Notes on Variational Quantum Imaginary Time Evolution (VarQITE)

Xiao Suyang
Summer Research Fellowship

July 11, 2025

Abstract

This is an expository note that builds up to the VarQITE algorithm, with the motivation of filling up various gaps of knowledge that I had during my Summer Research Fellowship(SRF). We start from the calculus of variations and try first to establish McLachlan's variational principle. Next, since the topic of my SRF is to investigate the application of VarQITE in solving SDEs, we also cover the process of executing VarQITE on quantum circuits, the derivation of the Feynman-Kac formula, which is used in embedding the SDE, and some further discussion in ansatz design for specific SDEs for optimization.

Contents

1 Calculus of Variation

We start by detailing the concepts of the calculus of variation and explore its role in defining the Action and Lagrangian in classical mechanics.

A functional is a function of a function, i.e. taking a function as an input and a number as an output. We define the functional using the following map from C^2 to $\mathbb{R}$:

$$C^2 \xrightarrow{I[y]} \mathbb{R}$$

For example, the following is a functional mapping $f(x)$ to a number:

$$I[y(x)] = \int_a^b y(x)\eta(x)\,dx$$
$$y(x) \mapsto \int_a^b y(x)\eta(x)\,dx$$

Where $\eta(x)$ is a smooth function. Technically, a functional is just the composition of two functions, like $f \circ g$, but g is a variable that we change. We can write it as something like $f \circ (-)$. The usefulness of a functional is rooted in its application. For example, when given a function that is a "path" between two points, we can construct a functional that takes in the path function y as its input, and outputs various properties regarding that path as a number, such as the length of the path, the average speed of something travelling along that path, etc..

Just some additional remarks, we see a lot of duality between a functional and a function. For example, a functional that is "path-independent" is like a function that is a horizontal line.

Additionally in physics, the potential between points is simply a "path-independent" functional. Notice how we defined the function $y(x)$ to be at least twice differentiable. In this context, when considering one variable only, it is helpful to treat $y(x)$ as a path between some points in space.

A form of a functional that we use very frequently is formulated in the following way:

$$I[y] = \int_a^b F(x, y, y')\,dx \tag{1}$$

where y is a function in x. This formulation is heavily motivated by problems in physics, where $y(x)$ and $y'(x)$ naturally translate to position and velocity along some path.

A central problem that we are interested in is investigating when a functional takes stationary values, i.e. for which $y(x)$ is the functional maximized or minimized. Just like for a function, where we consider a δx in the input, and inspect at which point

- TODO: variation

Lemma 1.1: Fundemental lemma of the calculus of variations

let $y(x)$ be a continuous function on the interval $[a, b]$ that satisfies

$$\int_a^b y(x)\eta(x)\,dx = 0$$

For all $\eta(x)$ that is compactly supported and smooth, then $y(x)$ is identically 0.

Proof. Assume to the contrary that there exists $x_0 \in (a, b)$ such that $y(x) \neq 0$. Then by the continuity of $y(x)$, we can find an interval $[c, d]$ containing x_0 where $y(x) > 0$.

Consider the following Bump function:

$$B(x) = \begin{cases} e^{-\dfrac{1}{(x-c)(d-x)}} & c < x < d \\ 0 & \text{otherwise} \end{cases}$$

With $a < c < x_0 < d < b$. This function satisfies the criteria for $\eta(x)$, where it is smooth (infinitely differentiable), and has a support of $[c, d]$, which is compact.

Then, we must have

$$\int_a^b y(x)B(x)\,dx = \int_c^d y(x)B(x)\,dx > \int_c^d 0 \cdot B(x)\,dx = 0$$

Which contradicts our assumption. □

We can extend this lemma further, which will be useful in proving the Euler-Lagrange equation:

Lemma 1.2: Extension of the fundamental lemma

let $y(x)$ be a continuous function on the interval $[a, b]$ that satisfies

$$c_1\eta(a) + c_2\eta(b) + \int_a^b y(x)\eta(x)\,dx = 0$$

For all $\eta(x)$ that is compactly supported and smooth, then $c_1 = c_2 = 0$ and $y(x)$ is identically 0.

