac{5}{6}.
\]

\[
w_q > w_c.
\]

### 3.3.6 \(N = 2, \ M = 3, \ K = 2\) Scenarios

The translation operator is \(T = |x+\rangle \langle x+| - |x-\rangle \langle x-|\). \(U^3 = T\). Our choice of \(U\) is

\[
U = |x+\rangle \langle x+| + e^{i\pi/3} |x-\rangle \langle x-|
\]

which generate two new bases, \(\left\{|v_j^{(k)}\rangle = U^k |j\rangle \mid j = 0, 1 \right\}\), other than the computational basis \(\left\{|v_j^{(0)}\rangle = |j\rangle \mid j = 0, 1 \right\}\).

The sequence generated by \(B\) is
\[ |0\rangle|0\rangle \rightarrow |v_0^{(1)}\rangle|0\rangle \rightarrow |v_0^{(1)}\rangle|v_0^{(1)}\rangle \rightarrow |v_0^{(2)}\rangle|v_0^{(1)}\rangle \rightarrow \\
|v_0^{(2)}\rangle|v_0^{(2)}\rangle \rightarrow |1\rangle|v_0^{(2)}\rangle \rightarrow |1\rangle|1\rangle \rightarrow |v_1^{(1)}\rangle|1\rangle \rightarrow \\
|v_1^{(1)}\rangle|v_1^{(1)}\rangle \rightarrow |v_1^{(2)}\rangle|v_1^{(1)}\rangle \rightarrow |v_1^{(2)}\rangle|v_1^{(2)}\rangle \rightarrow |0\rangle|v_1^{(2)}\rangle.\]

Eigenvalues of \( B \) are 1, \( e^{i\pi/3} \) and \( \pm e^{i\pi/6} \) none of which is degenerate. The eigenstate

\[ |e^{i\pi/6}\rangle = \frac{1}{\sqrt{2}}(|x\rangle - x) + e^{i\pi/6}|-x\rangle + x)\]

\[ = \frac{1}{2\sqrt{2}}[(1 + e^{i\pi/6})(|0\rangle|0\rangle - |1\rangle|1\rangle) \]
\[ + (1 - e^{i\pi/6})(|1\rangle|0\rangle - |0\rangle|1\rangle)]\]

yields the largest eigenvalue for \( A \), which is

\[ a_{\text{max}} = 12 \left| \langle e^{i\pi/6} | 0, 0 \rangle \right|^2 = \frac{3}{2} \left( 2 + \sqrt{3} \right).\]

To calculate the classical bound define \( a_k \) and \( b_k \) as quantities corresponding to the \( a^{(k)} \)-basis. Assume the existence of \( p(a_1, a_2, a_3, b_1, b_2, b_3) \) and derive the joint probabilities in the sequence a its marginals. The classical bound becomes, \( c = 5 \).

\[ \frac{3}{2} \left( 2 + \sqrt{3} \right) \approx 5.598 > 5 \]

In the non-local game context the questions that can be asked are \( (s,t) \) with \( s-t = 0, 1 \) in modulo 3. The correct answer for the question \( (s,t) = (0, 2) \) is returning opposite bit values and for the rest of the questions the same bit. Classical maximum winning probability is \( w_c = 5/6 \) and quantum winning probability is

\[ \frac{a_{\text{max}}}{2} = \frac{1}{2} \approx 0.933 \]

\[ w_q > w_c \]
3.3.7 \( N = 2, M \geq 3, K = 2 \) Scenarios

Choose \( U = |x+\rangle \langle x+| + e^{i\pi/M} |x-\rangle \langle x-| \). The measurement bases are \( \{ |v_j^{(k)}\rangle = U^k |j\rangle \mid j = 0, 1 \} \) for \( k = 0..M - 1 \). Eigenvalues of \( B \) are 1, \( e^{i\pi/M} \) and \( \pm e^{i\pi/2M} \). The eigenvalue \( e^{i\pi/2M} \) whose eigenstate is

\[
\frac{1}{\sqrt{2}} (|x+\rangle \langle x-| + e^{i\pi/2M} |x-\rangle \langle x+|)
\]

yields the largest eigenvalue for \( A \) which is

\[
a_{\text{max}} = M \left[ 1 + \cos \left( \frac{\pi}{2M} \right) \right].
\]

Define \( a_k \) and \( b_k \) as quantities corresponding to the \( v^{(k)} \)-basis, the sum of the probabilities from the sequence becomes

\[
\sum_{j=0,1} \left[ \sum_{k=0}^{M-1} p(a_k = j, b_k = j) + \sum_{k=0}^{M-1} p(a_k+1 = j, b_k = j) \right]
\]

\[+ p(a_0 = 0, b_{M-1} = 1) + p(a_0 = 1, b_{M-1} = 0).\]

If they come from a joint distribution \( p(a_0, a_1, \ldots, a_{M-1}, b_0, b_1, \ldots, b_{M-1}) \) will give a classical bound of \( 2M - 1 \).

\[
M \left[ 1 + \cos \left( \frac{\pi}{2M} \right) \right] > 2M - 1
\]

\[q > c.\]

In terms of nonlocal games, questions are \( s, t \in \{0..M - 1\} \) and \( s - t = 0, 1 \) modulo \( M \). The correct answer is to return opposite bits for the question \( (s, t) = (0, M - 1) \) and same bits for the rest. \( a(s) = 0, b(t) = 0 \) strategy gives the highest classical winning probability

\[
w_c = 1 - \frac{1}{2M}
\]

whereas the quantum winning probability is

\[
\frac{M \left[ 1 + \cos \left( \frac{\pi}{2M} \right) \right]}{2MK} K = \frac{1}{2} \left[ 1 + \cos \left( \frac{\pi}{2M} \right) \right] \approx 1 - \frac{\pi^2}{16M^2}
\]
3.3.8 Numerical Results for $N = 2$ Parties

The quantum value $q$ for different $M$ and $K$ values is calculated and shown in Table 3.

The heuristic formula found for the classical bound is

$$c(M, K) = 2M - 1.$$  

The difference between the quantum value and the classical bound, in other words the amount of violation is given in the Table 4.

Another important quantity is the mutual information. The quantum value $q$, the classical bound $c$, the value of each probability in the sequence of the quantum case $p$ (which is also the probability that Alice can correctly guess Bob’s outcome) and the mutual information between the parties

$$I(a_m, b_n) = \sum_{j,k=0}^{K-1} p(a_m = j, b_n = k) \log_2 \left[ \frac{p(a_m = j, b_n = k)}{p(a_m = j) p(b_n = k)} \right]$$

is calculated numerically and shown in Table 5.

Mutual information depends on the choice of observables ($m$ and $n$). We choose the pair with the highest mutual information, which happened to be $a_1$ and $b_1$, the observables in computational basis.

From the table we can see that while the size of the Bell violation is decreasing the mutual information
Table 5: Comparison of quantum value with the mutual information for increasing dimensionality

<table>
<thead>
<tr>
<th>$K$</th>
<th>$q$</th>
<th>$c$</th>
<th>$p$</th>
<th>$I_{ab}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.4142</td>
<td>3</td>
<td>0.8536</td>
<td>0.3991</td>
</tr>
<tr>
<td>3</td>
<td>3.3333</td>
<td>3</td>
<td>0.8333</td>
<td>0.8146</td>
</tr>
<tr>
<td>4</td>
<td>3.3066</td>
<td>3</td>
<td>0.8266</td>
<td>1.1482</td>
</tr>
<tr>
<td>5</td>
<td>3.2944</td>
<td>3</td>
<td>0.8236</td>
<td>1.4223</td>
</tr>
</tbody>
</table>

between Alice and Bob is increasing with $K$. This is because while the probability of correctly guessing Bob’s outcome is decreasing, the number of alternatives from which she is choosing is increasing. Thus, Alice and Bob share more information as $K$ increases.

### 3.4 Bell Inequalities from Group Actions: $N = 3$ Parties

In this section we will extend the method from the previous sections to the 3 parties case where each party is a qutrit system ($K = 3$) and they choose among two possible measurements: Alice $\{a_0, a_1\}$, Bob $\{b_0, b_1\}$, and Charlie $\{c_0, c_1\}$.

We are not going to use the swap operator $S$ anymore. This allows us to build the connection between abstract groups and states in the sequences. In this first example, the group of interest is the cyclic group of order 6, $Z_6$.

The translation operator $T$ acts on computational basis as $T|j\rangle = |j + 1\rangle$ $j = 0, 1, 2$ where the addition is modulo $K = 3$.

$U^2 = T$ is chosen as

$$U = |t_0\rangle\langle t_0| + e^{-i\pi/3}|t_1\rangle\langle t_1| + e^{i\pi/3}|t_2\rangle\langle t_2|$$

where eigenstates of $T$, $|t_j\rangle$ are, like before

$$|t_j\rangle = \frac{1}{\sqrt{3}} \sum_{k=0}^{2} e^{i\frac{2\pi}{3}jk} |k\rangle.$$

$U^6 = I$ implies that $\{U^m|m = 0..5\}$ is a representation of $Z_6 = (Z_6, +)$. Application of $U^m$ on $\mathbb{C}^3 \otimes \mathbb{C}^3 \otimes \mathbb{C}^3$ defines a group action.

Define a second basis as $\{|v_j\rangle = U|j\rangle |j = 0, 1, 2\}$. The first basis, computational basis, corresponds to the first observables ($a_0, b_0$ and $c_0$) and second basis corresponds to the second observables ($a_1, b_1$ and $c_1$).

The $B$ operator for the new scheme is

$$B = U \otimes U \otimes U.$$
Actually, this choice makes the $S$ operator unnecessary because the effect will be the same $(U \otimes U \otimes U) S = U \otimes U \otimes U$.

We choose 4 initial states $\{ |\xi_i \rangle | i = 0..3 \}$ which are

$$
|021\rangle \quad |00v_1\rangle \quad |0v_00\rangle \quad |v_020\rangle.
$$

Each initial state generates an orbit of 6 elements, in total we have $4 \times 6 = 24$ states.

In the sequence generated by the first initial state, all parties measure the same basis. In the sequences generated by the rest of the initial states, two of the parties measure the same state and the other one measures the other basis. These basis combinations exhaust all possible choices of measurement bases by the parties. This results in 24 three-qutrit states each of which is a product of single-qutrit states from one of the two bases.

Like before the Bell expression is the sum of probabilities corresponding to the outcome probabilities of the events corresponding to these states. For example $|00v_1\rangle$ corresponds to Alice measuring $a_0$ and obtaining 0, Bob measuring $b_0$ and obtaining 0, and Charlie measuring $c_1$ and obtaining 1.

Again we want to find the maximum eigenvalue of the $A$ operator. This time we have 4 initial states. We can do this by first calculating 4 $A_i$ operators for each orbit, and then calculate their sum.

$$
A = \sum_i A_i
= \sum_i \left( \sum_{j=0}^{5} B_j^i |\xi_i\rangle \langle \xi_i| B_j^{i} \right)
$$

Or, we can first define

$$
L = \sum_i |\xi_i\rangle \langle \xi_i|
= |021\rangle \langle 021| + |00v_1\rangle \langle 00v_1| + |0v_00\rangle \langle 0v_00| + |v_020\rangle \langle v_020|.
$$

Then, $A$ becomes
\[ A = \sum_{j=0}^{5} B^j L B^{\dagger j} \]
\[ = \sum_{j=0}^{5} (U \otimes U \otimes U)^j L (U^\dagger \otimes U^\dagger \otimes U^\dagger)^j. \]

A 3 qutrit quantum state \(|\phi\rangle\) that is shared by Alice, Bob and Charlie determines the 24 probabilities. Their sum is \(\langle \phi | A | \phi \rangle = \langle A \rangle\). The largest value for the sum is the largest eigenvalue of \(A\).

The eigenvalues of \(B\) are \(1, -1, e^{\pm i\pi/3}\) and \(e^{\pm 2i\pi/3}\), all of which are degenerate. Let \(P_b\) be the projection onto the subspace corresponding to the eigenvalue \(b\), \(B\) have an spectral decomposition

\[ B = \sum_b b P_b \]

with \([P_b, B] = 0\) and \(\sum_b P_b = I\). Note that

\[ B P_b = \sum_{b'} b' P_{b'} P_b \]
\[ = \sum_{b'} b' P_{b'} \delta_{b,b'} \]
\[ = b P_b. \]

Therefore we have

\[ A = \left( \sum_b P_b \right) \sum_{j=0}^{5} B^j L (B^\dagger)^j \left( \sum_{b'} P_{b'} \right) \]
\[ = \sum_b \sum_{b'} \left( \sum_{j=0}^{5} b^j (b'^*)^j \right) P_b L P_{b'} \]
\[ = \sum_b \sum_{b'} 6 \delta_{b,b'} P_b L P_{b'} \]
\[ = 6 \sum_b P_b L P_b. \]

In order the diagonalize \(A\), it is enough to diagonalize it within the eigensubspaces of \(B\). Call \(A^{(b)} = 6 P_b L P_b\).

