Contextuality in a Deterministic Quantum Theory

Atul Singh Arora

Department of Physical Sciences

Indian Institute of Science Education and Research, Mohali

This dissertation is submitted for the degree of

Master of Science

2016

I would like to dedicate this thesis to my mother, who introduced me to the most valuable notion, that of freedom.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

Atul Singh Arora
2016

Acknowledgements

I acknowledge the contribution of my project guide, Prof. Arvind, who provided me with an environment conducive to research, and motivation to facilitate completion of this project. He encouraged independent thinking and guided me through difficulties I encountered.

Discussions with the Quantum Computation & Quantum Information (QCQI) group members, especially those with Rajendra Bhati and Kishor Bharti have been particularly efficacious. Inputs from Jaskaran Singh have also been helpful. The QCQI group seminars, specifically, those by Samridhi Gambhir and Dr. Arun Shehrawat, have helped narrowing the thesis problem.

Dr. Abhishek Choudhuri and Prof. Sudeshna Sinha, although formally not related to the project, have been kind enough to assist me at various instances with challenges I faced while numerically simulating the equations.

I am grateful to my parents and my colleagues/friends, who have been pivotal to helping me maintain mental balance, inspired me and provided emotional support, in ways they perhaps don't realize, specifically Manu Jayadharan, Prashansa Gupta, Kishor Bharti, Ritu Roy Choudhuri, Vivek Sagar, Yosman Bapat-dhar, Saumya Gupta, Shwetha Srinivasan, Evelyn Abraham and Srijit Mukherjee.

I end with acknowledging Immanuel Kant, for his philosophical work on ethics, which has heavily affected my thoughts, actions and outlook to life.

Abstract

The Copenhagen Interpretation of Quantum Mechanics (QM) asserts that the wavefunction is the most complete description, which entails that there is an inherent fuzziness in our description of nature. There exists a completion of QM, known as Bohmian Mechanics (BM), which replaces this fuzziness with precision, and re-introduces notions of physical trajectories. Various interesting questions arise, solely by existence of such a description; doesn't it contradict the uncertainty principle, for instance. Most of these questions were found to have been addressed satisfactorily in the literature. There was however, one question, whose answer has become the subject of the thesis; that of the paradoxical co-existence of contextuality and BM. Contextuality arguments show that the value an operator takes, must depend, in addition to the state of the system (including any hidden variables), on the complete set of compatible operators, to be consistent with QM. BM being deterministic, is at complete odds with this notion.

We were able to show, after various attempts, that the notion of contextuality is infact not necessary, by identifying another 'classical property' and constructing a non-contextual toy-model, serving as a counter-example to the impossibility proof.

Contents

List of Figures

List of Tables

Chapter 1

Prologue

1.1 Overview

A course in Quantum Mechanics (QM) typically leaves the reader with a fuzzy view of the world. The founders of the subject, were able to abstract out the mathematics from it's implications and interpretation. The view that the wavefunction is the most complete description of a system, known as the 'Orthodox Copenhagen Interpretation', is among the most popular. Einstein and Bohm, both played an important role in challenging this belief.
Einstein showed that if one makes reasonable assumptions about nature (apparently these were in spirit with his relativity theories), then QM must be incomplete. By incompleteness, he meant that the results of certain experiments should be predictable precisely, but QM fails at doing this. Thus he concluded that, QM must be treated as an intermediate theory and that a more complete description of nature should be possible. We will come back to this discussion. Bohm challenged this view by realizing that one can't experimentally refute the interpretation. His argument was that if for instance, the predictions of QM fail to match with experiments, then one can always add appropriate terms in the Hamiltonian until the difficulty is resolved. He was already anticipating the modern form of particle physics, where to accurately account for interactions between particles, one can postulate new force mediating particles, which was also done in the history of the subject, but can't refute the interpretation due to inaccuracy in predictions. He therefore, aimed at, and succeeded at constructing a complete quantum theory, now known as Bohmian Mechanics (BM), existence of which, would show that we are not justified at believing the interpretation, simply because there exist other alternatives.

Returning to Einstein, as will be described in more detail, Bell was able to construct a physical situation, that would test the very requirements Einstein imposed on a complete quantum theory; he was able to show that according to predictions of QM, such a complete description is impossible. This was verified experimentally (and was confirmed to be true, without any

possible loop-holes, only recently in 2015).
It might appear that Bell's test shows that QM is inconsistent with BM, but upon looking at the details, one can be convinced, that this is not the case. Infact, Bell was among the first few to popularize BM.

The discussion so far, is well known in the literature. However, certain developments which followed the said work of Bohm, again lead to apparent inconsistencies between BM and QM, which hadn't been satisfactorily addressed in the literature. These are the subject of the thesis.
For the sake of completeness, we merely name these, for one must look at the details, to be able to appreciate them. Greenberger, Horne and Zeilinger (GHZ), constructed a test that apparently showed determinism, the notion that observables have pre-defined values, can't exist. Developments due to Gleason, Bell himself, Kochen, Specker, Peres and Mermin, showed that the value an operator takes, must depend on the context in which it is measured; context here refers to the complete set of compatible operators.
The apparent contradiction must now be clear; BM is deterministic and it is at odds with the notion of contextuality as well.