Proof. let the set of all $\eta(x)$ be S. Then, consider a subset $S_0 \subseteq S$ that contains all $\eta(x)$ satisfying $\eta(a) = \eta(b) = 0$.

For this subset of functions,

$$0 = c_1\eta(a) + c_2\eta(b) + \int_a^b y(x)\eta(x)\,dx = \int_a^b y(x)\eta(x)\,dx$$

And by Lemma 1.1, we have $y(x) = 0$ for all $x \in [a, b]$ for all $\eta(x) \in S_0$. Then, it must also be true that $y(x) = 0$ for all $x \in [a, b]$ for all $\eta(x) \in S$, since the only $y(x)$ we can pick from are those that already satisfy

the restricted case, which, is only the function $y(x) = 0$. Then, for all $\eta(x) \in S$, we have

$$0 = c_1\eta(a) + c_2\eta(b)$$

We can further construct subsets of S, specifically, $S_1 \in S$ with $\eta(a) = 1, \eta(b) = 0$ and $S_2 \in S$ with $\eta(a) = 0, \eta(b) = 1$ to force $c_1 = c_2 = 0$.

□

These two lemmas may feel arbitrary at first, but let's scrutinize it. For Lemma 1.1, the function that "varies" is $\eta(x)$, and we are saying that if $I[\eta] = \int_a^b y(x)\eta(x)\,dx = 0$, which is a criterion on the entire functional, we can translate it to a criterion on the function $y(x)$ which defines the functional, independent of the variation η.

I guess the reason why this is a lemma instead of a theorem is that, we can't interpret much from this result on it's own, but it's fundemental because it establishes some powerful results, which we cover in the following subsection.

1.1 Euler-Lagrange equation

Now we embark on the analysis of the stationary values of the functional $I[y]$. We might be tempted to try to vary y(x) by some infinitesimal function $'(x)$, but there are uncountably many possible functions and this can lead to many difficulties. To avoid needing to worry about these many possibilities, we instead focus on a single one-parameter family of variations. We fix a function $\eta(x)$, and consider

$$I[y + \epsilon\eta]$$

Where ϵ is an arbitarily small number. Along the function η, we want to find a y^* such that

$$I[y^*] \leq I[y^* + \epsilon\eta]$$

For all ϵ. Then, it is natural to apply the criterion for extremas from calculus, i.e., the minimum y^* must satisfy the following equation

$$\frac{d}{d\epsilon} I[y + \epsilon\eta]\bigg|_{\epsilon=0} = 0$$

This is what we call a weak formulation for the extrema, where we inspect variation along a specific path at a time, i.e., y^* is only the extrema along the slice η of the domain. It is important to note an alternative perspective, which is more useful in the setting of physics. We say that $\epsilon\eta$ is δy, the variation of y. Then, we define the first variation of the functional $I[y]$ as

$$\delta I[y] = I[y + \delta y] - I[y]$$

Another way to describe the criterion for the minimiser via this perspective is

$$\lim_{\delta y \to 0} \frac{\delta I[y]}{\delta y} = 0$$

To transform this into what we call a strong formulation, which is independent of the variation function η, we need to use Lemma 1.2 that we have previously established:

Theorem 1.3: Euler-Lagrange Equation

Let $I[y]$ be a functional defined in (1) for some smooth function F, where

$$I[y] = \int_a^b F(x, y, y')\, dx$$

The minimizers y satisfy the following:

$$\frac{\partial F}{\partial y} = \frac{d}{dx}\frac{\partial F}{\partial \dot{y}}$$

Proof. Omitted. □

In fact, we can arrive at the same criterion via an alternative proof using the first variation:

Proof. Take the Taylor Expansion of $I[y + \delta y]$ at y, we have

$$\begin{aligned} I[y + \delta y] &= \int_a^b F(x, y + \epsilon\eta, y' + \epsilon\eta')\, dx \\ &= \int_a^b F(x, y, y') + \frac{\partial F}{\partial y}\epsilon\eta + \frac{\partial F}{\partial y'}\epsilon\eta' + \mathcal{O}(\epsilon^2)dx \\ &= I[y] + \epsilon\eta \int_a^b \frac{\partial F}{\partial y} + \frac{\partial F}{\partial y'}\frac{\eta'}{\eta}dx + \mathcal{O}(\epsilon^2) \end{aligned}$$