It happens that the eigenstate corresponding to the maximum eigenvalue \(a_{\text{max}}\) lies in the subspace where \(B\) has the eigenvalue 1, which is a 7 dimensional. We are interested in \(A^{(1)}\).
Because of the form of the matrix

\[ M = \sum_{j=1}^{4} |\mu_j\rangle\langle \mu_j| \]

where \(|\mu_i\rangle = P_1 |\xi_i\rangle, |\mu_1\rangle = P_1 |021\rangle, |\mu_2\rangle = P_1 |00v_1\rangle, |\mu_3\rangle = P_1 |0v_00\rangle, \text{ and } |\mu_4\rangle = P_1 |v_020\rangle\), the problem can be reduced to a 4-dimensional one. \( M = A^{(1)}/6 \). If we express the eigenvector as \(|m\rangle = \sum_{j=1}^{4} c_j |\mu_j\rangle\), then the eigenvalue equation becomes

\[
M |m\rangle = m |m\rangle \\
M \sum_{k=1}^{4} c_k |\mu_k\rangle = m \sum_{k=1}^{4} c_k |\mu_k\rangle \\
\sum_{j=1}^{4} |\mu_j\rangle \langle \mu_j| \sum_{k=1}^{4} c_k |\mu_k\rangle = \\
\sum_{k=1}^{4} |\mu_k\rangle \left( \sum_{j=1}^{4} c_j \langle \mu_k| \mu_j\rangle \right) =
\]

Finding the overlaps of the vectors, we obtain

\[
\begin{pmatrix}
7 & 2 & -1 & -1 \\
2 & 7 & -1 & -1 \\
-1 & -1 & 7 & -1 \\
-1 & -1 & -1 & 7
\end{pmatrix}
\begin{pmatrix}
c_1 \\
c_2 \\
c_3 \\
c_4
\end{pmatrix}
= m
\begin{pmatrix}
c_1 \\
c_2 \\
c_3 \\
c_4
\end{pmatrix}.
\]

Largest eigenvalue \( m_{\text{max}} = 10/27 \), therefore \( a_{\text{max}} = 6m_{\text{max}} = 20/9 \). The corresponding eigenvector given in the computational basis is
\[
|\phi\rangle = \frac{1}{30\sqrt{3}}(-10(|000\rangle + |111\rangle + |222\rangle) + 14(|001\rangle + |112\rangle + |220\rangle) + 11(|002\rangle + |110\rangle + |221\rangle) - 7(|010\rangle + |121\rangle + |202\rangle) - 1(|011\rangle + |022\rangle + |100\rangle + |122\rangle + |200\rangle + |211\rangle) - 4(|012\rangle + |020\rangle + |101\rangle + |120\rangle + |201\rangle + |212\rangle) + 20(|021\rangle + |102\rangle + |210\rangle))
\]

Classical bound is found by assuming the existence of a joint probability distribution of all variables simultaneously, \(P(a_0, b_0, c_0; a_1, b_1, c_1)\), where each probability in the sequence is a marginal of this one. Like before the calculation is done by writing the Bell expression, the sum of probabilities, as

\[
\sum_{j=0}^{1} \sum_{a_j, b_j, c_j=0}^{2} c_{a_0, b_0, c_0; a_1, b_1, c_1} P(a_0, b_0, c_0; a_1, b_1, c_1).
\]

After translated in this form biggest coefficient \(c_{a_0, b_0, c_0; a_1, b_1, c_1}\) corresponds to the classical bound. Which is 2 in this case.

\[
\frac{20}{9} > 2 \Rightarrow q > c
\]

which is a Bell violation.

We can also do a non-local game analysis of the sequence. With 3 parties the questions are made of 3 bits \((s, t, u)\). The correct answers are given in the Table 6.

The deterministic strategy is \(a(s) = 0, b(t) = 2\) and \(c(u) = 1\) wins with the probability \(w_c = 1/4\). The fact that this is the classical bound can be shown if we let \(F(a, b, c; s, t, u)\), where \(a, b, c \in \{0, 1, 2\}\) be equal to 1 when \((a, b, c; s, t, u)\) is a winning condition for the game and 0 otherwise. Then

\[
w_c = \frac{1}{8} \sum_{a', b', c'=0}^{2} \sum_{s, t, u=0}^{1} F(a, b, c; s, t, u) \delta_{a', a(s)} \delta_{b', b(t)} \delta_{c', c(u)}.
\]
### Table 6: Winning conditions for the non-local game of $N = 3, N = 2, K = 3$ scenario

<table>
<thead>
<tr>
<th>s, t, u</th>
<th>Alice, Bob, Charlie</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>021, 102, 210,</td>
</tr>
<tr>
<td>111</td>
<td>021, 102, 210,</td>
</tr>
<tr>
<td>001</td>
<td>001, 112, 220,</td>
</tr>
<tr>
<td>110</td>
<td>002, 110, 221,</td>
</tr>
<tr>
<td>010</td>
<td>000, 111, 222,</td>
</tr>
<tr>
<td>101</td>
<td>010, 121, 202,</td>
</tr>
<tr>
<td>100</td>
<td>020, 101, 212,</td>
</tr>
<tr>
<td>011</td>
<td>012, 120, 201,</td>
</tr>
</tbody>
</table>

Compare this with the Bell expression which is the sum of our 24 probabilities, that can be expressed as

$$\text{LHS} = \sum_{a', b', c'=0}^{2} \sum_{s, t, u=0}^{1} F(a', b', c'; s, t, u)p(a_s = a', b_t = b', c_u = c').$$

We found that the classical bound of the Bell expression is 2. Noting that $\delta_{a,A(s)}\delta_{b,B(t)}\delta_{c,C(u)}$ can be derived from a joint distribution, in particular

$$P(a_0, b_0, c_0; a_1, b_1, c_1) = \delta_{a_0,A(0)}\delta_{a_1,A(1)}\delta_{b_0,B(0)}\delta_{b_1,B(1)}\delta_{c_0,C(0)}\delta_{c_1,C(1)}$$

we see that the sum in $8w_c$ is less than or equal to 2, which implies that the classical winning strategy must be less than or equal to $1/4$.

The quantum strategy in which each party responds with the outcome of a measurement done in the asked basis, the probability of winning is just $1/8$ times the sum of the probabilities of the winning configurations, which we have seen is $20/9$. This gives an overall probability of $5/18$, which is approximately 0.28, and this is greater than the winning probability of the classical strategy.

$$\frac{5}{18} \approx 0.28 > \frac{20}{9} \approx 0.22$$

which is a Bell violation.

### 3.5 Bell Inequalities from Group Actions: Non-Abelian Groups

In this section we incorporate the group theory ideas to the sequence generation method.
3.5.1 Reproduction of CH Using a Cyclic Group (Reprise)

We again start with a simple case and use a cyclic group to reproduce the CH inequality, which is the best way to learn how a Bell inequality generating scheme works.

A cyclic group of order \( n \), \( \mathbb{Z}_n \), is isomorphic to \((\mathbb{Z}_n, +)\). It has one generator \( a \) for which \( a^n = e \) and from which the rest of the group elements can be generated \( G = \{e, a, a^2, \ldots, a^{n-1}\} \). Therefore its presentation is \( \mathbb{Z}_n = \{a|a^n\} \).

Let us find a representation \( \Gamma : \mathbb{Z}_n \rightarrow \mathcal{H} \). The dimensionality of \( \mathcal{H} \) determines the number of outcomes of the measurements, \( \dim \mathcal{H} = K \). Therefore we choose \( \dim \mathcal{H} = 2 \).

We are dealing with non-degenerate systems. In general \( \dim \mathcal{H} \geq K \) but the observables have \( K \) distinct eigenvalues.

The representation matrix of the identity element is the identity matrix

\[
\Gamma(e) = I_{K \times K} = I_{2 \times 2}
\]

Therefore

\[
\Gamma(a^n) = I
\]

\[
\Gamma^n(a)
\]

Which means that \( \Gamma(a) \) is an \( n \text{th} \) root of identity matrix. How should we choose \( n \)?

Just like in the Section 3.3.1 where we used the translation-swap operators scheme we need 4 events corresponding to 2 outcome per measurement choice and 2 measurement choices where the corresponding quantum states are

\[
|x_1 = 0\rangle, |x_1 = 1\rangle, |x_2 = 0\rangle, |x_2 = 1\rangle.
\]

To get these states as the results of group action we need a group of order at least 4. Let us choose \( n = 4 \).

We want to construct a 2 dimensional representation of \( \mathbb{Z}_4 \). Because a cyclic group has only 1 generator, the representation matrix for the group generator is enough to construct the rest of the representation matrices,

\[
\Gamma(a^n) = \Gamma^n(a).
\]

We said that abelian groups have conjugacy classes made of single elements, and in general the number of unequivalent irreducible representations is equal to the number of conjugacy classes. Therefore a cyclic
group have \( n \) representation that are \( 1 \times 1 \) dimensional, namely complex numbers.

For \( n = 4 \) a set of irreducible representations are:

\[
\begin{align*}
\{ \Gamma^{(1)}(a^k) | k = 0, 1, 2, 3 \} &= \{1, 1, 1, 1\} \\
\{ \Gamma^{(2)}(a^k) | k = 0, 1, 2, 3 \} &= \{1, -1, 1, -1\} \\
\{ \Gamma^{(3)}(a^k) | k = 0, 1, 2, 3 \} &= \{1, i, -1, -i\} \\
\{ \Gamma^{(4)}(a^k) | k = 0, 1, 2, 3 \} &= \{1, -i, -1, i\}
\end{align*}
\]

In these different representations the elements that correspond to the generator are \(1, -1, i\) and \(-i\). All of them are 4th order roots of 1. We can construct a \( 2 \times 2 \) matrix by putting 2 of these 4 numbers on the diagonal of a matrix. Then take an arbitrary similarity transformation.

\[
\Gamma(a) = S \begin{pmatrix} (\sqrt[4]{1})_1 & 0 \\ 0 & (\sqrt[4]{1})_2 \end{pmatrix} S^{-1}
\]

\[
\Gamma^4(a) = S \begin{pmatrix} (\sqrt[4]{1})_1 & 0 \\ 0 & (\sqrt[4]{1})_2 \end{pmatrix} S^{-1} S \begin{pmatrix} (\sqrt[4]{1})_1 & 0 \\ 0 & (\sqrt[4]{1})_2 \end{pmatrix} S^{-1} \cdots
\]

\[
= S \begin{pmatrix} (\sqrt[4]{1})_1 & 0 \\ 0 & (\sqrt[4]{1})_2 \end{pmatrix}^4 S^{-1}
\]

\[
= S S^{-1}
\]

\[
= I
\]

Our choice for \( \Gamma(a) \) is

\[
\Gamma(a) = S \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} S^{-1}
\]

where 1 and \( i \) are chosen as \( \sqrt[4]{1} \). We needed at least one \( i \) (or \(-i\)) so that we can get a faithful representation (different matrices for different group elements). Note
The similarity transformation we apply is done by

\[
S = \frac{1}{\sqrt{2}}\begin{pmatrix}
1 & 1 \\
1 & -1 \\
\end{pmatrix}
\]

which is the transformation from \( z \)-basis, \( \{|z+\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |z-\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \} \), to \( x \)-basis.

\[
|x+\rangle \langle z+| + |x-\rangle \langle z-| = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}
\]

\[
S_{z\rightarrow x} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}
= S_{x\rightarrow z} = S^{-1}
\]

According to these choices the \( \Gamma (a) \) we get is

\[
\Gamma (a) = \begin{pmatrix} 1/2 + i/2 & 1/2 - i/2 \\ 1/2 - i/2 & 1/2 + i/2 \end{pmatrix} \\
= |x+\rangle \langle x+| + i |x-\rangle \langle x-|
\]

Note that \( \Gamma (a^2) = \Gamma ^2 (a) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \sigma_x = T \) which can be interpreted as the translation operator, for which \( T |0\rangle = |1\rangle, \quad T |1\rangle = |0\rangle \).

These choices are justified by getting a Bell violation at the end of the process.

When \( |\psi_0\rangle = |0\rangle \) is chosen as the initial state and \( U = \Gamma (a) \) is applied on it successively we get the sequence \( |\psi_i\rangle = \Gamma (a^i) |\psi_0\rangle \) where \( i = 0, 1, 2, 3 \).
\[
\Gamma (a^0) |\psi_0\rangle = I |0\rangle = |0\rangle \\
\Gamma (a^1) |\psi_0\rangle = \frac{1}{\sqrt{2}}(e^{i\pi/4}|0\rangle + e^{-i\pi/4}|1\rangle) \equiv |v_0\rangle \\
\Gamma (a^2) |\psi_0\rangle = T |0\rangle = |1\rangle \\
\Gamma (a^1) |\psi_0\rangle = \frac{1}{\sqrt{2}}(e^{-i\pi/4}|0\rangle + e^{i\pi/4}|1\rangle) \equiv |v_1\rangle .
\]

This is a cyclic sequence

\[
|0\rangle \rightarrow |v_0\rangle \rightarrow |1\rangle \rightarrow |v_1\rangle \rightarrow |0\rangle \rightarrow \cdots
\]

where we generated two orthonormal bases \{ |0\rangle, |1\rangle \}, \{ |v_0\rangle, |v_1\rangle \} after we classify them by their absolute inner products \| \langle \psi_i | \psi_k \rangle \|.

| \langle \psi_i | \psi_k \rangle | |0\rangle | |v_0\rangle | |1\rangle | |v_1\rangle |
|---|---|---|---|---|
| \langle 0 \rangle | 1 | \frac{1}{\sqrt{2}} | 0 | \frac{1}{\sqrt{2}} |
| \langle v_0 \rangle | \frac{1}{\sqrt{2}} | 1 | \frac{1}{\sqrt{2}} | 0 |
| \langle 1 \rangle | 0 | \frac{1}{\sqrt{2}} | 1 | \frac{1}{\sqrt{2}} |
| \langle v_1 \rangle | \frac{1}{\sqrt{2}} | 0 | \frac{1}{\sqrt{2}} | 1 |

This means that they are mutually unbiased bases (MUB). Two orthonormal bases, \( B_1 \) and \( B_2 \), are mutually unbiased when absolute inner products of vectors, \( b_{1,i} \in B_1 \) and \( b_{2,j} \in B_2 \), is always the same \| \langle b_{1,i} | b_{2,j} \rangle \| = c.