1.2 The EPR argument

The EPR argument requires an entangled state, over two particles, which was originally written as

$$\psi(x_1,x_2) = \int e^{-i(x_1-x_2+x_0)p/\hbar}dp = \delta(x_1-x_2+x_0).$$

In the modern notation, one can write this as

$$\begin{aligned} |\psi\rangle &= \int \delta(x_1-x_2+x_0)dx_1dx_2\,|x_1\rangle_1\,|x_2\rangle_2 \\ &= \int |x\rangle_1\,|x+x_0\rangle_2\,dx. \end{aligned}$$

Similarly, one can write

$$\begin{aligned} |\psi\rangle &= \int e^{-i(x_1-x_2)p/\hbar}dpdx_1dx_2\,|x_1\rangle_1\,|x_2\rangle_2 \\ &= \int e^{-ix_1p/\hbar}\,|x_1\rangle_1\,dx_1e^{ix_2p/\hbar}\,|x_2\rangle_2\,dx_2dp \\ &= \int |p\rangle_1\,|-p\rangle_2\,dp. \end{aligned}$$

where the first ket is that of the first particle, and second that of the second. An attempt to reproduce the precise argument, will not be made. Instead, we satisfy ourselves with the

following simplified argument. We assume the principle of locality holds, which is to say that if two particles are sufficiently far away, then any change made to one particle, should not influence the other instantaneously. Given this assumption, consider two particles, which are sufficiently far away and that their state is given by $|\psi\rangle$. Now if the momentum of particle 2 is measured by observer 2, then observer 2, according to QM, will be able to predict the momentum of particle 1. Similarly, if the position of particle 2 is measured, then observer 2 can predict the position of particle 1. However, since the choice made by observer 2, can't influence particle 1 (by the assumption of locality), it follows that particle 1 had both it's position and momentum well defined, without being measured. Since QM fails to yield the answer with precision, whereas as we have shown, the answer was precise, we are forced to conclude that QM is not a complete description of nature.

Einstein ended his paper with this remark: "While we have thus shown that the wave function does not provide a complete description of the physical reality, we left open the question of whether or not such a description exists. We believe, however, that such a theory is possible."

1.3 Bell's Theorem

Bell's work addresses and satisfactorily answers Einstein's open question, of whether a *such* a description exists. Consider the following scenario. There are two observers, A and B, each has a particle. Each observer, can measure two properties, call them $\hat{a}_1, \hat{a}_2$ for observer A and $\hat{b}_1, \hat{b}_2$ for observer B, which yield a ± 1 outcome. Let $\langle \hat{a}_i \hat{b}_j \rangle$ represent the average value obtained by measuring the property $\hat{a}_i$ and $\hat{b}_j$ on the first and second particle respectively. Consider the average given by $\langle \hat{B} \rangle = \langle \hat{a}_1 \hat{b}_1 \rangle + \langle \hat{a}_1 \hat{b}_2 \rangle + \langle \hat{a}_2 \hat{b}_1 \rangle - \langle \hat{a}_2 \hat{b}_2 \rangle$. If one assumes local realism[1], that is that the values a_i takes are unaffected by those taken by b_i and that $\hat{a}_i, \hat{b}_j$ have pre-defined values (respectively), then $\langle \hat{B} \rangle \leq 2$. One can check this quickly by a brute force substitution of ± 1 values, in place of a_i, b_j.

Consider now, the state $|\psi\rangle = (|+-\rangle - |-+\rangle)/\sqrt{2}$, where $\hat{\sigma}_x|\pm\rangle = \pm|\pm\rangle$, and $\hat{\sigma}_{x,y,z}$ are the Pauli matrices. Let $\hat{a}_1 = \hat{\sigma}_z$ and $\hat{a}_2 = \hat{\sigma}_x$, while $\hat{b}_1 = -\frac{\hat{\sigma}_z + \hat{\sigma}_x}{\sqrt{2}}$ and $\hat{b}_2 = \frac{\hat{\sigma}_z - \hat{\sigma}_x}{\sqrt{2}}$. Upon evaluating $\langle \hat{B} \rangle$ using the rules of QM, one obtains $\langle \hat{B} \rangle = 2\sqrt{2} \nleq 2$. This entails that our assumption, that of local realism, must be incorrect, which settles the question: A *local* hidden variable description can not exist. The word hidden is used to represent information about the state of the system, that is not contained in the wavefunction. It is instructive to emphasize that a non-local description, may still be possible. It is also important to note that, despite this non-locality, one can't use QM to send signals instantaneously (superluminal communication

[1] Infact, an argument similar to EPR can be used to show that locality entails realism.

is not possible). This is known as the no-signalling theorem and can be proven by showing that marginals in QM, don't depend on the system which is traced out (TODO).

1.4 Bohm's Theory, Bohmian Mechanics (condensed introduction)

Bohm gave a precise, but non-local description of quantum phenomena. Let us start with non-interacting particles. A particle is associated with (1) a position, q & momentum, p, precisely defined and (2) a wavefunction $\psi = Re^{iS/\hbar}$. The postulates of the theory are: (a) Evolution of the wavefunction, is governed by Schrodinger's equation: $i\hbar\partial\psi/\partial t = -(\hbar^2/2m)\nabla^2\psi + V\psi$. (b) The particle is guided by the wavefunction: $\dot{q} = p/m$ where $p = \nabla S = \hbar\mathrm{Im}(\nabla\psi/\psi)$. (c) The initial distribution of the particles is given by $\rho(x) = |\psi|^2$.