Then, we have

$$\begin{aligned} \lim_{\delta y \to 0} \frac{\delta I[y]}{\delta y} &= \lim_{\delta y \to 0} \frac{I[y + \delta y] - I[y]}{\delta y} \\ &= \lim_{\epsilon \to 0} \int_a^b \frac{\partial F}{\partial y} + \frac{\partial F}{\partial y'}\frac{\eta'}{\eta}dx + \frac{\mathcal{O}(\epsilon^2)}{\epsilon\eta} \\ &= \int_a^b \frac{\partial F}{\partial y} + \frac{\partial F}{\partial y'}\frac{\eta'}{\eta}dx \\ &= \int_a^b \frac{\partial F}{\partial y}dx + \frac{\partial F}{\partial y'}\bigg|_a^b - \int_a^b \frac{d}{dx}\frac{\partial F}{\partial y'}dx \end{aligned}$$

Setting this equal to 0 gives us the same result to the previous proof. □

1.2 Vector valued functions

Lemma 1.4: Fundemental lemma of the calculus of variations(Vector valued functions)

let $y(x)$ be a continuous function on the interval $[a, b]$ that satisfies

$$\int_a^b \langle \mathbf{y}(x), \boldsymbol{\eta}(x) \rangle\, dx = 0$$

For all $\eta(x)$ that is compactly supported and smooth, then $y(x)$ is identically 0.

Proof. Omitted. □

Lemma 1.5: Extension of the fundamental lemma(Vector valued functions)

let $y(x)$ be a continuous function on the interval $[a, b]$ that satisfies

$$\langle \mathbf{c_1}, \boldsymbol{\eta}(a) \rangle + \langle \mathbf{c_2}, \boldsymbol{\eta}(b) \rangle + \int_a^b \langle \mathbf{y}(x), \boldsymbol{\eta}(x) \rangle \, dx \, dx = 0$$

For all $\eta(x)$ that is compactly supported and smooth, then $c_1 = c_2 = 0$ and $y(x)$ is identically 0.

Proof. Omitted. □

Theorem 1.6: Euler-Lagrange Equation(Vector valued functions)

Let $I[\mathbf{y}]$ be a functional defined in (1) for some smooth function F, where

$$I[\mathbf{y}] = \int_a^b F(x, \mathbf{y}, \mathbf{y}') \, dx$$

The minimizers $\mathbf{y}$ satisfy the following:

$$\frac{\partial F}{\partial \mathbf{y}} - \frac{d}{dx}\frac{\partial F}{\partial \mathbf{y}'} \iff \frac{\partial F}{\partial y_i} - \frac{d}{dx}\frac{\partial F}{\partial y_i'} = 0$$

Proof. First, for the vector valued function, we inspect the domain and codomain rigorously. Let $\mathcal{C}^2_{P,Q}$ be a subspace of C^2, containing curves $\mathbf{y} : [a, b] \to \mathbb{R}^n$ with $\mathbf{y}(a) = P$ and $\mathbf{y}(b) = Q$. Then, for the function $F(x, \mathbf{y}, \mathbf{y}')$, it's input space is (2n+1) dimensional, where we let $F : \mathbb{R}^{2n+1} \to \mathbb{R}$ be a sufficiently differentiable function (smooth, infinitely differenciable). Hence for the functional, we have $I[-] : \mathcal{C}^2_{P,Q} \to \mathbb{R}$.