Now we define the group action on \( \mathbb{C}^2 \otimes \mathbb{C}^2 \),

\[
\alpha (g, |\Psi\rangle) : G \times \mathcal{H} \rightarrow \mathcal{H}, \quad \mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2
\]

as

\[
\alpha (g, |\Psi\rangle) = \Gamma (g) \otimes \Gamma (g) |\Psi\rangle.
\]

Remember that the subset \{ \alpha (g, x) | g \in G \} for a group action \( \alpha : G \times X \rightarrow X \) is called an orbit. And orbits partition the set \( X \) into disjoint subsets. In our case the different orbits generated by different \( |\psi^{(i)}_0\rangle \) will partition \( \mathbb{C}^2 \otimes \mathbb{C}^2 \).
We choose $|\psi_0^{(1)}\rangle = |0\rangle |0\rangle$ and $|\psi_0^{(2)}\rangle = |0\rangle |v_0\rangle$ and calculate their orbits

$$\left\{ \Gamma (g) \otimes \Gamma (g) \left| \psi_0^{(i)} \right\rangle \big| g \in Z_4, i = 1, 2 \right\}.$$ 

The orbits are $|0\rangle |0\rangle \rightarrow |v_0\rangle |v_0\rangle \rightarrow |1\rangle |1\rangle \rightarrow |v_1\rangle |v_1\rangle$ and $|0\rangle |v_0\rangle \rightarrow |v_0\rangle |1\rangle \rightarrow |1\rangle |v_1\rangle \rightarrow |v_1\rangle |0\rangle$. The rest is the same as the previous schemes.

$$L = \sum_i |\psi_0^{(i)}\rangle \langle \psi_0^{(i)}|$$

$$A = \sum_{g \in G} (\Gamma (g) \otimes \Gamma (g)) L (\Gamma (g) \otimes \Gamma (g))^\dagger$$

The goal is to calculate largest eigenvalue of $A$ and compare it with the classical bound.

$$A = \frac{1}{2} \begin{pmatrix} 5 & -i & i & -1 \\ i & 3 & 1 & -i \\ -i & 1 & 3 & i \\ -1 & i & -i & 5 \end{pmatrix}$$

of which eigenvalues are $2 + \sqrt{2}$, $2 - \sqrt{2}$, 2 and 2, where $a_{\max} = 2 + \sqrt{2}$.

The probabilities from two orbits are $p(a_0 = 0, b_0 = 0), p(a_1 = 0, b_1 = 0), p(a_0 = 1, b_0 = 1), p(a_1 = 1, b_1 = 1)$ and $p(a_0 = 0, b_1 = 0), p(a_1 = 0, b_0 = 1), p(a_0 = 1, b_1 = 1), p(a_1 = 1, b_0 = 0)$. When their sum is expressed as a convex linear combination of the elements of probability distribution $p(a_0 = i, a_1 = j, b_0 = k, b_1 = l) \equiv p_{ijkl}$

$$\beta \cdot p = \sum_{ijkl} \beta_{ijkl} p_{ijkl}$$

with $\beta = (3, 3, 1, 1, 3, 1, 1, 1, 3, 3, 3, 1, 1, 1)$ where the biggest component is 3, which is the classical bound.

$$2 + \sqrt{2} > 3$$

$$q > c$$

which is a Bell violation.
3.5.2 Properties of Dihedral Group

In this section we are going to use a simple non-abelian group, Dihedral Group, to generate our orbits. Remember the dihedral group of degree \( n \) is the group of rotational and reflectional symmetries of \( n \)-sided regular polygon. The polygon is centered at the coordinate center, and the rotations and reflections are around the coordinate center too.

Symmetric group, \( S_n \), includes all possible permutations of vertices, whereas Dihedral group consists of transformations that keep the rigid body structure.

It has 2 generators, \( r, \) \( 1/n \) fraction of full rotation and \( s, \) reflection. Its presentation is \( \langle r, s|r^n, s^2, (sr)^2 \rangle \).

\( r^n = e \) means that if we do \( 1/n \) fraction of full rotation \( n \) times we get a full rotation and return the initial state. \( s^2 = e \) means that if we do reflection twice we return to the initial state. \( (sr)^2 = e \) or \( srs = r^{-1} \) means that a rotation that is applied in between two reflections is a rotation in the opposite direction. These three relations defines the dihedral group.

\( D_n \) has \( 2n \) elements.

\[
D_1 = \{e, r\} = Z_2 \\
D_2 = \{e, r, s, sr\} = V_4 \\
D_3 = \{e, r, r^2, s, sr, sr^2\} = S_3 \\
D_4 = \{e, r, r^2, r^3, s, sr, sr^2, r^3\} \\
D_5 = \{e, r, r^2, r^3, r^4, s, sr, sr^2, r^3, sr^4\}
\]

etc.

Let us analyze the conjugacy classes of \( D_n \), which will be helpful in calculating the character tables.

For \( n = \text{odd} \), There is one conjugacy class for the identity element,

\[
C_e = \{geg^{-1}|g \in D_n\} = \{e\}.
\]

For elements in the subgroup \( \langle r \rangle \),

\[
C_{r^j} = \{gr^jg^{-1}|g \in D_n\}
\]

we can express the elements of \( D_n \) in two different forms, \( r^k \) and \( sr^k \), \( k = 0..n-1 \).

\[
C_{r^j} = \{r^k r^j r^{-k}|k = 0..n-1\} \cup \{sr^k r^j (sr^k)^{-1}|k = 0..n-1\}.
\]
For the elements that are outside of the subgroup \( \langle r \rangle \), which are of the form \( sr^j \)

\[
C_{sr^j} = \{ sr^j, sr^j^{-1} \} \cup \{ sr^j r^k | k = 0..n-1 \}
\]

For odd \( n \), these two cases will cover all elements outside of \( \langle r \rangle \) independent of \( j \).

\[
C_{sr^j} = \{ sr^{j-2k} | k = 0..n-1 \} \cup \{ sr^{-j+2k} | k = 0..n-1 \} = \{ sr^k | k = 0..n-1 \}.
\]

So, the elements outside of \( \langle r \rangle \) form a single conjugacy class of order \( n \). In brief there are 1 class of order 1, \( (n-1)/2 \) classes of order 2, and 1 class of order 1, in total \( (n+3)/2 \) conjugacy classes.

If we do a similar analysis for \( n = \text{even} \), we’ll have \( C_e = \{ e \} \).

Again group elements of the form \( r^j \) will form classes of order 2, \( \{ r^j, r^{-j} = r^{n-j} \} \), except \( r^{n/2} \) because its inverse is itself. So, we have \( C_{r^j} = \{ r^j, r^{n-j} \} \) for \( j = 1..(n-2)/2 \) and \( C_{r^{n/2}} = \{ r^{n/2} \} \).

For the elements outside of \( \langle r \rangle \)

\[
C_{sr^j} = \{ sr^j r^k | k = 0..n-1 \} \cup \{ sr^j (sr^k)^{-1} | k = 0..n-1 \} = \{ sr^{j-2k} | k = 0..n-1 \} \cup \{ sr^{-j+2k} | k = 0..n-1 \}
\]

For even \( n \), there are two distinct \( C_{sr^j} \) for odd and even \( j \).

\[
C_{sr,\text{even}} = \{ sr^{2k} | k = 0..n-1 \}
\]

\[
C_{sr,\text{odd}} = \{ sr^{2k+1} | k = 0..n-1 \}.
\]

So, we have 2 classes of order \( n/2 \).
In brief, there are 2 classes of order 1, \((n - 2) / 2\) classes of order 2, and 2 classes of order \(n / 2\), in total \((n + 6) / 2\) classes.

These information is reflected in the group representation characters. The number of conjugacy classes is equal to number of irreducible representations (irreps).

For example for \(n = \text{odd}\) there are 2 1-dimensional irreducible representations and \((n - 1) / 2\) 2-dimensional irreducible representations. First 1-dim irrep is the trivial irrep

\[
\Gamma^{(1)}(g) = 1.
\]

Second 1-dim irrep sends elements in \(\langle r \rangle\) to 1 and the rest to \(-1\).

\[
\Gamma^{(2)}(g) = \begin{cases} 
1 & g \in \langle r \rangle \\
-1 & g \notin \langle r \rangle.
\end{cases}
\]

The two common interpretations for the dihedral representations are rotations on real-plane and rotations on complex plane.

The \(\nu^{\text{th}}\) 2-dim representation of \(D_{n=\text{odd}}\), \(\Gamma^{(3+\nu)}(g)\), can be:

<table>
<thead>
<tr>
<th>(g)</th>
<th>real orthogonal</th>
<th>complex unitary</th>
<th>(\chi(g))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(r)</td>
<td>(\begin{pmatrix} \cos(2\pi \nu / n) &amp; -\sin(2\pi \nu / n) \ \sin(2\pi \nu / n) &amp; \cos(2\pi \nu / n) \end{pmatrix})</td>
<td>(\begin{pmatrix} \exp(2\pi \nu / n) &amp; 0 \ 0 &amp; \exp(-2\pi \nu / n) \end{pmatrix})</td>
<td>(2 \cos(2\pi \nu / n))</td>
</tr>
<tr>
<td>(r')</td>
<td>(\begin{pmatrix} \cos(2\pi \nu / n) &amp; -\sin(2\pi \nu / n) \ \sin(2\pi \nu / n) &amp; \cos(2\pi \nu / n) \end{pmatrix})</td>
<td>(\begin{pmatrix} \exp(2\pi \nu / n) &amp; 0 \ 0 &amp; \exp(-2\pi \nu / n) \end{pmatrix})</td>
<td>(2 \cos(2\pi \nu / n))</td>
</tr>
<tr>
<td>(s)</td>
<td>(\begin{pmatrix} 1 &amp; 0 \ 0 &amp; -1 \end{pmatrix})</td>
<td>(\begin{pmatrix} 0 &amp; 1 \ 1 &amp; 0 \end{pmatrix})</td>
<td>0</td>
</tr>
<tr>
<td>(sr^l)</td>
<td>(\begin{pmatrix} \cos(2\pi \nu / n) &amp; -\sin(2\pi \nu / n) \ \sin(2\pi \nu / n) &amp; \cos(2\pi \nu / n) \end{pmatrix})</td>
<td>(\begin{pmatrix} 0 &amp; \exp(2\pi \nu / n) \ \exp(-2\pi \nu / n) &amp; 0 \end{pmatrix})</td>
<td>0</td>
</tr>
</tbody>
</table>

Similarly, for \(n = \text{even}\) there are 4 1-dimensional irreps and \((n - 2) / 2\) 2-dimensional irreps.

First two 1-dim representations are the same as the odd case.

Third 1-dim irrep is

\[
\Gamma^{(3)}(g) = \begin{cases} 
1 & g \in \langle r^2, s \rangle \\
-1 & g \notin \langle r^2, s \rangle.
\end{cases}
\]
And fourth one is

$$\Gamma^{(4)}(g) = \begin{cases} 
1 & g \in \langle r^2, s \rangle \\
-1 & g \notin \langle r^2, sr \rangle.
\end{cases}$$

The $\nu^{th}$ 2-dimensional representation of $D_{n=\text{even}}$, $\Gamma^{(5+\nu)}(g)$, are the same as the odd case, except $\nu = 0..(n-2)/2$ instead of $(n-1)/2$.

Using this information it is possible to fill the character tables. Remember that the first row is always 1, due to trivial representation. In general, the 1 dimensional representations can be written down directly.

For example for $D_3$

<table>
<thead>
<tr>
<th></th>
<th>$C_e$</th>
<th>$C_r$</th>
<th>$C_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Gamma^{(1)}$</td>
<td>$\text{Tr}(\Gamma^{(1)}(e))$</td>
<td>$\text{Tr}(\Gamma^{(1)}(r))$</td>
<td>$\text{Tr}(\Gamma^{(1)}(s))$</td>
</tr>
<tr>
<td>$\Gamma^{(2)}$</td>
<td>$\text{Tr}(\Gamma^{(2)}(e))$</td>
<td>$\text{Tr}(\Gamma^{(2)}(r))$</td>
<td>$\text{Tr}(\Gamma^{(2)}(s))$</td>
</tr>
<tr>
<td>$\Gamma^{(3)}$</td>
<td>$\text{Tr}(\Gamma^{(3)}(e))$</td>
<td>$\text{Tr}(\Gamma^{(3)}(r))$</td>
<td>$\text{Tr}(\Gamma^{(3)}(s))$</td>
</tr>
</tbody>
</table>

$$= \begin{array}{ccc}
D_3 & C_e & C_r & C_s \\
\Gamma^{(1)} & 1 & 1 & 1 \\
\Gamma^{(2)} & 1 & 1 & -1 \\
\Gamma^{(3)} & 2 & -1 & 0
\end{array}$$

where $2 \cos\left(\frac{2\pi\nu}{n}\right) = -1$ for $n = 3, \nu = 1$.

If we do the same calculation for $D_6$ the character table we get is

<table>
<thead>
<tr>
<th></th>
<th>$C_e$</th>
<th>$C_r$</th>
<th>$C_{r2}$</th>
<th>$C_{r3}$</th>
<th>$C_s$</th>
<th>$C_{rs}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Gamma^{(1)}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\Gamma^{(2)}$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>$\Gamma^{(3)}$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>$\Gamma^{(4)}$</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>$\Gamma^{(5)}$</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\Gamma^{(6)}$</td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

These tables will be helpful in future calculations of eigenvalues of the $A$ operator etc.
3.5.3 A Bell Violation Using $D_3$

In our work we are going to use the representation with the interpretation of rotations on the real plane. The matrices corresponding to the generators are

$$U = \Gamma(3) (r) = \begin{pmatrix} \cos(\frac{2\pi n}{3}) & -\sin(\frac{2\pi n}{3}) \\ \sin(\frac{2\pi n}{3}) & \cos(\frac{2\pi n}{3}) \end{pmatrix}, \quad n = 3, \nu = 1$$

$$= \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}$$

and

$$V = \Gamma(3) (s) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$ 

The group elements are \{e, r, r^2, s, rs, r^2 s\} and the corresponding representation matrices are \{I, U, U^2, V, UV, U^2 V\}.