The astute reader would've noticed that ∇S is just the probability current, which entails that if the initial distribution satisfies $|\psi(t_0)|^2$, then it will do so at all times t, thereafter. Before generalizing this to multiple particles, let us first see a quick consequence of this formulation. In the Orthodox Copenhagen interpretation, the double slit experiment is a source of mystery and the which slit question, that of confusion. In BM, since trajectories are well defined, one observes (TODO: add image) that the particle goes through precisely one slit and then later, forms the interference pattern. The wavefunction goes through both slits, and interferes to create the pattern. If one of the slits is blocked, then as expected, since the wavefunction can't interfere, the pattern is lost. We haven't spoken about measuring the particle yet, but if one notes that to measure the particle, the potential at one of the slits must be changed, then it is immediate that the interference pattern will be effected, since the wavefunction is affected by the potential at both slits, even if the particle passes through only one. One finds that by using the following multi-particle generalization, and an appropriate measuring scheme, one can show more precisely how the process of measuring effectively destroys the pattern, replacing the mystery with clarity.

For N interacting particles, we have $p_i = \nabla_i S(q_1, q_2, \ldots, q_N)$ where note that the momentum of the i^{th} particle, depends on the instantaneous positions of all the particles. Consequently, BM is an explicitly non-local, but complete description.

As a final remark, it must be added that spins can also be included in BM, however, a particle is not associated with a specific spin, only it's wavefunction is. For a spinor, say $\psi \equiv (\psi_+, \psi_-)^T$, the generalization is that $p = \hbar\mathrm{Im}((\psi, \nabla\psi)/(\psi, \psi))$ where $(.,.)$ represents inner product in the spin space $\mathbb{C}^2$. A quick illustration is in order. Consider the following Stern-Gerlach setup; Quantum effects are considered only along the x-axis. A particle moves

along the z-axis, with speed v_z, and it's initial wavefunction, given by a Gaussian centred at the origin, say $\psi(x) = (1/\sqrt{2\pi}\sigma)e^{-x^2/2\sigma^2}$, viz. $|\psi\rangle = \int dx\psi(x)|x\rangle$. It's spin state is given by $|\chi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, where $|0\rangle$ and $|1\rangle$ represent computational basis. Along the z-axis, a strong heterogeneous magnetic field is present, whose action maybe captured by $H_{\text{int}} = a\hat{p}_x \otimes \hat{\sigma}_z$ where a is a constant that quantifies the strength of the field. Why this particular form works, will become clear momentarily. Assuming that a is large enough to neglect effects of free-evolution, we have $\hat{U}(t) = e^{-ia\hat{p}_x \otimes \hat{\sigma}_z t/\hbar}$. Thus, if the initial state is $|\Psi\rangle = |\psi\rangle \otimes |\chi\rangle$, then

$$\begin{aligned} |\Psi(t)\rangle &= \hat{U}(t)|\Psi\rangle \\ &= \frac{e^{-ia\hat{p}_x t/\hbar}|\psi\rangle \otimes |0\rangle + e^{ia\hat{p}_x t/\hbar}|\psi\rangle \otimes |1\rangle}{\sqrt{2}} \\ &= \frac{|\psi_{at}\rangle \otimes |0\rangle + |\psi_{-at}\rangle \otimes |1\rangle}{\sqrt{2}} \end{aligned}$$

where $|\psi_{x_0}\rangle \equiv \int dx\psi(x-x_0)|x\rangle$, viz. a Gaussian wavepacket centered at x_0. One can plot $|\Psi|^2$, as a function of z, using $z = v_z t$, schematically as shown in the figure (TODO: put the figure), where regions enclosing, say 70% of $|\Psi|^2$ have been outlined. So far, according to QM, if we measure the position of the particle, then (given $\sigma \ll at$), obtaining the particle near $x = at$, is as probably as finding it near $x = -at$. QM doesn't make a deterministic prediction at this stage. According to BM, if in addition to the wavefunction, we assume that the particle was initially at some $x > 0$, then we can predict precisely where the particle will turn up. Infact, one can accomplish this with essentially no calculation. Since $\dot{x} = \nabla S/m$, a single valued function, it follows that trajectories of particles won't intersect, where recall that S is given by $\psi = Re^{iS/\hbar}$. Consequently, a particle that starts with $x > 0$, must follow the 'up' trajectory, else by symmetry, the trajectories will have to intersect. We conclude therefore, that if $x > 0$ initially, the particle has spin $|0\rangle$ and if $x < 0$, the spin must be $|1\rangle$. This appears very clear and intuitive. A non-intuitive aspect of this, is that we can't associate spins with particles, even though we can predict precisely, the outcome of the experiment. This is manifested by the observation that if $H_{\text{int}} \to -H_{\text{int}}$, which practically amounts to reversing the heterogeneity of the magnet, then $\sqrt{2}|\Psi(t)\rangle = |\psi_{-at}\rangle \otimes |0\rangle + |\psi_{at}\rangle \otimes |1\rangle$. Again, if $x > 0$ initially for a particle, it will follow the 'up' trajectory. However, now this corresponds to spin $|1\rangle$, as opposed spin $|0\rangle$. We have thus demonstrated that we can't associate spin uniquely to a particle, and that it must only be associated with the wavefunction.