Recall that a path $\mathbf{y}^*$ is a critical point for the functional $I[-]$ if, for all endpoint-fixed variations $\boldsymbol{\eta}$ (where $\boldsymbol{\eta}(a) = P$, $\boldsymbol{\eta}(b) = Q$), we have

$$\frac{d}{d\epsilon} I[\mathbf{y} + \epsilon\boldsymbol{\eta}]\Big|_{\epsilon=0} = 0 \tag{2}$$

Substituting in the definition for the functional, we have

$$\frac{d}{d\epsilon} \int_a^b F(x, \mathbf{y} + \epsilon\boldsymbol{\eta}, \mathbf{y}' + \epsilon\boldsymbol{\eta}') \, dx\Big|_{\epsilon=0} = 0$$
$$\int_a^b \frac{d}{d\epsilon} F(x, \mathbf{y} + \epsilon\boldsymbol{\eta}, \mathbf{y}' + \epsilon\boldsymbol{\eta}')\Big|_{\epsilon=0} dx = 0$$

Using chain rule, where

$$\begin{aligned}\frac{d}{d\epsilon}F(x,\mathbf{y}+\epsilon\boldsymbol{\eta},\mathbf{y}'+\epsilon\boldsymbol{\eta}') &= \frac{\partial F(x,\mathbf{u},\mathbf{u}')}{\partial \mathbf{u}}\frac{d\mathbf{u}}{d\epsilon}+\frac{\partial F(x,\mathbf{u},\mathbf{u}')}{\partial \mathbf{u}'}\frac{d\mathbf{u}'}{d\epsilon}\\ &= \sum_{i=1}^{n}\frac{\partial F}{\partial y_i}\eta_i+\frac{\partial F}{\partial y_i'}\eta_i'\end{aligned}$$

Next, we have the following identity from the chain rule:

$$\begin{aligned}&\frac{d}{dx}(\frac{\partial F}{\partial y_i'}\eta_i)=\frac{\partial F}{\partial y_i'}\eta_i'+\eta_i\frac{d}{dx}\frac{\partial F}{\partial y_i'}\\ \iff &\frac{\partial F}{\partial y_i'}\eta_i'=\frac{d}{dx}(\frac{\partial F}{\partial y_i'}\eta_i)-\eta_i\frac{d}{dx}\frac{\partial F}{\partial y_i'}\end{aligned}$$

This is useful in eliminating the η_i' term. Putting all of this together, we have (2) is equivalant to

$$\begin{aligned}\int_a^b\sum_{i=1}^{n}\frac{\partial F}{\partial y_i}\eta_i+\frac{\partial F}{\partial y_i'}\eta_i'\Big|_{\epsilon=0}dx=0\\ \int_a^b\sum_{i=1}^{n}\frac{\partial F}{\partial y_i}\eta_i+\frac{d}{dx}(\frac{\partial F}{\partial y_i'}\eta_i)-\eta_i\frac{d}{dx}\frac{\partial F}{\partial y_i'}\,dx=0\\ (\sum_{i=1}^{n}\frac{\partial F}{\partial y_i'}\eta_i)\Big|_a^b+\int_a^b\sum_{i=1}^{n}(\frac{\partial F}{\partial y_i}-\frac{d}{dx}\frac{\partial F}{\partial y_i'})\eta_i\,dx=0\\ \langle\frac{\partial F}{\partial \mathbf{y}'}(b),\boldsymbol{\eta}(b)\rangle+\langle-\frac{\partial F}{\partial \mathbf{y}'}(a),\boldsymbol{\eta}(a)\rangle+\int_a^b\langle\frac{\partial F}{\partial \mathbf{y}}-\frac{d}{dx}\frac{\partial F}{\partial \mathbf{y}'},\boldsymbol{\eta}\rangle\,dx=0\end{aligned}$$

And by Lemma 1.5, we arrive at

$$\frac{\partial F}{\partial \mathbf{y}}-\frac{d}{dx}\frac{\partial F}{\partial \mathbf{y}'}=0\iff\frac{\partial F}{\partial y_i}-\frac{d}{dx}\frac{\partial F}{\partial y_i'}=0$$

□

The motivation of the Euler-Lagrange Equation is to express the condition for a extrema in a strong formulation, i.e. it is free of the arbitary function η, and only depends on the choice of the path $\mathbf{y}$.

Two further corollaries follows from this theorem:

"Ignorable coordinate":

"Beltrami identity":

Another perspective

1.3 The Action and the Lagrangian

Up to this point, we have established the Euler-Lagrange Equation, and one important application of this is in expressing the energy of a physical system, and in defining equation of motion in Lagrangian mechanics.