The sequence of quantum states when started from the initial state $|x+\rangle$ generates the measurement bases.\{ |x+\rangle, |u_0\rangle, |v_0\rangle, |x-\rangle, |u_1\rangle, |v_1\rangle \}. First observable is in $x$-basis \{|x+\rangle, |x-\rangle\}, the second one is in $u$-basis \{|u_0\rangle, |u_1\rangle\}, and the third one is in $v$-basis \{|v_0\rangle, |v_1\rangle\}.

Figure 18 shows the vectors corresponding to the quantum states. Because the components of the states are all real, these qubit states can be shown on a cartesian plane instead of a Bloch sphere.

In this scheme the application of representing matrix $U$ of first generator $r$ changes the measurement basis while keeping the outcome the same. $|x+\rangle \rightarrow |u_0\rangle \rightarrow |v_0\rangle \rightarrow |x+\rangle \rightarrow \cdots$ and likewise $|x-\rangle \rightarrow |u_1\rangle \rightarrow |v_1\rangle \rightarrow |x-\rangle \rightarrow \cdots$.

The representation matrix $V$ of the second generator $s$ changes the measurement outcome while keeping the basis the same. $|x+\rangle \leftrightarrow |x-\rangle, |u_0\rangle \leftrightarrow |u_1\rangle$ and $|v_0\rangle \leftrightarrow |v_1\rangle$.

This structure can also be seen in the Cayley graph of $D_3$ in Figure 19. A Cayley graph is a graph where vertices represent group elements, and edges represent the connection between the elements in terms of the generators. The generators are chosen as $r = (1, 2, 3)$ and $g = (1, 3)$.

According to this choice, the group elements end their disjoint cycle notation expressions are

<table>
<thead>
<tr>
<th></th>
<th>e</th>
<th>r</th>
<th>r^2</th>
<th>s</th>
<th>rs</th>
<th>r^2s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>()</td>
<td>(1,2,3)</td>
<td>(1,3,2)</td>
<td>(1,3)</td>
<td>(2,3)</td>
<td>(1,2)</td>
</tr>
</tbody>
</table>

As an example, a directed edge from $a$ to $b$ via $g$ means that $g \ast a = b$. For example the red edge
Figure 18: The quantum states generated by the group action. The bases are indicated with color. $x$-basis is red, $u$-basis is green and $v$-basis is blue. The group elements give the state is indicated at the tip of the state vector.

$r = (1, 2, 3)$ from $(1, 2)$ to $(1, 3)$ means that $r * r^2 s = es = s$ etc.

Now we can generate the joint outcomes. For that, we use a representation from $D_3$ to $\mathbb{C}^2 \otimes \mathbb{C}^2$ given by $\Gamma^{(3)}(g) \otimes \Gamma^{(3)}(g)$ where $\Gamma^{(3)}(g)$ is the two dimensional representation that we just constructed. For the joint events the group action is $\alpha(g, |\Psi\rangle) = \Gamma^{(3)}(g) \otimes \Gamma^{(3)}(g) |\Psi\rangle$ where $|\Psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$. The orbit, if we start from $|x+\rangle |x+\rangle$, is

$$\{ |x+, x\rangle, |u_0, u_0\rangle, |v_0, v_0\rangle, |x-, x\rangle, |u_1, u_1\rangle, |v_1, v_1\rangle \},$$

and another orbit if we start from $|x-\rangle |v_0\rangle$ is

$$\{ |x-, v_0\rangle, |u_1, +x\rangle, |v_1, u_0\rangle, |+x, u_1\rangle, |u_0, v_1\rangle, |v_0, -x\rangle \}.$$
and due to the second orbit are

\[ \{ p(a_0 = 1, b_2 = 0), p(a_1 = 1, b_0 = 0), \]
\[ p(a_2 = 1, b_1 = 0), p(a_0 = 0, b_1 = 1), \]
\[ p(a_1 = 0, b_2 = 1), p(a_2 = 0, b_0 = 1) \}. \]

The corresponding \( A \) operator is

\[ A = \sum_{g \in D_3} \left( \Gamma^{(3)}(g) \otimes \Gamma^{(3)}(g) \right) L \left( \Gamma^{(3)}(g) \otimes \Gamma^{(3)}(g) \right)^\dagger \]

where

\[ L = |x+, x+\rangle \langle x+, x+ | + |x-, v_0\rangle \langle x-, v_0|. \]

To calculate the largest eigenvalue of \( A \), \( a_{\text{max}} \), we make use of group representation theory. The representation \( \Gamma^{(3)}(g) \otimes \Gamma^{(3)}(g) \) is reducible. We first split it into a direct sum irreducible components. To do that we first employ orthogonality theorem (See Section 3.2.5) for characters.
\[ n_p = \frac{1}{|G|} \sum_g \chi(g) \chi^{(p)\ast}(g), \]

which gives the number of times, \( n_p \), an irreducible representation of a group \( G \), \( \Gamma^{(p)} \) appears in the decomposition of a representation \( \Gamma \). \( |G| \) is the order (number of elements) of \( G \), \( \chi(g) \) is the character of \( \Gamma(g) \), and \( \chi^{(p)}(g) \) is the character of \( \Gamma^{(p)}(g) \).

\( \chi(g) \) is the character for \( \Gamma^{(3)}(g) \otimes \Gamma^{(3)}(g) \), and \( \chi^{(p)}(g) \) is the character of \( \Gamma^{(p)} \) for \( p = 1, 2, 3 \), the 3 irreducible representations of \( D_3 \) that we calculated.

Because for two matrices \( E \) and \( F \), \( \text{Tr}(E \otimes F) = \text{Tr}(E) \text{Tr}(F) \),

\[
\chi(g) = \text{Tr} \left( \Gamma^{(3)}(g) \otimes \Gamma^{(3)}(g) \right) \\
= \text{Tr}(\Gamma^{(3)}(g)) \text{Tr}(\Gamma^{(3)}(g)) \\
= \text{Tr}(\Gamma^{(3)}(g))^2 \\
= \chi^{(3)}^2 
\]

which gives

<table>
<thead>
<tr>
<th>( D_3 )</th>
<th>( C_e )</th>
<th>( C_r )</th>
<th>( C_s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Gamma )</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
n_1 = \frac{1}{6} (4 \times 1 + 1 \times 1 + 1 \times 1 + 0 \times 1 + 0 \times 1 + 0 \times 1) = \frac{6}{6} = 1 \\
n_2 = \frac{1}{6} (4 \times 1 + 1 \times 1 + 1 \times 1 + 0 \times (-1) + 0 \times (-1) + 0 \times (-1)) = \frac{6}{6} = 1 \\
n_3 = \frac{1}{6} (4 \times 2 + 1 \times (-1) + 1 \times (-1) + 0 \times 0 + 0 \times 0 + 0 \times 0) = \frac{6}{6} = 1 
\]

This means the decomposition of \( \Gamma^{(3)}(g) \otimes \Gamma^{(3)}(g) \) includes 1 \( \Gamma^{(1)} \), 1 \( \Gamma^{(2)} \) and 1 \( \Gamma^{(3)} \).

\[
\Gamma^{(3)}(g) \otimes \Gamma^{(3)}(g) = \Gamma^{(1)} \oplus \Gamma^{(2)} \oplus \Gamma^{(3)}. 
\]

As we said in Section 3.2.8 each subrepresentation in the decomposition only acts in a subspace and the decomposition of the bigger representation into irreducible representations also corresponds to the decom-
position of $\mathbb{C}^2 \otimes \mathbb{C}^2$ into subspaces

$$\mathbb{C}^2 \otimes \mathbb{C}^2 = \mathcal{V}(1) \oplus \mathcal{V}(2) \oplus \mathcal{V}(3)$$

where $\mathcal{V}^{(p)}$ is the subspace on which $\Gamma^{(p)}$ acts. These invariant subspaces are basically the eigensubspaces of the linear operator $\Gamma (g) = \Gamma^{(3)} (g) \otimes \Gamma^{(3)} (g)$. The eigenstates of $\Gamma (g)$ is the Bell basis, where

$$\begin{align*}
|\Phi^+\rangle &= (|00\rangle + |11\rangle) / \sqrt{2} \\
|\Phi^-\rangle &= (|00\rangle - |11\rangle) / \sqrt{2} \\
|\Psi^+\rangle &= (|01\rangle + |10\rangle) / \sqrt{2} \\
|\Psi^-\rangle &= (|01\rangle - |10\rangle) / \sqrt{2}
\end{align*}$$

When all $\Gamma (g)$ is applied on these vectors, we might get different eigenvalues, or eigenvectors but overall we observe

$$\Gamma (r) |\Phi^+\rangle = |\Phi^+\rangle \quad \Gamma (s) |\Phi^+\rangle = |\Phi^+\rangle$$
$$\Gamma (r) |\Psi^-\rangle = |\Psi^-\rangle \quad \Gamma (s) |\Psi^-\rangle = -|\Psi^-\rangle$$

In other words

$$\Gamma (g) |\Phi^+\rangle \in \text{span} \{ |\Phi^+\rangle \} = \mathcal{V}(1)$$
$$\Gamma (g) |\Psi^-\rangle \in \text{span} \{ |\Psi^-\rangle \} = \mathcal{V}(2)$$
$$\Gamma (g) (\alpha |\Phi^-\rangle + \beta |\Psi^+\rangle) \in \text{span} \{ |\Phi^-\rangle, |\Psi^+\rangle \} = \mathcal{V}(3)$$

$|\Phi^+\rangle$ always goes to itself with the eigenvalue 1. Hence it is the invariant subspace of the trivial representation, $\mathcal{V}(1)$. $|\Psi^-\rangle$ gets the eigenvalue $-1$ when $g \in \{ s, rs, r^2 s \}$. Hence it is the subspace related to the sign representation, $\mathcal{V}(2)$. $\Gamma (g)$ rotates $|\Phi^-\rangle$ and $|\Psi^+\rangle$ in the plane they span, $\mathcal{V}(3)$.

Now we will use the great orthogonality theorem that was explained in Section 3.2.3

$$\frac{1}{|G|} \sum_{g \in G} \Gamma^{(p)}(g)^* \Gamma^{(q)}(g)_{j'k'} = \frac{1}{d_p} \delta_{pq} \delta_{j'j} \delta_{k'k}.$$ 

Let $\{ |\alpha_j^{(p)}\rangle \}$ be an orthonormal basis of a carrier space for the irreducible representation $\Gamma^{(p)}$, $|X_p\rangle$ a vector in that space, $\{ |\beta_j^{(q)}\rangle \}$ an orthonormal basis for a carrier space for the irreducible representation $\Gamma^{(q)}$, and $|X_q\rangle$ a vector in that space. Start with this expression
Thanks to the orthogonality theorem this expression is equal to

$$\sum_{k,k'} \frac{1}{d_p} \delta_{pq} \delta_{jj'} \delta_{kk'} \langle \alpha_k^{(p)} | X_p \rangle \langle X_q | \beta_k^{(q)} \rangle = \sum_k \frac{1}{d_p} \delta_{pq} \delta_{jj'} \langle \alpha_k^{(p)} | X_p \rangle \langle X_q | \beta_k^{(q)} \rangle.$$

Note that the irreducible representations of the direct sum decomposition only appeared twice $\Gamma^{(3)}(g) \otimes \Gamma^{(3)}(g) = \Gamma^{(1)} \oplus \Gamma^{(2)} \oplus \Gamma^{(3)}$. There is only 1 $\Gamma^{(p)}$ for every $p$, in other words $n_p = 1 \forall p$. This does not have to be the case.

The general case is

$$\mathcal{V} = \mathcal{V}^{(1)} \oplus \mathcal{V}^{(1)} \oplus \ldots \oplus \mathcal{V}^{(1)} \oplus \mathcal{V}^{(2)} \oplus \mathcal{V}^{(2)} \oplus \ldots \oplus \mathcal{V}^{(2)} \oplus \ldots$$

$$= \bigoplus_{p=1}^{n_p} \bigoplus_{j=1}^{n_p} \mathcal{V}_j^{(p)}$$

where we denoted $\mathcal{V}_j^{(p)}$ as the carrier space of $\Gamma_j^{(p)}$ which is the $j$-th appearance of the irreducible representation $\Gamma^{(p)}$ in the decomposition. Therefore we have two cases for $p = q$. One in which the carrier spaces are the same $\text{span} \left\{ \langle \alpha_j^{(p)} \rangle \right\} = \text{span} \left\{ \langle \beta_j^{(p)} \rangle \right\}$, and other one in which the carrier spaces are different $\text{span} \left\{ \langle \alpha_j^{(p)} \rangle \right\} \neq \text{span} \left\{ \langle \beta_j^{(p)} \rangle \right\}$.