1.5 Determinism: The GHZ test

One drawback of Bell's test, was that it's statistical nature. The GHZ test, takes this a step further and shows that one can't even conceive of having pre-defined values. The construction requires three observers with one particles each. Each observer can measure two properties, X and Y, with outcomes ± 1. We start with the state $\sqrt{2}|\chi\rangle = |000\rangle - |111\rangle$ and note that for $\hat{A} := \hat{\sigma}_x \otimes \hat{\sigma}_y \otimes \hat{\sigma}_y$, $\hat{A}|\chi\rangle = |\chi\rangle$, where the property X is the projection of spin of the particle along the x-axis and Y is that along the y-axis. Thus a measurement of $\hat{A}$ (the first observer measures X and the other measure Y), will yield $+1$ with certainty. Next, we define $\hat{B} := \hat{\sigma}_y \otimes \hat{\sigma}_x \otimes \hat{\sigma}_y$ and $\hat{C} := \hat{\sigma}_y \otimes \hat{\sigma}_y \otimes \hat{\sigma}_z$, which by symmetry, must also yield $+1$ for the said state. If these observables, X and Y had predefined values, then a measurement of $\hat{A}\hat{B}\hat{C}$ would be equivalent to measuring $\hat{D} := \hat{\sigma}_x \otimes \hat{\sigma}_x \otimes \hat{\sigma}_x$, since Ys appear twice for each particle, and $Y^2 = 1$. Since a measurement of each $\hat{A}$, $\hat{B}$ and $\hat{C}$ yields a $+1$, it follows that a measurement of $\hat{D}$ therefore, must also yield $+1$. However, $\hat{D}|\chi\rangle = -|\chi\rangle$, which entails that a measurement of $\hat{D}$ must yield a -1. Thus, we arrive at a contradiction and conclude that the assumption that the properties X and Y had predefined values, must be incorrect.

1.6 Contextuality: The Peres Mermin test

There've been various developments, of which the simplest, with effectively the same consequence, is discussed: the Peres Mermin test. In this test also, we will show that pre-defined values can't exist, but in addition, also define the notion of contextuality. Consider the following set of operators

$$\hat{A}_{ij} \doteq \begin{bmatrix} \hat{\mathbb{I}} \otimes \hat{\sigma}_x & \hat{\sigma}_x \otimes \hat{\mathbb{I}} & \hat{\sigma}_x \otimes \hat{\sigma}_x \\ \hat{\sigma}_y \otimes \hat{\mathbb{I}} & \hat{\mathbb{I}} \otimes \hat{\sigma}_y & \hat{\sigma}_y \otimes \hat{\sigma}_y \\ \hat{\sigma}_y \otimes \hat{\sigma}_x & \hat{\sigma}_x \otimes \hat{\sigma}_y & \hat{\sigma}_z \otimes \hat{\sigma}_z \end{bmatrix}$$

which have the peculiar property that all operators along a row (and along a column) commute. It is trivial to see that this holds for the first two rows and the first two columns. To see that this holds also for the last row and column, note the anti-commutation relation, $\{\hat{\sigma}_x, \hat{\sigma}_y\} = 0$ and that $\sigma_z = i\sigma_y\sigma_x$, which one may check explicitly. Another interesting property is that the product of rows (columns) yield $\hat{R}_i = \mathbb{I}$ and $\hat{C}_j = \mathbb{I}\,(j \neq 3)$, $\hat{C}_3 = -\mathbb{I}$, $(\forall\, i, j)$ where $\hat{R}_i \equiv \prod_j \hat{A}_{ij}$, $\hat{C}_j \equiv \prod_i \hat{A}_{ij}$. This can be verified easily by using the aforesaid relations and the fact that $\sigma^2 = 1$, for every Pauli matrix. Using this property of R_i and C_j, it is easy to show that no pre-defined values for operators can exist. Let us assume that pre-defined values do exist. Not that to get $C_3 = -1$, we must have an odd number of -1 assignments in the third column. In the remaining columns, the number of -1 assignments must be even for each column. Thus, in

the entire square, the number of -1 assignments must be odd. Let us use the same reasoning, but along the rows. Since each $R_i = 1$, we must have even number of -1 assignments along each row. Thus, in the entire square, the number of -1 assignments must be even. We have arrived at a contradiction and therefore we conclude that our assumption that operators have predefined values, must be wrong. Note that unlike the GHZ test, the Peres Mermin test is state independent.

We are now in a position to discuss the notion of contextuality. We begin with defining a non-contextual assignment to be one, where the assignment depends only on the state ($|\psi\rangle$ + hidden variables, if any) and the operator to which the assignment is being made. It follows that we had tacitly assumed a non-contextual assignment, that resulted in a contradiction. If we allow for the value of an operator, to depend on the context in which it is measured, where the context is meant to refer to a set of compatible observables, then the contradiction won't arise. For instance, if we imagine that all observables yield a $+1$, except $\hat{A}_{33}$ which yields a -1 if measured with $\hat{A}_{32}, \hat{A}_{31}$ and yields a $+1$ if measured with $\hat{A}_{23}, \hat{A}_{13}$. If this doesn't appear entirely unsatisfactory, then you're on the right track and are encouraged to read the more detailed description given in Section [??].