Theorem 1.7: Hamilton's principle of least action

Let $\mathbf{q}(t) \in \mathbb{R}^n$ be a parameterized path in the generalized coordinates. we define the Action $\mathcal{S}$ as the following:

$$\mathcal{S}[\mathbf{q}(t)] = \int_{t_1}^{t_2} \mathcal{L}(t, \mathbf{q}(t), \mathbf{q}'(t))\, dt$$

Where $\mathcal{L}$ is a function that we call the Lagrangian of the system. Then, a particle follows a trajectory, i.e. a parameterised path that minimizes the action.

Using this principle, we can actually show that for a classical physical system, the Lagrangian being KE - PE satisfies the Newton's laws of motion, i.e. the definition of the Lagrangian is more fundamentally motivated by minimizing the Action.

Theorem 1.8: Lagrangian in classical mechanics

Assuming Newton's Second Law of Motion, where

$$-\frac{\partial V}{\partial \mathbf{q}} = m\mathbf{q}'' \iff -\frac{\partial V}{\partial q_i} = mq_i''$$

and define the kinetic energy as

$$T(\mathbf{q}') = \frac{1}{2} m \left\|\mathbf{q}'\right\|^2$$

We have

$$\mathcal{L}(t, \mathbf{q}, \mathbf{q}') = T(\mathbf{q}') - V(\mathbf{q})$$

Satisfying Hamilton's principle of least action.

Proof. Here we use both Theorem 1.7 and Theorem 1.6 to show that this formula for the Lagrangian is valid for classical mechanical systems.

First, to minimize the action, the path $\mathbf{q}$ must satisfy Theorem 1.6, where

$$\frac{\partial L}{\partial \mathbf{q}} - \frac{d}{dt}\frac{\partial L}{\partial \mathbf{q}'} = 0$$

Substituting in the formula for $\mathcal{L}$, we have

$$\frac{\partial L}{\partial \mathbf{q}} = -\frac{\partial V}{\partial \mathbf{q}}, \quad \frac{\partial L}{\partial \mathbf{q}'} = m\mathbf{q}'$$

Then the Euler-Lagrange Equation becomes

$$\begin{aligned} -\frac{\partial V}{\partial \mathbf{q}} - \frac{d}{dt}(m\mathbf{q}') &= 0 \\ -\frac{\partial V}{\partial \mathbf{q}} &= m\mathbf{q}'' \end{aligned}$$

Which is exactly Newton's second law. □

Action is a very fundemental physical quantity, with a dimension of energy x time. In particular, the Planck's constant is a quantum action.

2 Quantum Time Evolution

Here, we skip over some details regarding the mathematical formalism of Quantum Mechanics, and assumes some knowledge in this area, such as the state space of quantum systems. This section borrows heavily from the Berkeley 221A course notes.

Let the pure state of a system at some initial time t_0 be described by the state vector $|\psi(t_0)\rangle$, and at some later time t, to be described by $|\psi(t)\rangle$. We postulate that these two state vectors are related by a linear operator $U(t, t_0)$, where

$$|\psi(t)\rangle = U(t, t_0)\,|\psi(t_0)\rangle$$

From this definition, we have the following properties:

Theorem 2.1: Properties of the Time Evolution Operator

For the operator U, we must have

$$U(t_0, t_0) = 1$$

and to preserve probability of the state vector, it must be unitary, satisfying

$$UU^\dagger = I$$

and also that

$$U(t_2, t_1)U(t_1, t_0) = U(t_2, t_0)$$

Proof. Omitted. □

Consider an infinitesimal time evolution from time t to $t + \epsilon$, and consider the first order Taylor expansion of $U(t + \epsilon, t)$:

$$\begin{aligned} U(t + \delta t, t) &= 1 + \delta t \frac{\partial}{\partial t'} U(t', t)\bigg|_{t'=t} + \cdots \\ &= 1 - i\delta t \Omega(t) + \cdots \end{aligned}$$

Where $\Omega(t) = i\dfrac{\partial}{\partial t'} U(t', t)\bigg|_{t'=t}$ is defined in this way to be a self-adjoint/Hermitian operator. We call $\Omega(t)$ the infinitesmal generator of the time translation, because it gives the small correction that takes the state from $|\psi(t)\rangle$ to $|\psi(t + \delta t)\rangle$. From the postulates of quantum mechanics, any Unitary operator is a time evolution, and any Hermitian operator is an observable.