In our case $\Gamma(g)$ is a representation which is the direct sum of irreducible representations each of which only appears once. For $p = q$ the carrier spaces are the same, namely $|\alpha_j^{(p)}\rangle = |\beta_j^{(q)}\rangle$ when we have $p = q$. In this case the expression reduces to
\[
\frac{1}{d_p} \delta_{pq} \delta_{jj'} \|X_p\|^2.
\]

Then any vector \(|\psi\rangle\) can be expressed as a linear combination of vectors in invariant subspaces

\[
|\psi\rangle = \sum_q c'_q |\psi_q\rangle
\]

where \(|\psi_q\rangle \in V^{(q)}\), if expressed in terms of \(\alpha_j^{(q)}\)

\[
|\psi_q\rangle = \sum_q \sum_j c_{q,j} |\alpha_j^{(q)}\rangle.
\]

Then

\[
\frac{1}{|G|} \sum_{g \in G} \Gamma(g)|X\rangle \langle X|\Gamma^\dagger(g)|\psi\rangle = \frac{1}{|G|} A |\psi\rangle
\]

\[
= \sum_p \frac{1}{d_p} \|X_p\|^2 |\psi_p\rangle
\]

where \(|X_p\rangle \in V^{(p)}\) is the projection of \(|X\rangle\) on \(V^{(p)}\).

Which tells us the result of application of \(A\) on any vector \(|\psi\rangle\) in terms of \(|\psi\rangle\)'s projections on carrier spaces, given \(L = |X\rangle \langle X|\).

\[
A |\psi\rangle = |G| \sum_p \frac{1}{d_p} \|X_p\|^2 |\psi_p\rangle.
\]

Let us apply this to find the eigenstates of \(A\). Set \(|X^{(1)}\rangle = |x+, x+\rangle\) and \(|X^{(2)}\rangle = |x-, v_0\rangle\), the first states in the orbit. Hence \(L = |X^{(1)}\rangle \langle X^{(1)}| + |X^{(2)}\rangle \langle X^{(2)}|\) we have that

\[
A |\psi\rangle = +6(\|X_1^{(1)}\|^2 + \|X_1^{(2)}\|^2) |\psi_1\rangle \\
+6(\|X_2^{(1)}\|^2 + \|X_2^{(2)}\|^2) |\psi_2\rangle \\
+3(\|X_3^{(1)}\|^2 + \|X_3^{(2)}\|^2) |\psi_3\rangle
\]

Because \(\langle \psi_i | \psi_j \rangle = 0\), from this expression we see that eigenvectors of \(A\) are just vectors lying in the
invariant subspaces, and the eigenvalues are

\[
6(\|X^{(1)}\|_2^2 + \|X^{(2)}\|_2^2) = \frac{21}{4}
\]
\[
6(\|X^{(1)}_2\|_2^2 + \|X^{(2)}_2\|_2^2) = \frac{3}{4}
\]
\[
3(\|X^{(1)}_3\|_2^2 + \|X^{(2)}_3\|_2^2) = 3.
\]

Therefore, the largest eigenvalue is $21/4$ and the corresponding eigenvector is $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$.

The classical bound on the sum of the 12 probabilities is found as before. We assume that the probabilities can be derived from a joint distribution, $P(a_0, b_0; a_1, b_1; a_2, b_2)$ and calculate their sum in terms of the joint distribution. The largest coefficient multiplying a probability from the joint distribution gives the upper bound to the sum, and in this case it is 5.

\[
\frac{21}{4} > 5
\]
\[
q > c
\]

The quantum result violates the classical inequality.

### 3.5.4 Nonlocal Game Properties of a $Z_6$ System

In this section we invent a nonlocal game using the $D_3$ system and compare it to another nonlocal game invented using the cyclic group $Z_6$.

First, the $Z_6$-game. $Z_6$ has a single generator, $a^6 = e$.

The translation operator is $T|z\pm\rangle = |z\mp\rangle = \sigma_x$. $U^3 = T$ where

\[
U = |x+\rangle \langle x+| + e^{i\pi/3} |x-\rangle \langle x-|.
\]

$U$ defines two bases other than the computational basis \{0,1\}. \{|u_j\rangle = U^j |j\rangle \mid j = 0,1\}$ and \{|v_j\rangle = U^{2j} |j\rangle \mid j = 0,1\}.

These three bases are eigenstates of three observables $a_0, a_1$ and $a_2$.

To investigate joint events we use the $U \otimes U$ representation of $Z_6$ that acts on $\mathbb{C}^2 \otimes \mathbb{C}^2$. We choose two orbits, one that starts with $|0,0\rangle$

\[
\{|0,0\rangle, |u_0, u_0\rangle, |v_0, v_0\rangle, |1,1\rangle, |u_1, u_1\rangle, |v_1, v_1\rangle\}.
\]
and other one starts with $|0, u_0\rangle$

$$\{0, u_0\}, |u_0, v_0\rangle, |v_0, 1\rangle, |1, u_1\rangle, |u_1, v_1\rangle, |v_1, 0\rangle\}.$$  

The set of probabilities due to the first orbit are

$$\{p(a_0 = 0, b_0 = 0), p(a_1 = 0, b_1 = 0),$$

$$p(a_2 = 0, b_2 = 0), p(a_0 = 1, b_0 = 1),$$

$$p(a_1 = 1, b_1 = 1), p(a_2 = 1, b_2 = 1)\}$$

and due to the second one are

$$\{p(a_0 = 0, b_1 = 0), p(a_1 = 0, b_2 = 0),$$

$$p(a_2 = 0, b_0 = 1), p(a_0 = 1, b_1 = 1),$$

$$p(a_1 = 1, b_2 = 1), p(a_2 = 1, b_0 = 0)\}.$$  

Construct the $A$ operator

$$L = |0, 0\rangle \langle 0, 0| + |0, u_0\rangle \langle 0, u_0|$$

$$A = \sum_{j=0}^{5} (U \otimes U)^j L (U^\dagger \otimes U^\dagger)^j.$$  

We find that the eigenstate of $A$ with the largest eigenvalue lies in the space corresponding to the $e^{i\pi/3}$ eigenvalue of $U \otimes U$ which is two dimensional and spanned by $|x+\rangle |x-\rangle$ and $|x-\rangle |x+\rangle$. In this subspace $A$ reduces to a $2 \times 2$ matrix

$$\frac{1}{4} \begin{pmatrix} 1 & 1 + e^{i\pi/3} \\ 1 + e^{-i\pi/3} & 2 \end{pmatrix}.$$  

Hence the largest eigenvalue is $3 + (3/2)\sqrt{3}$ with the corresponding eigenvector

$$|\phi\rangle = \frac{1}{\sqrt{6}}[(1 + e^{i\pi/3}) |x, -x\rangle + \sqrt{3} |x, +x\rangle].$$
The classical bound for the sum of the joint probabilities is 5.

\[ 3 + (3/2)\sqrt{3} > 5 \]

\[ q > c \]

hence the quantum probabilities violate the classical bound.

If we express this $Z_6$ system as a nonlocal game, where each player is asked a 3-bit question $(s, t) \in \{0, 1, 2\} \times \{0, 1, 2\}$. But in this scheme not all pairs of $(s, t)$ can be asked. Either $s = t$ or $(s, t)$ must be $(0, 1)$, $(1, 2)$, or $(2, 0)$, so that six out of the nine possibilities are allowed, and they will be assumed to be equally probable.

For the case $(s, t) = (2, 0)$ they win if their bit values differ, and for the other any of the other allowed values of $(s, t)$ they win if their bit values are the same. Note that for each allowed value of $(s, t)$ there are two winning possibilities.

Classically their winning probability is $5/6$, and it can be achieved if Alice and Bob each always send the bit value 0.

In the quantum case, each probability has the same value $(2 + \sqrt{3})/8$ and there are two correct answers per question hence the winning probability is $(2 + \sqrt{3})/4$.

\[ \left(2 + \sqrt{3}\right)/4 \approx 0.93 > 5/6 \approx 0.83 \]

\[ w_q > w_c \]

so there is a quantum advantage.

### 3.5.5 Nonlocal Game Properties of the $D_3$ System

The structure of $D_3$ nonlocal game is different than the structure of $Z_3$ one. Again the 3-bit questions are $(s, t) \in \{0, 1, 2\} \times \{0, 1, 2\}$, but in this case all 9 questions are possible. Alice and Bob win if $s = t$ and they return the same bit value or if $s \neq t$ they return the bit values $(a, b)$ that are shown in the Table 7.

Note that in this case when $s = t$ there are two winning possibilities for $(a, b)$, but for $s \neq t$ there is only one. This is different from the previous game where for each allowed value of $(s, t)$ there were two winning possibilities.

The highest classical winning probability $w_c = 5/9$ can be achieved with the following strategy. If their deterministic responses are $a(s)$ for Alice and $b(t)$ for Bob, $a(0) = b(0) = 1$, $a(1) = b(1) = 0$, and $a(2) = b(2) = 0$. 
<table>
<thead>
<tr>
<th>(s, t)</th>
<th>(a, b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,1</td>
<td>0,1</td>
</tr>
<tr>
<td>1,0</td>
<td>1,0</td>
</tr>
<tr>
<td>0,2</td>
<td>1,0</td>
</tr>
<tr>
<td>2,0</td>
<td>0,1</td>
</tr>
<tr>
<td>1,2</td>
<td>0,1</td>
</tr>
<tr>
<td>2,1</td>
<td>1,0</td>
</tr>
</tbody>
</table>

Table 7: Winning conditions for the $D_3$ non-local game. Questions on the left column and correct answers are on the right column.

In the quantum case, when Alice and Bob share $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ the probabilities of different orbits have different values. The probabilities of the first orbit, which correspond to the case $s = t$, are $1/2$, which implies they always give to correct answer. Whereas the probabilities of the second orbit, which correspond to the case $s \neq t$, the probabilities are all equal to $3/8$. $w_q$ is $1/9$ times the sum of all probabilities which is $7/12$.

$$7/12 \approx 0.583 > 5/9 \approx 0.556$$

$$w_q > w_c$$

which is a Bell violation.

3.5.6 **A Bell Violation using $D_6$**

As an example of a larger non-abelian group we chose $D_6$. It has the same generators, but the presentation is different. $D_6 = \{r, s|r^6, s^2, (rs)^2\}$.

As a Dihedral group of even degree, $n = 6$, the group has six conjugacy classes as explained in Section 3.5.2. They are $C_e = \{e\}$, $C_r = \{r, r^5\}$, $C_{r^2} = \{r^2, r^4\}$, $C_{r^3} = \{r^3\}$, $C_s = \{s, r^2s, r^4s\}$, and $C_{rs} = \{rs, r^3s, r^5s\}$. From which we can deduce that it has 6 irreducible representations 4 of which $\Gamma^{(j)} j = 1, 2, 3, 4$ are 1-dimensional and 2 of which $\Gamma^{(j)} j = 5, 6$ are 2-dimensional.

We will make use of the following representation of $D_6$ on $\mathbb{C}^3$. The computational basis is $\{|j\rangle | j = 0, 1, 2\}$, and let us define another basis

$$|u_j\rangle = \frac{1}{\sqrt{3}} \sum_{k=0}^{2} e^{2\pi i j k/3} |k\rangle.$$
Corresponding to the group element \( r \), we choose

\[
\Gamma (r) = U
\]

\[
= |u_0\rangle\langle u_0| + e^{-i\pi/3}|u_1\rangle\langle u_1| + e^{i\pi/3}|u_2\rangle\langle u_2|
\]

\[
= \frac{1}{3}\begin{pmatrix}
1 - i\sqrt{3} & 1 & 1 + i\sqrt{3} \\
1 + i\sqrt{3} & 1 - i\sqrt{3} & 1 \\
1 & 1 + i\sqrt{3} & 1 - i\sqrt{3}
\end{pmatrix}
\]

and corresponding to \( s \) we choose

\[
\Gamma (s) = V
\]

\[
= |u_0\rangle\langle u_0| + i(|u_1\rangle\langle u_2| - |u_2\rangle\langle u_1|)
\]

\[
= \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0
\end{pmatrix}
\]

which is the same \( U \) that was used in Section 3.3.5. Note that it has the property \( T = U^2, \ T|j\rangle = |j + 1\rangle \), where the addition is in modulo 3. If we denote the representation generated by \( U \) and \( V \) by \( \Gamma \) then orthogonality for characters theorem gives the decomposition

\[
\Gamma = \Gamma^{(1)} \oplus \Gamma^{(5)}.
\]

Application of powers and products of the operators \( U \) and \( V \) to the computational basis yield three additional bases, \( \{|v_j\} = U|j\rangle|j = 0, 1, 2\}, \( \{|w_j\} = V|j\rangle|j = 0, 1, 2\}, \) and \( \{|x_j\} = UV|j\rangle|j = 0, 1, 2\} \).

We can now define four observables that take values in the set \( \{0, 1, 2\} \)

\[
a_0 = \sum_{j=1}^{2} j|j\rangle\langle j|
\]

\[
a_1 = \sum_{j=1}^{2} j|v_j\rangle\langle v_j|
\]

\[
a_2 = \sum_{j=1}^{2} j|w_j\rangle\langle w_j|
\]

\[
a_3 = \sum_{j=1}^{2} j|x_j\rangle\langle x_j|
\]
Figure 20: The quantum states generated by the group action. The bases are indicated with color. Computational basis is black (which also indicates the axes), $v$-basis is red and $w$-basis is green and $x$-basis is blue. (Note that this is not the Bloch sphere. The states are qutrits. But the components of state vectors are only real numbers.)