Chapter 2

Determinism Tests and Bohmian Mechanics

2.1 Rationale

2.2 GHZ

2.2.1 Compatibility with BM

2.2.2 Phase Space

2.2.2.1 Known Extension

2.2.2.2 Optimized Extension

2.2.2.3 Challenges

2.3 BM Simulator

2.3.1 Design of Numerics

2.3.1.1 Simulation of Schrödinger's Equation

2.3.1.2 Simulation of trajectories

2.3.2 Performance against analytic results

2.3.3 Conclusion

2.4 Measurements in BM (I)

2.4.1 Observable with discrete spectrum | Hamiltonain Approach

2.4.1.1 Generalization to observables with continuum spectrum

2.4.1.2 Consistency check; measurement of position

Chapter 3

Contextuality and Bohmian Mechanics

3.1 Rationale

3.2 Intensifying Contextuality

3.2.1 GHZ escalated to Contextuality

3.2.2 Peres Mermin escalated to Continuous Variables

3.3 Measurements in BM (II)

3.3.1 Stern Gerlach based measurements

3.3.1.1 Single Spin

3.3.1.2 Entangled Spin

3.3.2 Hamiltonian Approach

3.3.2.1 Single Spin

3.3.2.2 Entangled Spin

3.3.2.3 GHZ Entangled Spin

3.4 Peres Mermin Revisited

3.4.1 Assumptions and Definitions

3.4.2 BM explanation of Peres Mermin

3.5 Conclusion

Chapter 4

An Alternative to Contextuality

4.1 Rationale

4.2 Multiplicativity

4.3 Simple Models

4.3.1 Contextual, Memory Model

4.3.2 Non-Contextual, Toy Model

4.4 Generic Models

4.4.1 Generalized Toy Model

4.4.2 C-ingle Theory

4.5 Conclusion

Chapter 5

Epilogue

5.1 The Verdict: Contextuality vs. Bohmian Mechanics

5.2 Summary of Results

5.3 Conclusion

5.4 Digressions

5.4.1 Attempt at superluminal communication with BM

5.4.2 Identical Particles, alternative approach

References

Appendix A

Bohmian Mechanics, from Bohm's papers

The usual interpretation of quantum mechanics (QM) is self consistent. It assumes however that the most complete possible specification of an individual system is in terms of ψ that yields only probabilistic results. To investigate if this assumption is accurate, it is reasonable to expect an experimental test to exist. However, as will be shown, no experiment can test this interpretation. The other alternative then, is to attempt constructing a theory, which has extra variables which in principle completely specify the system but in practice get averaged over. If such a theory existed then it is obvious that the assumption false. In his paper, Bohm describes such a theory. This theory yields the same results as QM in all physical situations studied and is shown to be broader conceptually. The mathematics used is shown to be more general. Bohm's idea was that while QM fails at high energies, his theory might work. In retrospection though, one realizes how both BM and QM suffer from the same problem which requires quantum field theory (QFT) to remedy.

A.1 Introduction

The usual interpretation assumes that the wavefunction is the most complete description. This description was criticized by Einstein for he believed that there must exist a better, precise theory that describes nature as opposed to QM that claims nature is fundamentally probabilistic. Einstein's views have themselves been criticised to be irrelevant because the usual QM interpretation is in excellent agreement with experiments.

Appendix B

Bohmian Mechanics, from Holland

It is worthwhile spending some time trying reformulate classical mechanics (CM) into a form that is closest to quantum mechanics. Although this is not necessary for understanding BM (and may very well be skipped), doing so brings with it it's own perspective and excitement.

B.1 The Hamilton Jacobi Theory

B.1.1 Introduction

Let's start with a Lagrangian, $L = L(q,\dot{q},t)$, where $p_i := \partial L/\partial \dot{q}_i$, $H := p_i\dot{q}_i - L$, q_i represents the position of the ith particle and $q,\dot{q}$ refer to the set $\{q_1,q_2,\dots\}$, $\{\dot{q}_1,\dot{q}_2,\dots\}$ respectivley. If we imagine a path that the particle takes from a point q_0,t_0 to q,t, then we can define a quantity $I(q,t;q_0,t_0) := \int_{q_0,t_0}^{q,t} L(q,\dot{q},t)dt$ where the integral is taken along the aforesaid path. We know from experiments that the path the particle actually takes, given L, is one which extremizes I. This entails that, just like at the extremum value of the function, it's derivative vanishes, the first order variation of I for a path that extremizes I must vanish; viz. $\delta I = 0$ for any extremal path. As the reader would recall, after some calculation, this yields

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0.$$

This can be expressed in terms of H instead of L by substituting for L in the definition of I and changing the independent variables from $(q,\dot{q})$ to (q,p) & varying them, where p refers to $\{p_1,p_2,\dots\}$. The result is the familiar Hamilton equations (this again requires some calculation and may not always be possible)

$$\dot{q}_i = \frac{\partial H}{\partial p_i},\ \dot{p}_i = -\frac{\partial H}{\partial q_i}.$$

Defn: Canoncial Transformation: $q,p \to Q = Q(q,p,t), P = P(q,p,t)$ is a a canonical transformation if Q, P preserve the form of the Hamilton Equations, viz. if $K = K(Q,P,t)$ is the new H, then we must have $\dot{Q}_i = \frac{\partial K}{\partial P_i}, \dot{P}_i = -\frac{\partial K}{\partial Q_i}$ which could've come from $I = \int P_i dQ_i - K dt$.