In the case of classical mechanics, the Hamiltonian is the generator of time translations. For any classical observable, we can use the Hamiltonian to calculate the changes of such observables over time, in the language of the Poisson bracket:

Theorem 2.2: Evolution of observable over time

Let $A(t, \mathbf{q}, \mathbf{p})$ be an observable in generalized coordinates. Then, the evolution of A is governed by

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \{A, H\}$$

where

$$\{A, H\} = \sum_i (\frac{\partial A}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial H}{\partial q_i})$$

Proof. By chain rule, the total differential of the observable is given as the following:

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \sum_i (\frac{\partial A}{\partial q_i}q_i' + \frac{\partial A}{\partial p_i}p_i')$$

By Theorem **??**, this is equivalant to

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \sum_i (\frac{\partial A}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial H}{\partial q_i}) = \frac{\partial A}{\partial t} + \{A, H\}$$

□

For instance, let $f(x, p)$ be a classical observable, a function of position and momentum. Then

Because of the role of the classical Hamiltonian as the generator of time translations, we guess that $\Omega(t)$ is closely related to the operator that we should define as the quantum Hamiltonian. We cannot have $\Omega = H$ because Ω has units of inverse time, while H has units of energy. We fix this by defining

$$H = \hbar\Omega(t)$$

2.1 Commutator and Poisson bracket

The Commutator is the equivalant of a Poisson bracket in classical mechanics.

$$AB - BA = [A, B] \mapsto i\hbar\{A, B\} = i\hbar\sum_{i=0}^{n}(\frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i}) \tag{3}$$

3 Variational principle for quantum evolution

3.1 Pertubation Theory and dynamics of closed systems

Definition 3.1: Rayleigh Quotient

For an observable A, the function

$$\mathbb{E}[-] : \mathcal{H} \to \mathbb{R} \tag{4}$$

$$\psi \mapsto \frac{\langle\psi| A |\psi\rangle}{\langle\psi|\psi\rangle} \tag{5}$$

Is the Rayleigh quotient for A.

Essentially, this is the expectation of the observable A, but we consider ψ as a smooth function on $\mathcal{H}$. Another useful perspective is to consider $\mathbb{E}_A[-]$ as a functional, and the "path" of the quantum particle is traced out by $\psi(t)$ as time evolves (hence the use of [] brakets xD).

An important lemma follows from this definition, which re-establishes the extrema condition for a functional in the language of linear algebra:

Lemma 3.2: Perpendicular condition

Given a linear subspace $\mathcal{K} \subseteq \mathcal{H}$, then the condition for extremas:

$$\frac{d}{d\epsilon}\mathbb{E}_A[\psi + \epsilon\eta]\bigg|_{\epsilon=0} = 0 \quad \forall \eta \in \mathcal{K}$$

is equivalant to

$$(A - \mathbb{E}_A[\psi])\,|\psi\rangle \in \mathcal{K}^\perp \iff \langle\phi|\, A - \mathbb{E}_A[\psi]\,|\psi\rangle = 0 \quad \forall\, |\phi\rangle \in \mathcal{K}$$

In other words, $\psi \in \mathcal{K}$ is a stationary vector/extrema for the functional $\mathbb{E}_A[-]$ if and only if it is an eigenvector of the restricted Observable $A|_{\mathcal{K}}$ with an eigenvalue of $\mathbb{E}_A[\psi]$. In practice, the key is to pick a suitable family of functions η

Proof. Omitted. □

For static problems, we use the following method to find the ground state energy. Because the size of the Hilbert space grows exponentially to the system size, it is in general computationally hard to brutal forcely search the whole Hilbert space. Instead, the following method only searches states from a subset of the whole Hilbert space to find an approximate solution.