To generate joint events we use $\Gamma(g) \otimes \Gamma(g)$ representation where

$$\Gamma \otimes \Gamma = \Gamma^{(1)}(g) \oplus \Gamma^{(1)}(g) \oplus \Gamma^{(2)}(g) \oplus \Gamma^{(5)}(g) \oplus \Gamma^{(5)}(g) \oplus \Gamma^{(6)}(g).$$

The invariant subspaces we found for this decompositions are

- $\text{span} \{|u_0, u_0\rangle\} = \mathcal{V}^{(1)}_1$ for the first $\Gamma^{(1)}$
- $\text{span} \{(|u_1, u_2\rangle + |u_2, u_1\rangle)/\sqrt{2}\} = \mathcal{V}^{(1)}_2$ for the second $\Gamma^{(1)}$
- $\text{span} \{(|u_1, u_2\rangle - |u_2, u_1\rangle)/\sqrt{2}\} = \mathcal{V}^{(2)}$ for $\Gamma^{(2)}$
- $\text{span} \{|u_0, u_1\rangle, |u_0, u_2\rangle\} = \mathcal{V}^{(5)}_1$ for the first $\Gamma^{(5)}$
- $\text{span} \{|u_1, u_0\rangle, |u_2, u_0\rangle\} = \mathcal{V}^{(5)}_2$ for the second $\Gamma^{(5)}$
- $\text{span} \{|u_1, u_1\rangle, |u_2, u_2\rangle\} = \mathcal{V}^{(6)}$ for the $\Gamma^{(6)}$.

The two orbits we choose start with the states $(U^4 \otimes U^2 V|0, 0\rangle = |2, w_2\rangle$ and $(I \otimes U^3 V)|0, 0\rangle = |0, x_1\rangle$. Because $|D_6| = 12$ each orbit contains 12 states that give rise to 12 joint measurement probabilities. Those probabilities are given in Table 8.

$$L = |2, w_2\rangle \langle 2, w_2| + |0, x_1\rangle \langle 0, x_1|$$
<table>
<thead>
<tr>
<th></th>
<th>[2, w_2]</th>
<th>[0, x_1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(I)</td>
<td>(p(a_0 = 2, b_2 = 2))</td>
<td>(p(a_0 = 0, b_3 = 1))</td>
</tr>
<tr>
<td>(U)</td>
<td>(p(a_1 = 2, b_3 = 2))</td>
<td>(p(a_1 = 0, b_2 = 0))</td>
</tr>
<tr>
<td>(U^2)</td>
<td>(p(a_0 = 0, b_2 = 1))</td>
<td>(p(a_0 = 1, b_3 = 0))</td>
</tr>
<tr>
<td>(U^3)</td>
<td>(p(a_1 = 0, b_3 = 1))</td>
<td>(p(a_1 = 1, b_2 = 2))</td>
</tr>
<tr>
<td>(U^4)</td>
<td>(p(a_0 = 1, b_2 = 0))</td>
<td>(p(a_0 = 2, b_3 = 2))</td>
</tr>
<tr>
<td>(U^5)</td>
<td>(p(a_1 = 1, b_3 = 0))</td>
<td>(p(a_1 = 2, b_2 = 1))</td>
</tr>
<tr>
<td>(V)</td>
<td>(p(a_2 = 2, b_0 = 2))</td>
<td>(p(a_2 = 0, b_1 = 0))</td>
</tr>
<tr>
<td>(UV)</td>
<td>(p(a_3 = 2, b_1 = 2))</td>
<td>(p(a_3 = 0, b_0 = 1))</td>
</tr>
<tr>
<td>(U^2V)</td>
<td>(p(a_2 = 1, b_0 = 0))</td>
<td>(p(a_2 = 2, b_1 = 1))</td>
</tr>
<tr>
<td>(U^3V)</td>
<td>(p(a_3 = 1, b_1 = 0))</td>
<td>(p(a_3 = 2, b_0 = 2))</td>
</tr>
<tr>
<td>(U^4V)</td>
<td>(p(a_2 = 0, b_0 = 1))</td>
<td>(p(a_2 = 1, b_1 = 2))</td>
</tr>
<tr>
<td>(U^5V)</td>
<td>(p(a_3 = 0, b_1 = 0))</td>
<td>(p(a_3 = 1, b_0 = 0))</td>
</tr>
</tbody>
</table>

Table 8: Probabilities generated by orbits for \(D_6\). Starting states of the orbits are at the header row.

and

\[
A = \sum_{g \in D_6} (\Gamma (g) \otimes \Gamma (g)) L (\Gamma^\dagger (g) \otimes \Gamma^\dagger (g)).
\]

We found that the largest eigenvalue of \(A\) corresponds to an eigenvector that lies in the subspace spanned by the two vectors in \(V^{(1)} = V_1^{(1)} \cup V_2^{(1)}\). We say the eigenvector transforms as \(\Gamma^{(1)}\).

The rest of the eigenvalues are as follows. For \(\Gamma^{(2)}\), we find that the component of \([2, w_2]\) in this subspace is

\[
|X_2^{(1)}\rangle = \frac{-i}{6} (|u_1, u_2\rangle - |u_2, u_1\rangle)
\]

while the component of \([0, x_1]\) is just

\[
|X_2^{(2)}\rangle = -|X_2^{(1)}\rangle.
\]

The eigenvalue corresponding to the \(\Gamma^{(2)}\) space \(V^{(2)}\) is, then, 4/3.

The components of \([2, w_2]\) and \([0, x_1]\) in the \(\Gamma^{(6)}\) subspace are

\[
|X_6^{(1)}\rangle = \frac{i}{3}(-|u_1, u_1\rangle + |u_2, u_2\rangle) \\
|X_6^{(2)}\rangle = -\frac{1}{3\sqrt{3}}[(1 - e^{-2\pi i/3})|u_1, u_1\rangle + (1 - e^{2\pi i/3})|u_2, u_2\rangle]
\]

respectively. This gives 8/3 as the eigenvalue corresponding to \(\Gamma^{(6)}\), and this eigenvalue is two-fold degenerate.

The \(\Gamma^{(5)}\) subspace \(V^{(5)} = V_1^{(5)} \cup V_2^{(5)}\) is more complicated. It is four dimensional and consists of two copies of the \(\Gamma^{(5)}\) irreducible representation. We first note that because \(|u_0\rangle\) is invariant under the actions of \(U\) and \(V\), the states \(|u_0, u_1\rangle\) and \(|u_1, u_0\rangle\) transform in the same way, and the states \(|u_0, u_2\rangle\) and \(|u_2, u_0\rangle\) transform in the same way.

Now suppose that \(|X_5\rangle\) is a vector in \(V^{(5)}\). Setting \(|\alpha_j\rangle = |u_0, u_j\rangle\) and \(|\beta_j\rangle = |u_j, u_0\rangle\), for \(j = 1, 2\), we
find from orthogonality relation, that

\[
\sum_{g \in D_6} \Gamma(g) |X_5\rangle \langle X_5| \Gamma^\dagger(g) = 6 \begin{pmatrix}
\|X_{5\alpha}\|^2 & 0 & z & 0 \\
0 & \|X_{5\alpha}\|^2 & 0 & z \\
z^* & 0 & \|X_{5\beta}\|^2 & 0 \\
0 & z^* & 0 & \|X_{5\beta}\|^2
\end{pmatrix}
\]

where the matrix is in the \{\alpha_1, \alpha_2, \beta_1, \beta_2\} basis, and \(\|X_{5\alpha}\|^2 = \sum_{j=1}^{2} |\langle X_5|\alpha_j\rangle|^2\) and \(\|X_{5\beta}\|^2 = \sum_{j=1}^{2} |\langle X_5|\beta_j\rangle|^2\)
and \(z = \sum_{j=1}^{2} \langle X_5|\beta_j\rangle \langle \alpha_j|X_5\rangle\).

The component of \(|2, w_2\rangle\) transforming as \(\Gamma(5)\) is

\[
|X_5^{(1)}\rangle = \frac{1}{3\sqrt{3}}[(1 - e^{-2\pi i/3})|u_0, u_1\rangle + (1 - e^{2\pi i/3})|u_0, u_2\rangle + \frac{1}{3}(e^{-2\pi i/3}|u_1, u_0\rangle + e^{2\pi i/3}|u_2, u_0\rangle)]
\]

and the component of \(|0, x_1\rangle\) transforming as \(\Gamma(5)\) is

\[
|X_5^{(2)}\rangle = -\frac{1}{3\sqrt{3}}[(1 - e^{-2\pi i/3})|u_0, u_1\rangle + (1 - e^{2\pi i/3})|u_0, u_2\rangle + \frac{1}{3}(|u_1, u_0\rangle + |u_2, u_0\rangle)].
\]

For both \(|X_5^{(1)}\rangle\) and \(|X_5^{(2)}\rangle\) we find \(\|X_{5\alpha}\|^2 = \|X_{5\beta}\|^2 = 2/9\) and \(z = -1/(3\sqrt{3})\). Putting these together, we find that eigenvalues of \(A\) in the \(\Gamma(5)\) subspace are \((4/3)(2 \pm \sqrt{3})\) each of which is two-fold degenerate.

If we go back to the invariant subspace of \(\Gamma(1)\), which corresponds to the largest eigenvalue of \(A\), the components of both \(|2, w_2\rangle\) and \(|0, x_1\rangle\) that lie in \(V^{(1)}\) are the same and are given by

\[
|X_1\rangle = \frac{1}{3}|u_0, u_0\rangle - \frac{1}{2\sqrt{3}}(|u_1, u_2\rangle + |u_2, u_1\rangle).
\]

Because it transforms as \(\Gamma(1)\), the trivial representation that sends all group elements to 1, this vector is invariant under the actions of \(U\) and \(V\), and this implies that in the \(\Gamma(1)\) space, \(A\) is just \(2(12)|X_1\rangle \langle X_1|\). Therefore, the two eigenvectors of \(A\) in this subspace are the vector orthogonal to \(|X_1\rangle\), which has an eigenvalue
of 0, and a normalized version of $|X_1\rangle$, which is $|\phi\rangle = 3(\sqrt{2/5})|X_1\rangle$, whose eigenvalue is $2(12)\|X_1\|^2 = 20/3$.

$$\frac{20}{3} \approx 6.67 > 6$$

$$q > c$$

So, when Alice an Bob share the state $|\phi\rangle$ the sum of probabilities violates the classical bound, hence the sum of 24 probabilities gives us a Bell inequality.
4 Conclusion

We have shown how the technique of Jordan bases for two subspaces can be used to find maximum quantum violations of a class of Bell inequalities. We do not need to make any assumptions on the dimensionality of the Hilbert space, and the technique gives us the observables and the states that produce the maximum violations.

We also have shown how certain group actions can be used to generate Bell inequalities. In particular, we provided an example of a three-party Bell inequality using an Abelian group, and two examples of two-party inequalities via non-Abelian groups.

The orbits of the group action are used to generate events, the sum of their probabilities is the Bell expression appearing in the Bell inequality.

There are a number of areas in which the research presented here could be extended. The choice of the orbits that led to the Bell inequalities was done by using a random search. It would be useful to have a criterion for choosing them. This would also allow us to gain a better understanding of how the structures of Bell inequalities are related to the underlying groups.

The Bell inequalities depend on both the group and the choice of orbits, and at the moment we do not have a good way of disentangling these two effects. A better understanding of how to choose the orbits would, we hope, lead to a better idea of the relation between group theory and Bell’s theorem.
5 Appendix

5.1 Computer Calculations

For group theoretical and abstract linear algebraic calculations that are hard to do by hand we used the Python programming language with its libraries Numpy, Sympy and the SAGE computer algebra and symbolic manipulation system based on Python http://www.sagemath.org/.

Usually, we invented a scheme for small parameters (small numbers in Bell scenarios or small groups) by hand and then implemented the scheme in the computer and do searches with bigger parameters that give Bell violations.

5.1.1 Calculations for the Jordan Angle Method with $N$ parties

We used the “quantum” module of “sympy” library in the SAGE environment. sympy is a symbolic manipulation and calculation module extension for Python programming language. Quantum module has programming objects for generic kets. It can work on abstract kets and operators without defining their dimensionality or assigning numbers to its components and entries.

$(2,2,2)$ Case Demonstration of the idea with the simplest case of $(2,2,2)$. We have 2 parties $a$ and $b$. There are two observables that we measure on each party: $a_1$, $a_2$ and $b_1$, $b_2$. They have two outcomes of which values are not important and that we will label as + and −. The corresponding eigenvectors belonging their individual Hilbert spaces are $\{|a_1\pm\rangle, |a_2\pm\rangle\} \in \mathcal{H}^A$, $\{|b_1\pm\rangle, |b_2\pm\rangle\} \in \mathcal{H}^B$. The eigenvectors have these relations between them

$$\langle a_m \pm | b_n \pm \rangle = 0$$

$$\langle a_m = j | a_m = k \rangle = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$$

$$\langle a_1 + | a_2 + \rangle = \cos \theta_a \equiv c_a$$

$$\langle b_1 + | b_2 + \rangle = \cos \theta_b \equiv c_b$$

where $\theta_a$ and $\theta_b$ are the Jordan Principle Angles between the subspaces defined by the positive eigenvector of the observables on two parties. These numbers uniquely define the relation between subspaces (up to an isometry).
We will choose a non-orthogonal basis in $H = H^A \otimes H^B$, $\{|a_1 + b_1+\rangle, |a_1 + b_2+\rangle, |a_2 + b_1+\rangle, |a_2 + b_2+\rangle\} = \{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}$. We want to express the Bell operator, $B$, in this basis. Because of the non-orthogonality $B_{ij} \neq \langle i | B | j \rangle$. $B_{ij}$ is found by applying $B$ on $|j\rangle$ and then picking the coefficient of the term $|i\rangle$. Such as

$$B |j\rangle = B_{0j} |0\rangle + B_{1j} |1\rangle + B_{2j} |2\rangle + B_{2j} |3\rangle$$

Sympy/SAGE code

Import these programming libraries to include the ability of manipulation of abstract kets and operators:

```python
from sympy import *
from sympy.physics.quantum import *
```

Create the abstract kets. ($u_1 = |a_1+\rangle$, $v_2 = |b_2+\rangle$ etc.)

```python
u1 = Ket('u1')
u2 = Ket('u2')
v1 = Ket('v1')
v2 = Ket('v2')
```