It is known that if two Ls differ by $\frac{df}{dt}$, where f [check: where $f = f(q,\dot{q},t)$] is some function, then they yield identical equations of motion. [) This follows at once by recalling that $I_1 = \int_{q_0,t}^{q,t} L_1 dt$ and noting that $I_2 = \int_{q_0,t}^{q,t} L_1 dt + F - F_0$ entails that the variation $\delta I_2 = \delta I_1 + \cancelto{0}{\delta F} - \cancelto{0}{\delta F_0} = \delta I_1$(]. Thus if we write the Lagrangian in terms of the Hamiltonian and impose

$$P_i dQ_i - K(Q,P)dt + \frac{dF}{dt} = p_i dq_i - H(q,p)dt,$$

then we're guaranteed that the equation of motion (EOM) will be identical and hence describe the same physical system. Assuming a specific functional form, viz. $F = F(q,Q,t)$, one can evaluate

$$\begin{aligned} p_i &= \frac{\partial F}{\partial q_i}, \\ P_i &= -\frac{\partial F}{\partial Q_i}, \\ K &= H + \frac{\partial F}{\partial t}. \end{aligned}$$

[) This follows from the restatement $\left(P_i + \frac{\partial F}{\partial q_i}\right)\dot{Q}_i - \left(p_i - \frac{\partial F}{\partial q_i}\right)\dot{q}_i + \left(K - H - \frac{\partial F}{\partial t}\right) = 0$ and by choosing q, Q to be the independent variables. (] Note that the first equation will yield Q in terms of p and q. The following equation yields P in terms of Q and q. The last equation yields K, since Q and P are known (it's not obvious though that writing $K = K(Q,P)$ will be straight forward).

Thus, given $F(q,Q,t)$ we can find the canonical transformation $Q = Q(q,p,t), P = P(q,p,t)$ and even $K = K(q,p,t)$. Defn: (by example) F is a generating function.

Next let us consider a case where the transformed variables are almost the same as the original set, viz.

$$\begin{aligned} Q_i &= q_i + \delta q_i, \\ P_i &= p_i + \delta p_i. \end{aligned}$$

The claim is that $\delta q_i = \varepsilon \partial G/\partial p_i$ and $\delta p_i = -\varepsilon \partial G/\partial q_i$, where $G = G(q,p,t)$ is the generating function and ε is small. [) This is plaussible by simply taking a generating function which is a function of (q,p,t) instead of (q,Q,t) as was F (] Note that for $\varepsilon = dt$, $G = H$, it follows

$\partial q = dt\partial H/\partial p_i = dt\dot{q}_i = dq_i$ and similarly $\delta p_i = dp_i$. The central idea is that infinitesimal canonical transformations $\leftrightarrow$ infinitesimal motion. Successive infinitesimal canonical transformations yield a finite canonical transformation. Successive infinitesimal motion describes observed motion. Consequently, any observed classical motion can be described by an appropriate canonical transformation. This can be used to find the motion itself, in fact.

One tries to find a canonical transformation corresponding to $q = q(q_0, p_0, t)$, $p = p(q_0, p_0, t)$ which can be inverted to yield q_0, p_0(at t_0) given q, p, t. [] Thus, all points along the trajectory, viz. the set (q, p, t), all map to the same point (q_0, p_0, t_0). [] The mapped point, must not evolve with time. Thus, we demand (or require) that $K = 0$ (basically must be constant, can be $\neq 0$) so that $\dot{Q} = 0, \dot{P} = 0$. Conventionally, the associated generating function is denoted by S. Assume that $S = S(q, Q, t)$. From the earlier discussion, it follows that $p = \partial S/\partial q$:= Jacobi Law (Defn). Also, $K = H + \partial S/\partial t \implies \partial S/\partial t + H(q, \partial S/\partial q, t) = 0$:= Hamilton Jacobi Equation (Defn). This is a first order partial differential equation in $(n+1)$ variables, q, t. It is sufficient to define S. Defn: S :=Hamilton's principle function. Since only derivatives of S appear in the EOM, only n non-additive constants are relevant, viz. $S = S(q, \alpha, t)$. By assumption, $\alpha = Q$; however in general, one may use a combination of Qs and Ps.

Scheme 1: Consider the following overview of how this is implemented in practice. (1) Substitute $\partial S/\partial q = p$ in $\partial S/\partial t + H = 0$. (2) Solve for S and obtain n non-additive integration constants, $\alpha_1, \ldots \alpha_n$. (3) Define independent linear combinations of α_i as $Q_i \equiv \gamma_i(\alpha_1, \ldots \alpha_n)$. (4) Now recall that $P_i = -\partial S/\partial \gamma_i \equiv \beta_i$. From this equation, in principle, q can be obtained in terms of (β, γ), granted $\det\left(\partial^2 S/\partial q_i \partial \gamma_j\right) \neq 0$. [] Where did the Q, β get defined? Well, you specify it by matching the initial conditions t_0, q_0, p_0 with Q_i and $P_i (= \beta_i)$ using appropriate linear combinations of α_i. []

B.1.2 The Action Function

Assume that the system starts from (q_0, t_0) and reaches (q, t) through the path γ. Accordingly, the (extremized) action will be $I(q, t; q_0, t_0) = \int_\gamma L(q, \dot{q}, t) dt$. The idea is to look at variations in I, induced by varying q, t and q_0, while holding t_0 fixed. The motivation for this is the Hamilton Jacobi formalism, as was just set up. Upon substituting for H in the expression for I, one can evaluate $dI = pdq - Hdt - p_0 dq_0$. [] This step needs some work to derive, but take it as a claim for the moment [] Also, $dI = (\partial I/\partial q)dq + (\partial I/\partial t)dt + (\partial I/\partial q_0)dq_0$. Equating these yields $\partial I/\partial q = p$, $H = -\partial I/\partial t$, $p_0 = -\partial I/\partial q_0$. Using the first two, it surprisingly follows that I is a valid Hamilton's principle function. [] Ofcourse, now it's clear from the aforesaid equations that I generates canonical transformations which trivializes the motion ($K = 0$, nothing evolves). Here $q_0 \equiv Q$ and $p_0 \equiv P$ []