Theorem 3.3: Rayleigh-Ritz method

Consider a self-adjoint Hamiltonian H acting on a Hilbert space $\mathcal{H}$, with the following family of energy eigenstates:

$$E_0 \leq E_1 \leq \cdots E_k \leq \cdots$$

Here, we assume that H is a linear combination of tensor products of local operators σ_j with coefficients h_j, where $H = \sum_j \sigma_j h_j$. The ground state E_0 is the solution to the following problem:

$$E_0 = \min_{\forall |\psi\rangle \in \mathcal{H}} \mathbb{E}_H[\psi]$$

We choose a k-dimensional subspace $\mathcal{K} \subseteq \mathcal{H}$ (A trial subspace), and now consider the Hamiltonian restricted to this subspace, $H|_{\mathcal{K}}$. Then, we find an approximation to the ground state by solving the following problem:

$$E_0 \leq E_0^{\mathcal{K}} = \min_{\forall |\psi\rangle \in \mathcal{K}} \mathbb{E}_{H|_{\mathcal{K}}}[\psi]$$

Then, finding $E_0^{\mathcal{K}}$ is equivalent to finding the minimal λ that satisfy

$$\det(H - \lambda S) = 0$$

Where H and S are defined as the following

$$H_{i,j} = \langle \phi_i | H | \phi_j \rangle$$
$$S_{i,j} = \langle \phi_i | \phi_i \rangle$$

Proof. For the subspace $\mathcal{K}$, we can further consider a basis $\{|\phi_0\rangle, \cdots |\phi_{k-1}\rangle\}$, then for every $\phi \in \mathcal{K}$, we can write it as $\phi(\boldsymbol{\theta}) = \sum_{i=0}^{k-1} \theta_i \phi_i$, where $\boldsymbol{\theta} = (\theta_0, \cdots \theta_{k-1})$.

Then, we can rewrite the minimum energy state condition as

$$\begin{aligned} E_0 \leq E_0^{\mathcal{K}} &= \min_{\forall \boldsymbol{\theta}} \mathbb{E}_{H|_{\mathcal{K}}}[\phi(\boldsymbol{\theta})] \\ &= \min_{\forall \boldsymbol{\theta}} \frac{\langle \phi(\boldsymbol{\theta}) | H|_{\mathcal{K}} | \phi(\boldsymbol{\theta}) \rangle}{\langle \phi(\boldsymbol{\theta}) | \phi(\boldsymbol{\theta}) \rangle} \\ = \min_{\forall \boldsymbol{\theta}} \frac{\sum_{i,j} \bar{\theta_i} \theta_j \langle \phi_i | H|_{\mathcal{K}} | \phi_j \rangle}{\sum_{i,j} \bar{\theta_i} \theta_j \langle \phi_i | \phi_i \rangle} \end{aligned}$$

Next, we conduct the first derivative test, where we calculate the partial derivative of $\mathbb{E}_{H|_{\mathcal{K}}}[\phi(\boldsymbol{\theta})]$ with respect to $\bar{\theta_i}$ and set that derivative to 0. This is the normal criteria for an extrema, where we want to find a parameter $\boldsymbol{\theta}$ that brings the estimated energy as close to the actual energy as possible.

$$
\begin{aligned}
0 = \frac{\partial \mathbb{E}_{H|_{\mathcal{K}}}[\phi(\boldsymbol{\theta})]}{\theta_i} &= \frac{(\sum_{j=0}^{k-1} \theta_j \langle \phi_i | H|_{\mathcal{K}} | \phi_j \rangle)(\sum_{i,j} \bar{\theta}_i \theta_j \langle \phi_i | \phi_i \rangle) - (\sum_{j=0}^{k-1} \theta_j \langle \phi_i | \phi_i \rangle)(\sum_{i,j} \bar{\theta}_i \theta_j \langle \phi_i | H|_{\mathcal{K}} | \phi_j \rangle)}{(\sum_{i,j} \bar{\theta}_i \theta_j \langle \phi_i | \phi_i \rangle)^2} \\
&\iff (\sum_{j=0}^{k-1} \theta_j \langle \phi_i | H|_{\mathcal{K}} | \phi_j \rangle)(\sum_{i,j} \bar{\theta}_i \theta_j \langle \phi_i | \phi_i \rangle) = (\sum_{j=0}^{k-1} \theta_j \langle \phi_i | \phi_i \rangle)(\sum_{i,j} \bar{\theta}_i \theta_j \langle \phi_i | H|_{\mathcal{K}} | \phi_j \rangle) \\
&\iff \sum_{j=0}^{k-1} \theta_j \langle \phi_i | H|_{\mathcal{K}} | \phi_j \rangle = \mathbb{E}_{H|_{\mathcal{K}}}[\phi(\boldsymbol{\theta})] \sum_{j=0}^{k-1} \theta_j \langle \phi_i | \phi_i \rangle \ \ \forall 0 \leq j \leq k-1
\end{aligned}
$$