Define the non-orthogonal basis $e = \{|a_1 + b_1+\rangle, |a_1 + b_2+\rangle, |a_2 + b_1+\rangle, |a_2 + b_2+\rangle\} = \{|0\rangle, |1\rangle, |2\rangle, |3\rangle\}$ and $ed = \langle e |$. TensorProduct($u_1,v_1$) = $|u_1\rangle \otimes |v_1\rangle = |a_1 + b_1+\rangle$ etc.

```python
e = [TensorProduct(u1,v1), TensorProduct(u1,v2),
    TensorProduct(u2,v1), TensorProduct(u2,v2)]
ed = [TensorProduct(u1.dual,v1.dual), TensorProduct(u1.dual,v2.dual),
    TensorProduct(u2.dual,v1.dual), TensorProduct(u2.dual,v2.dual)]
```

Create projection operators onto eigensubspaces. $qa1 = Q (a_1+) = |a_1+\rangle \langle a_1+|$ etc. acting on $H^A$

```python
qa1 = u1*u1.dual
qa2 = u2*u2.dual
qb1 = v1*v1.dual
qb2 = v2*v2.dual
```

Express the Bell operator $B$ acting on $H$, only using these operators (and identity).

$$\text{LHS} = Q (a_1 + b_1+) - Q (a_2 + b_2+)
\text{RHS} = Q (a_2 - b_1+) + Q (a_1 + b_2-)$$
\[ Q(a_1 + b_1^+) = Q(a_1^+) \otimes Q(b_1^+) \]

\[ Q(a_2 - b_1^+) = Q(a_2^-) \otimes Q(b_1^+) = [I - Q(a_2^+)] \otimes Q(b_1^+) = I \otimes Q(b_1^+) - Q(a_2^+) \otimes Q(b_1^+) \]

Similarly

\[ Q(a_1 + b_2^-) = Q(a_1^+) \otimes I - Q(a_1^+) \otimes Q(b_2^+) \]

Therefore

\[ B = \text{LHS} - \text{RHS} \]
\[ = Q(a_1 + b_1^+) - Q(a_2 + b_2^+) - Q(a_2 - b_1^+) - Q(a_1 + b_2^-) \]
\[ = +Q(a_1^+) \otimes Q(b_1^+) - Q(a_2^+) \otimes Q(b_2^+) - I \otimes Q(b_1^+) \]
\[ = +Q(a_2^+) \otimes Q(b_1^+) - Q(a_1^+) \otimes I + Q(a_1^+) \otimes Q(b_2^+) \]

\[ \langle B \rangle = p(a_1, b_1) + p(a_1, b_2) + p(a_2, b_1) - p(a_2, b_2) - p(a_1) - p(b_1) \]

which is classically always greater than zero.

bopterms includes all terms of the \( B \) and boptermssigns includes their signs.

```
1 bopterms = [TensorProduct(qa1,qb1), TensorProduct(qa1,qb2),
2 TensorProduct(qa2,qb1), TensorProduct(qa2,qb2),
3 TensorProduct(qa2,1), TensorProduct(1,qb1)]
4 boptermssigns = [1,-1,1,1,-1,-1]
```

Define the principle angles between the subspaces belonging to the same party. \( a = c_a, b = c_b \). Assign cosines of angles to brackets. “u1.dual*u1: 1” means \( \langle a_1^+ \mid a_1^+ \rangle = 1 \) and “u1.dual*u2: a” means \( \langle a_2^+ \mid a_1^+ \rangle = c_a = \cos \theta_a \).

```
1 [a,b] = var('a\_b')
2 rules = {u1.dual*u1: 1, u1.dual*u2: a, u2.dual*u1: a,
3 u2.dual*u2: 1, v1.dual*v1: 1, v1.dual*v2: b,
4 v2.dual*v1: b, v2.dual*v2: 1}
```
The function that applies $B$ on any vector. Apply each term on the vector, then multiply with the sign and add up all results.

```python
def applyB(vec):
    sum = 0
    for i in range(len(bopterms)):
        sum = sum +
        boptermssigns[i] * qapply(tensor_product_simp(bopterms[i] * vec)).subs(rules)
    return sum
```

Print the results of $B|i\rangle$ for $i = 0, 1, 2, 3$

```python
print applyB(e[0])
print applyB(e[1])
print applyB(e[2])
print applyB(e[3])
```

which gives

```plaintext
a*b*|u2>x|v2> - b*|u1>x|v2>
2 a*b*|u2>x|v1> - |u1>x|v2>
3 -a*b*|u1>x|v2> + a*|u1>x|v1> + b*|u2>x|v2> - |u2>x|v1>
4 a*b*|u1>x|v1> - a*|u1>x|v2>
```

which means

\[
B|0\rangle = -b|1\rangle + ab|3\rangle \\
B|1\rangle = -|1\rangle + ab|2\rangle \\
B|2\rangle = a|0\rangle - ab|1\rangle - |2\rangle + b|3\rangle \\
B|3\rangle = ab|0\rangle - a|1\rangle
\]

code that gets the matrix elements of $B$. where “expr.coeff(el)” gives the coefficient of the element el in expression expr.

```python
mat = Matrix(4, 4, lambda i, j: 0)
for i in range(4):
    for j in range(4):
```
val = applyB(e[j]).coeff(e[i])

if val == None:
    mat[i, j] = 0
else:
    mat[i, j] = val

which gives the matrix

\[
\begin{bmatrix}
0 & 0 & a & a*b \\
-b & -1 & -a*b & -a \\
0 & a*b & -1 & 0 \\
a*b & 0 & b & 0 \\
\end{bmatrix}
\]

(3, 2, 2) Case In order to add the third part manually two new eigenvectors are defined \{\(|c_1+\rangle, |c_2+\rangle\}\).

The non-orthogonal basis on which we represent the bell operator is made of tensor products of three eigenvectors.
\[
|a_m+\rangle \otimes |b_n+\rangle \otimes |c_l+\rangle = |a_m + b_n + c_l+\rangle \equiv |mnl\rangle
\]

\(B\) is determined by the \(N = 3\) inequality.

\[
\text{LHS} = Q(a_1 + b_1 + c_1+) - Q(a_2 + b_2 + c_2+) \\
\text{RHS} = Q(a_2 - b_1 + c_1+) + Q(a_1 + b_2 - c_1+) + Q(a_1 + b_1 + c_2-)
\]

After everything is expressed using \(Q(a_m+), Q(b_n+), Q(c_l+)\) we have \(B\) in the intended form.

\[
B = Q(a_1) \otimes Q(b_1) \otimes Q(c_1) - Q(a_2) \otimes Q(b_2) \otimes Q(c_2) \\
- I \otimes Q(b_1) \otimes Q(c_1) + Q(a_1) \otimes Q(b_1) \otimes Q(c_1) \\
- Q(a_1) \otimes I \otimes Q(c_1) + Q(a_1) \otimes Q(b_1) \otimes Q(c_1) \\
- Q(a_1) \otimes Q(b_1) \otimes I + Q(a_1) \otimes Q(b_1) \otimes Q(c_2)
\]

and its representation in \(...\) basis is

\[
\begin{bmatrix}
-2, & -c, & -b, & 0, & -a, & 0, & 0, & a*b*c \\
c, & 0, & b*c, & 0, & a*c, & 0, & a*b*c, & 0 \\
b, & b*c, & 0, & 0, & a*b, & a*b*c, & 0, & 0 \\
\end{bmatrix}
\]
\( (N, 2, 2) \) **General Case**  Then we have a code for the general case. First set the number of parties and number of outcomes \( np = N \), \( no = K \). (number of observables per party is still fixed to the value 2.)

```
np = 3
no = 2
```

Create the eigenvectors using the labels \( |u_{i,j}\rangle \) where \( i \) indicates the index of the party, and \( j \) the index of the outcome. They all correspond to only one of the outcomes, say +.

```
u = [[Ket("u"+str(i)+"","+str(j)") for j in range(no)] for i in range(np)]
```

These are the vectors for \( N = 3 \).

```
[[|u0,0>, |u0,1>], [|u1,0>, |u1,1>], [|u2,0>, |u2,1>]]
```

Now it is time to create the non-orthogonal basis in which we will represent \( B \). Basically we need all combinations with a tensor product of three eigenvectors, one from each party. \( \{ |u_{0,0}u_{1,0}u_{2,0}\rangle, |u_{0,0}u_{1,1}u_{2,0}\rangle, |u_{0,0}u_{1,0}u_{2,1}\rangle, \ldots, |u_{0,1}u_{1,1}u_{2,1}\rangle \} \)

The order of the first indices is always the same 0, 1, 2 indicating the parties. The second indices go from 0, 0, 0 to 1, 1, 1.

“bin(i)” gives the binary representation of the integer \( i \). “bin(2)” is “0b10”. To get rid off the first two characters use “bin(i)[2::]”. Then the left side must be padded with zeros until the number of digits is equal to \( N \). The number of zeros needed is \( N \) minus the number of digits of the binary representation. “len(x)” gives the number of characters in \( x \). “len(bin(i)[2::])” gives the number of digits of the binary representation. “np-len(bin(i)[2::])” is the number of zeros to be padded. “0*n” creates \( n \) zeros. “x+y” concatenate two strings.

```
binaries = [ '0'*(np-len(bin(i)[2::]))+bin(i)[2::] for i in range(2*np)]
```

The result for \( N = 3 \)

```
[ '000', '001', '010', '011', '100', '101', '110', '111']
```
Use this array of all combinations of 3 binary digits in creating the non-orthogonal basis. Use each item as the second indices.

\[
\text{binaries\_int} = \left[ \left[ \text{int(c)} \right] \text{for \ for} \ \text{c in binaries[i]} \right] \text{for \ i in range(2**np)}
\]

\[
e = [\text{TensorProduct (}*[u[i][v[i]] \text{for \ i in range(np)]}) \text{for \ v in binaries\_int]}
\]

The result is

\[
[| u0,0 > x | u1,0 > x | u2,0 > , \ | u0,0 > x | u1,0 > x | u2,1 > ,
\]

\[
[| u0,1 > x | u1,0 > x | u2,0 > , \ | u0,1 > x | u1,0 > x | u2,1 > ,
\]

\[
[| u0,1 > x | u1,1 > x | u2,0 > , \ | u0,1 > x | u1,1 > x | u2,1 > ]
\]

Create the projection operators projecting onto the space spanned by each positive eigenvector of observables.

\[
\{\{u_{0,0}\} \langle u_{0,0} | u_{0,1}\} \langle u_{0,1} | \}, \{\{u_{1,0}\} \langle u_{1,0} | u_{1,1}\} \langle u_{1,1} | \}, \{\{u_{2,0}\} \langle u_{2,0} | u_{2,1}\} \langle u_{2,1} | \}
\]

\[
o = [\left[ | u[i][j] > x | u[i][j] > x ... | u[i][j] > x \right] \text{dual for \ j in range(no)] \ for \ i in range(np)]}
\]

Create the LHS:

\[
| u_{0,0}\} \langle u_{0,0} | \otimes | u_{1,0}\} \langle u_{1,0} | \otimes | u_{2,0}\} \langle u_{2,0} | - | u_{0,1}\} \langle u_{0,1} | \otimes | u_{1,1}\} \langle u_{1,1} | \otimes | u_{2,1}\} \langle u_{2,1} |
\]

“[o[i][0] for i in range(np)]” is an array of projection operators onto the first observable’s positive eigenspace of each party.

“function(*array)” gives an array as arguments of a function. “TensorProduct(*[o[i][0] for i in range(np)])” gives the tensor product of the projection operators in the array.

\[
[\text{TensorProduct (}*[o[i][0] \text{for \ i in range(np)]})],
\]

\[
\text{TensorProduct (}*[o[i][1] \text{for \ i in range(np)]})]
\]

and their signs are \{+,−\}.

Create RHS: RHS is made of \(N\) terms. All of the arguments of each term include projection onto positive value of the first observable, except one of them is onto negative (actually non-positive) value of the second observable.

\[
Q(a_1+) \otimes Q(b_1+) \otimes ... \otimes Q(n_2−) \otimes ... \otimes Q(z_1+)
\]
Instead of $Q(n_2^-)$ we want to use $I - Q(n_2^+)$. Hence our term becomes

$$Q(a_1+) \otimes Q(b_1+) \otimes \cdots \otimes I \otimes \cdots \otimes Q(z_1^+) - Q(a_1+) \otimes Q(b_1+) \otimes \cdots \otimes Q(n_2^+) \otimes \cdots \otimes Q(z_1^+)$$

We need to prepare $2N$ terms in pairs. Positive term has identity instead of a measurement on the $n^{th}$ party and a negative term where second measurement is done on the $n^{th}$ party.