In Scheme 1 (Sec. B.1.1), the Hamilton Jacobi equation is integrated to solve Hamil-

ton's equation. The converse, viz. solving the Hamilton's equation to solve the Hamilton Jacobi equation, subject to $S(q,t_0) = S_0(q)$, can be achieved (it's a "Cauchy Problem"[1]); this is Scheme 2. $S(q,t_0) = S_0(q)$. (after this is not quite clear. Ask Arvind sir)

B.1.3 The Basic Law of Motion

There's an issue: if a complete solution of the Hamilton-Jacobi equation is not available, then the Jacobi Law can't be used. We assume that $S(q,t)$ is known, but the constants are not, viz. $\partial S/\partial \gamma_i = \beta_i$ doesn't help. Recall that $p_i = \partial S/\partial q_i = \partial L/\partial \dot{q}_i$. From L, one can evaluate $p_i = p_i(q_i,\dot{q}_i,t)$ and then use $p_i = \partial S/\partial \gamma_i$ to integrate for $q(t)$ with $q_0 = q(0)$.

Remarks: (1) For consistency, we must impose $p_0 = (\partial S/\partial q)_{q=q_0,t=t_0}$. (2) Even when the complete integral is known, p_i as a function of L is used to specify the initial conditions. (3) Since $p_i = \partial S/\partial q_i$ always holds, it maybe regarded as the basic law of motion.

Illustration (1) [using Scheme 1]: Recall that $p = \nabla S$. Assume $L = m\dot{q}^2/2 - V(q,t)$. $\implies$ $p = m\dot{q}$. S must satisfy $\partial S/\partial t + \nabla S^2/2m + V = 0$. For a free particle, we may use $S = P_i P_i t/2m + P_i q_i$ as a solution, where P_i is some constant. $P = \nabla S = p = m\dot{q}$. This integrates to $q = q_0 + P(t-t_0)/m$. We can also use the action, $S = m(q-q_0)^2/2t$ as a solution. [) That can be worked out as follows: Recall that $I = \int_\gamma L dt$. For a free particle, $= \int_\gamma (m\dot{q}^2/2)dt = \int_\gamma (1/2)m(dq/dt)(dq/\cancel{dt})\cancel{dt}$. Now to impose γ as the path, we must put $dq/dt = v$ (constant). Thus, $I = mv(q-q_0)/2 = m(q-q_0)^2/2t$. It might appear pointless to solve the motion, when we have already evaluated γ. The point is that, given S, how to solve the motion. This was just one way of finding S. (] From $p = \nabla S$, we get $p = m(q-q_0)/t$; $P = \partial S/\partial Q$, with $Q = q_0$, yields $P = p$. Thus we recover $q = q_0 + P(t-t_0)/m$. Remarks: (1) S in the former case, is constant in q,t while in the latter, it explicitly depends on q,t. (2) S may not have a constant value along the particle trajectory; eg. $S = m(q-q_0)^2/2t = mv^2t^{\cancel{2}}/\cancel{t}$ for a free particle. (3) While S may not be single valued, physical quantities will be well defined, such as momentum (∇S), energy ($-\partial S/\partial t$) etc. [) Note how we started by saying that $S = S(q,Q,t)$ but suddenly, there's P in the first definition of S but not Q, though in the second definition, $S = S(q,Q,t)$ where $Q = q_0$. So how is the former a valid S (in the sense that how can we use results we derived by assuming $S(q,Q,t)$ as opposed to $S(q,P)$ (]

[Ask Arvind sir: super confusing] Let us now explore how..

Illustration (2): Consider $S(q,E,t) = W(q,E) - Et$

[1] A PDE problem with constraints defined on some hyper-surface of the domain

B.2 Densities

1. To understand the continuity equation $\rho + \nabla.(\rho v) = 0$, examples of v are taken

   (a) $v = v(x)$, a solution is obtained as

   $$\rho(x,t) = \frac{1}{v(x)} v \left[x \left(t - \int \frac{dx}{v} \right) \right] \rho_0 \left[x \left(t - \int \frac{dx}{v} \right) \right]$$

   in which further assuming that $\rho(x)$ results in $\rho = A/|v(x)|$

   i. $v = v(t)$ then we get

   $$\rho(x,t) = \rho_0 \left(x - \int v dt \right)$$

   which means that ρ is constant along particle trajectories

   (b) connection with Liuoville's equation

   i. $f(x,p,t)$ is defined instead of $\rho(x,t)$. Pure and mixed states are defined accordingly as $f(x,p,t) = \rho(x,t)\delta(p - \nabla S(x,t))$ being pure and the remaining as mixed.

   ii. $\frac{df}{dt} = \partial_t f + \frac{1}{m}\sum p_i \partial_{x_i} f - \sum \partial_{x_i} V \partial_{p_i} f = 0$ is the Liuoville's equation (which holds since we can show that the volume doesn't change under Hamiltonian evolution and particles inside the volume stay inside; $f(p',q',t+\delta t) = f(p,q,t)$ is essentially the statement $\frac{df}{dt} = 0$) which is linear in f.

   iii. One may project out the moment space. They define equivalent of ρ as $P(x) = \int f d^3p$, mean momentum as $\overline{p_i(x)} = \frac{\int p_i f d^3p}{P(x)}$ and $\overline{p_i p_j(x)} = \frac{\int p_i p_j f d^3p}{P(x)}$. The louvielle equation can then be expressed in terms of these spatial variables. Integrating it we get

   $$\partial_t P + \frac{1}{m}\sum \partial_{x_i}(P\overline{p_i}) = 0.$$

   To get the momentum transport equation, after multiplying the louvielle equation with p_i and integrating, we get

   $$\partial_t(P\overline{p_i}) + \frac{1}{m}\sum \partial_{x_j}(P\overline{p_i p_j}) + P\partial_{x_i} V = 0$$