Following code prepares the terms with second measurements. It prepares $N$ terms, where at $i^{th}$ term the $i^{th}$ multiplicand of the tensor product is $Q(i_2^+)=\text{q}[i][1]$ and the rest is $Q(j_1^+)=\text{q}[j][0]$.

```python
for i in range(np):
    lst = []
    for j in range(np):
        if i == j:
            lst.append(q[j][1])
        else:
            lst.append(q[j][0])
    bopterm.append(TensorProduct(*lst))
    boptermssigns.append(1)
```

And following code prepares terms with identity in it. It prepares $N$ terms, where at $i^{th}$ term the $i^{th}$ multiplicand of the tensor product is $I=1$ and the rest is $Q(j_1^+)=\text{q}[j][0]$.

```python
for i in range(np):
    lst = []
    for j in range(np):
        if i == j:
            lst.append(1)
        else:
            lst.append(q[j][0])
    bopterm.append(TensorProduct(*lst))
    boptermssigns.append(-1)
```

We need $N$ jordan angles

```python
c = [var('c'+str(i)) for i in range(np)]
```

which gives:
The relationship between eigensubspaces are defined via Jordan angles. \( \langle u_{i,0} \mid u_{i,0} \rangle = 1 \leftrightarrow \langle (u[i][0].\text{dual}*u[i][0],1) \rangle \). \( \langle u_{i,1} \mid u_{i,1} \rangle = 1 \leftrightarrow \langle (u[i][1].\text{dual}*u[i][1],1) \rangle \). \( \langle u_{i,0} \mid u_{i,1} \rangle = c_i \leftrightarrow \langle (u[i][0].\text{dual}*u[i][1],c[i]) \rangle \\

### Rules

1. \( \text{rules} = [\langle u[i][0].\text{dual}*u[i][0],1 \rangle \text{ for } i \in \text{range}(np)] \)
2. \( + [\langle u[i][1].\text{dual}*u[i][1],1 \rangle \text{ for } i \in \text{range}(np)] \)
3. \( + [\langle u[i][0].\text{dual}*u[i][1],c[i] \rangle \text{ for } i \in \text{range}(np)] \)
4. \( + [\langle u[i][1].\text{dual}*u[i][0],c[i] \rangle \text{ for } i \in \text{range}(np)] \)
5. \( \text{rules} = \text{dict}(\text{rules}) \)

We have similar functions to apply \( B \) on the basis elements.

\( B \) for \( N = 3 \)

### Matrices

1. \[
\begin{bmatrix}
-2, & -c_2, & -c_1, & 0, & -c_0, & 0, & 0, & c_0*c_1*c_2
\end{bmatrix}
\]
2. \[
\begin{bmatrix}
c_2, & 0, & c_1*c_2, & 0, & c_0*c_2, & 0, & c_0*c_1*c_2, & 0
\end{bmatrix}
\]
3. \[
\begin{bmatrix}
c_1, & c_1*c_2, & 0, & 0, & c_0*c_1, & c_0*c_1*c_2, & 0, & 0
\end{bmatrix}
\]
4. \[
\begin{bmatrix}
0, & 0, & 0, & 0, & 0, & 0, & 0, & 0
\end{bmatrix}
\]
5. \[
\begin{bmatrix}
c_0, & c_0*c_2, & c_0*c_1, & c_0*c_1*c_2, & 0, & 0, & 0, & 0
\end{bmatrix}
\]
6. \[
\begin{bmatrix}
0, & 0, & 0, & 0, & 0, & 0, & 0, & 0
\end{bmatrix}
\]
7. \[
\begin{bmatrix}
0, & 0, & 0, & 0, & 0, & 0, & 0, & 0
\end{bmatrix}
\]
8. \[
\begin{bmatrix}
-c_0*c_1*c_2, & -c_0*c_1, & -c_0*c_2, & -c_0, & -c_1*c_2, & -c_1, & -c_2, & -1
\end{bmatrix}
\]

We have matrices of \( N = 4, 5, 6, 7 \) (The last one is \( 128 \times 128 \) dimensional. \( 128 = 2^7 \))

### 5.1.2 Group Theoretical Calculations in Computer Algebra System SAGE

SAGE comes with the GAP (Groups, Algorithms, Programming) which is a system for computational discrete algebra [http://www.gap-system.org/](http://www.gap-system.org/). We used GAP to generate groups and do calculations with them.

Here is how groups are generated in GAP.

1. \( \text{S3} = \text{SymmetricGroup}(3) \)
2. \( \text{D6} = \text{DihedralGroup}(6) \)
3. \( \text{Z8} = \text{CyclicPermutationGroup}(8) \)

The outputs of these commands are

1. Symmetric group of order 3! as a permutation group
Dihedral group of order 12 as a permutation group
Cyclic group of order 8 as a permutation group

Several methods of a group object are

- `G.list()`
- `G.order()`
- `G.subgroups()`
- `G.conjugacy_classes()`
- `G.gens()`

These lines return the elements of the group, the group order, the list of all subgroups, the list of all conjugacy classes, a list of generators.

If two variables are elements of the group, then GAP can apply group operation on them. For example

```gap
1 g1, g2 = G.list()[2]
2 g2 in G
3 g1 * g2
4 g2.inverse()
5 (g1**n)*(g2**m)
```

Line 1 assigns the first two group elements to \( g_1 \) and \( g_2 \). Line 2 returns true because \( g_2 \in G \). Line 3 is the group operation on the pair. Line 4 returns the \( g_2^{-1} \). Line 5 is \( g_1^n * g_2^m \) where \( n \) and \( m \) can be any integer.

The Cayley tables are calculated using the `cayley_table()` method.

```gap
1 G.cayley_table()
2 G.cayley_table(names='digits')
3 G.cayley_table(names='elements')
4 G.cayley_table(names=['e', 'x', 'y', 'z', 'u', 'v'])
```

where first one labels the group elements in alphabetic order, second one with digits, third one with permutation operators in disjoint-cyclic notation and the last one with arbitrary labels.

```gap
1 show(G.cayley_graph(), color_by_label=True, edge_labels=True)
```

plots the Cayley graph of the group, puts the permutation labels of the group elements at the vertices.

```gap
1 [list(c) for c in G.conjugacy_classes()]
```
Lists the elements of each conjugacy class.

1 \texttt{G.character_table()}

Calculates the character table of the group.

1 \texttt{[g.matrix() for g in G]}

shows the permutation matrix representation of the group elements.

GAP also comes with presentations of some classes of groups. For example the outputs of these commands

1 \texttt{groups.presentation.Alternating(5)}
2 \texttt{groups.presentation.Symmetric(4)}

are

1 \texttt{Finitely presented group <a,b | a^{-3}, b^{-5}, (a^{-1}*b^{-2})^{-2}, (b*a^{-1})^{-3}>}
2 \texttt{Finitely presented group <a,b | b^{-2}, a^{-4}, (a*b)^{-3}>}

GAP also calculates irreducible representations of a group.

1 \texttt{libgap.IrreducibleRepresentations( G )}

returns the list of homomorphisms from the group elements to representation matrices

1 \texttt{irreps = libgap.IrreducibleRepresentations( G )}
2 \texttt{hom = irreps[P]}
3 \texttt{img = libgap.Image(hom, g)}
4 \texttt{gammaP = matrix(img)}
5 \texttt{gammaTildeP = S*gammaP*S.inverse()}

Here line 1 calculates the irreducible representations of the group. Line 2 chooses the \(p\)'th representation (note that indices start from 0, not 1), which is a homomorphism from the group to matrices. Line 3 calculates the matrix that corresponds to the group element \(g\). Line 4 that returns \(\Gamma^{(p)}(g)\) is necessary to convert the matrix in GAP format to a matrix of SAGE format. Line 5 is a similarity transformation \(\hat{\Gamma}^{(p)} = S\Gamma^{(p)}(g)S^{-1}\) to calculate equivalent representations.

5.1.3 Calculation of Classical Bounds

To calculate classical bounds we need a system of several parts.
First part calculates joint probabilities of $N$ events that occur at $N$ parties in terms of the probability distribution of all events. For $(N, M, K) = (2, 2, 2)$ this means

$$p(a_m = i, b_n = j) = \sum_{o_x}^{2} \sum_{o_y}^{2} P(a_1 = o_{a1}, a_2 = o_{a2}, b_1 = o_{b1}, b_2 = o_{b2})$$

where $x \neq m$, $y \neq n$.

As a specific example

$$p(a_2 = 0, b_1 = 1) = \sum_{k} \sum_{t} P(a_1 = k, a_2 = 0, b_1 = 1, b_2 = t)$$

$$= \sum_{k} \sum_{l} \sum_{s} \sum_{t} \beta_{klst} P(a_1 = k, a_2 = l, b_1 = s, b_2 = t)$$

We can immediately say that $\beta_{klst} = 0$ whenever $l \neq 0$ or $s \neq 1$, in other words

$$\beta_{klst} = \begin{cases} 1 & l = 0 \text{ and } s = 1 \\ 0 & \text{otherwise} \end{cases}$$

If we define

$$P_{klst} = P(a_1 = k, a_2 = l, b_1 = s, b_2 = t)$$

Any joint probability can be expressed as a convex linear combination of $p_{klst}$

$$p(a_m = i, b_n = j) = \sum_{klst} \beta_{klst} P_{klst}$$

$$= \beta \cdot P$$

which can be expressed as a dot product.

So, first part of the system should calculate $\beta$ for a given joint measurement outcome probability for any $(N, M, K)$ scenario.
$$p(a_m = o_m, b_m = o_m, \ldots, z_m = o_m) = \sum_{o_{a1} \cdots o_{zM}} \beta_{o_{a1} \cdots o_{zM}} P \left( \begin{array}{c} a_1 = o_{a1}, a_2 = o_{a2}, \ldots, a_M = o_{zM}, \\ b_1 = o_{b1}, b_2 = o_{b2}, \ldots, b_M = o_{bM}, \\ \ldots \\ z_1 = o_{z1}, z_2 = o_{z2}, \ldots, z_M = o_{zM}, \end{array} \right)$$

$$= \sum_{o_{a1} \cdots o_{zM}} \beta_{o_{a1} \cdots o_{zM}} P(o_{a1} \cdots o_{zM})$$

$$= \beta \cdot P$$

Second part calculates the joint measurement outcomes from the sequence of quantum states. It should interpret the inner product relations of the quantum states and decide which outcome of which observable that they correspond. $|\psi\rangle \rightarrow p(a_m = i, b_m = j)$.

Last part calculates a Bell expression in terms of $\beta$. Say, a Bell expression is a linear combination of probabilities of joint measurements, $P = \{p_j\}$. $B = \sum_j c_j p_j$. In our Bell expressions $c_j = 1$ always.

For each $p_j$ there is a corresponding $\beta_j$. And $\beta$ for the Bell expression has the same linear combination

$$\beta = \sum c_j \beta_j$$

The classical bound is the biggest component of $\beta$.

### 5.1.4 Search for Bell violations

Here we provide more detail about how the random search to determine the orbits for the group $D_3$ was performed. The two orbits that yield a Bell inequality were found by a random search in the space of all possible orbit pairs.

First, using SAGE the $D_3$ group is generated and its elements $g \in D_3$ are calculated. Then the group generators $r$ and $s$ are associated with the corresponding representation matrices, $U = \Gamma(r), V = \Gamma(s)$.

We know how the rest of the group elements are generated from the generators $\{g_i| i = 0, 1, \ldots 5\} = \{e, r, r^2, s, rs, r^2s\}$. The associated representation matrices are calculated accordingly, $\{\Gamma(g_i)| i = 0, 1, \ldots 5\} = \{I, U, U^2, V, UV, U^2V\}$.

To associate the representation matrices with quantum measurement outcome states, the matrices are applied to a chosen initial state, which in our case was $|+x\rangle$, giving $|\psi_i\rangle = \Gamma_i|+x\rangle$.

In the code, then, the orthogonality relations among these states are analyzed. A table $T_{ij} = |\langle \psi_i | \psi_j \rangle|$ of the absolute values of inner products is calculated.
From the table these states are classified into different orthonormal bases, with each basis corresponding to the different possible eigenstates of a single observable. States for which the inner products are 0 or 1 are in the same basis.

In this way the Bell scenario for the number of measurements and outcomes is determined. The choice of initial state is essential to be able to get useful orthonormal bases. To be specific, each state is associated with an event $E$, namely an observable and its outcome, $|\psi_i\rangle \leftrightarrow a_{m(g_i)} = o(g_i)$, where $m$ is the choice of observable, and $o$ is the outcome. For our choice of $U$, $V$ and the initial state, $|+\rangle$, the 6 states $|\psi_i\rangle$ belong to 3 two dimensional orthonormal bases. $\{E_i| i = 0, \ldots, 5 \} = \{a_0 = 0, a_1 = 0, a_2 = 0, a_0 = 1, a_1 = 1, a_2 = 1 \}$.

We have two parties, and we want to see whether two orbits are sufficient. For each orbit we need two group elements, $g_\mu$ and $g_\nu$, to set the initial joint state $|\Psi_{\mu,\nu}\rangle = \Gamma(g_\mu)|+\rangle \otimes \Gamma(g_\nu)|+\rangle$. Then, the orbit will give us the $A$ operator $A_{\mu,\nu} = \sum_i (\Gamma(g_i) \otimes \Gamma(g_i)) |\Psi_{\mu,\nu}\rangle \langle \Psi_{\mu,\nu}| (\Gamma(g_i)^\dagger \otimes \Gamma^\dagger(g_i))$. The $A$ corresponding to both orbits is $A = A_{\mu_1,\nu_1} + A_{\mu_2,\nu_2}$. The choice of $\mu_1, \nu_1, \mu_2, \nu_2$ also determines the set of joint probabilities

$$\mathcal{P} = \left\{ P \left( a_{m(g_i g_\mu)}, b_{m(g_i g_\nu)} = o(g_i g_\nu) \right) \ | i = 0, \ldots, 5, j = 1, 2 \right\}.$$

Because the size of the search space increases exponentially with respect to the group size a random search is implemented. The size is $|G|^{N_o N_p}$ where $|G|$ is the order of the group, $N_o$ is the number of orbits we want, and $N_p$ is the number of parties. For a random choice of $\{\mu_1, \nu_1, \mu_2, \nu_2\}$ the biggest eigenvalue of $A$, $a_{max}$, is compared with the classical bound of the sum of the joint probabilities in $\mathcal{P}$, $c$. A violation is found when $a_{max} > c$.

The code can be downloaded from http://www.github.com/vug/bell-group-actions.
References