   (apparently integrated by parts and assumed $f \to 0$ as $p_i \to \infty$)
   While f is constant along a phase space trajectory, the spatial density P (equivalent of ρ) is not. It's apparent from the derivation of the continuity equation; either we start with a fixed volume or a fixed number of particles, not both.
   If you substitute $f = \rho\delta(p - \nabla S)$ as stated earlier, you'd get $P = \rho$, $\overline{p_i} = \partial_{x_i} S$,

$\overline{p_i p_j} = \partial_{x_i} S \partial_{x_i} S$ as expected. The substitution also yields what's called a field theoretic version of Newton's Laws given by

$$\partial_t \rho + \frac{1}{m} \nabla . (\rho \nabla S) = 0$$

and

$$\left[\partial_t + \frac{1}{m} \sum \partial_{x_i} S \partial_{x_j} \right] \partial_{x_i} S = 0$$

iv. Remarks:

A. It's not obvious that if we start with a state that has well defined momentum (delta distribution) but the positions are given by $\rho(x)$, then they will continue to be well defined in momentum. This happens only exceptionally. In general, a pure state maybe sent to a mixed state. We'll see examples of these. [todo: ensure examples make sense]

B. Can we decompose any mixed ensamble into a linear combination of pure ones? The answer's no. [proof?] Say there are many solutions of the Hamilton-Jacobi equation, given by S_i. Thus, we can construct a linear combination as $f(x,p,t) = \sum P_i \rho_i(x,t) \delta(p - \nabla S_i(x,t))$ where P_i (degenerate notation) refers to the distribution of momenta at a given point. $\sum P_i = 1$ is assumed for normalization. Claim is that this is not in general possible to decompose a state into this form. An explicit example is that of reflecting through a potential barrier (in CM) [todo: ensure the example works]

C. While this is not particularly useful in CM (the pure and mixed states), the formalism helps in comparison with QM.

(c) Pure and Mixed States

i. Illustration: We see that $f_0(x,p) = \delta(x - x_0)\delta(p - p_0)$ remains sharp
(it can be checked by inserting it in the louviel equation) to yield $f(x,p) = \delta(x - x(t,x_0,p_0))\delta(p - p(t,x_0,p_0))$ [this is expected, since you're in essence saying there's only one particle]

ii. Illustration 2: We want to see what happens to a Gaussian like state, does it spread?
We start with $\rho_0 = \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$ and $S_0 = px$ with σ and p constant. This form of S_0 has already been solved for and tells us $\rho = \frac{e^{-(x-vt)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$. There's no spreading classically! We'll see for the same initial conditions, what happens quantum mechanically.

iii. Illustration 3: What initial conditions yield a spreading Guassian? We start with the same ρ_0 but use $S_0 = \frac{m(x-x_0)^2}{2t}$, in which case the solution we saw the result is

B.3 Aside

So we start with $\oint dS = \oint \nabla S.dx = \oint p.dx = \int \nabla \times p.da$. If $\oint dS = nh$, then we must have $\nabla \times p = \sum_a \Gamma_a \int_{\gamma_a} \delta(x - x_a) dx_a$ where γ_a is the nodal line. If we assume $\oint dS = nh$ holds, then can we construct some example of the same? Let's first see how $\oint dS = nh$ can be derived. If the only condition is that ψ is single valued, then we know that at any point, S' and S both yield the same ψ, where $S' = S + nh$. If one considers a loop, then say we start from a point S_a. Then after completing some distance, the change in S is given by ΔS. So the value of S starting from $S_a$will be $S_a + \Delta S$. Now if we come back to the point a, then from uniqueness of ψ, we only demand $S_a + \Delta S = S_a + nh$. If S itself was unique, then we'd say $S_a + \Delta S = S_a$. Now at this point itself I seem to have trouble. I have tacitly assumed that S is single valued when I'm evaluating the 'change is S' along the curve.

Talked to manu for a while and made some progress, then figured it was non-sense and made some more progress. Finally, Manu found a document that helped clarify a few things. The issue was still that they had used a vector field and not a potential. And it wasn't clear to me what potential must I use in that case.

The potential is $V = k\theta$. Note how this is itself, as a function of position is multi valued and yet we never have any issues integrating this (as we'll see shortly). While V is multivalued, $\nabla V = \frac{k}{r}\hat{\theta}$ is happily single valued :) And not just that, check this; $\oint_\gamma \nabla V.dx = 2\pi$ (simply because γ is chosen to be a circle and then $dx = rd\theta\hat{\theta}$). Since in the domain of interest, everything is well defined, I can write $\oint_\gamma \nabla V.dx = \oint_\gamma dV = 2\pi$. And one can show independently (I know only a simple minded proof with discritizing the function) that $\oint dV = 0$ whenever V is single valued (or a function). So what does this example show? Various rather peculiar things. (I) that $\oint dV$ maybe non zero for a reasonable physical situation by virtue of multivaluedness of V. Yes, V is multivalued and yet we can integrate the said expression without ambiguity. (II) that there happens to be a singularity within the loop, over which the integral is non-zero. (III) The curl, $\nabla \times \nabla V \neq 0$ at the center and $= 0$ else.
Now we've made plausible various things which would've seemed arbitrary otherwise.