Now we define H and S as above, we get

$$
\begin{aligned}
\langle H_{i,}, \boldsymbol{\theta} \rangle &= \langle \mathbb{E}_{H|_{\mathcal{K}}}[\phi(\boldsymbol{\theta})] S_{i,}, \boldsymbol{\theta} \rangle \ \ \forall 0 \leq j \leq k-1 \\
\iff \langle H_{i,} - \mathbb{E}_{H|_{\mathcal{K}}}[\phi(\boldsymbol{\theta})] S_{i,}, \boldsymbol{\theta} \rangle &= 0 \ \ \forall 0 \leq j \leq k-1 \\
\iff (H - \mathbb{E}_{H|_{\mathcal{K}}}[\phi(\boldsymbol{\theta})] S) \boldsymbol{\theta} &= 0
\end{aligned}
$$

Since for non-trivial answers we require $\boldsymbol{\theta}$ to be non-zero, we would require the matrix $(H - \mathbb{E}_{H|_{\mathcal{K}}}[\phi(\boldsymbol{\theta})] S)$ to have a null space with dimension greater than 0, where

$$
\det(H - \mathbb{E}_{H|_{\mathcal{K}}}[\phi(\boldsymbol{\theta})] S) = 0
$$

□

The idea is to

This can only be solved by some classical optimization algorithm.

Theorem 3.4: Approximating time evolution with parameter variation

To approximate the evolution of a pure state $|\psi(t)\rangle \in \mathcal{H}$ that evolves according to the Schrodinger's equation using a parameterised trial state $|\phi(\boldsymbol{\theta}(t))\rangle \in \mathcal{K}$ with $\dim \mathcal{K} = k$, i.e. to solve

$$
|\psi(t)\rangle \approx |\phi(\boldsymbol{\theta}(t))\rangle
$$

over time, it is equivalent to solving the following target problem

$$
\sum_{j=0}^{k-1} \frac{\partial |\phi(\boldsymbol{\theta}(t))\rangle}{\partial \theta_j} \delta\theta_j \approx -i\delta t H |\phi(\boldsymbol{\theta}(t))\rangle \tag{6}
$$

Proof. The derivation of the new statement comes from comparing the variation of $|\phi(\boldsymbol{\theta}(t))\rangle$ from first according to Schrodinger's equation, and second by the variation of the term $\boldsymbol{\theta}(t)$ (i.e. we are saying that when we vary the parameter $\boldsymbol{\theta}(t)$, the trial state should also abide Schrodinger's equation). According to Schrodinger's equation, where

$$
\frac{d |\psi(t)\rangle}{dt} = -iH |\psi(t)\rangle
$$

We have the following first order approximation:

$$|\phi(\boldsymbol{\theta}(t+\delta t)\rangle \approx |\phi(\boldsymbol{\theta}(t))\rangle - i\delta t H\,|\phi(\boldsymbol{\theta}(t))\rangle$$

And by simply consider the first order variation, we also have

$$|\phi(\boldsymbol{\theta}(t+\delta t)\rangle \approx |\phi(\boldsymbol{\theta}(t))\rangle + \sum_{i=0}^{k-1} \frac{\partial\,|\phi(\boldsymbol{\theta}(t))\rangle}{\partial\theta_i}\delta\theta_i$$

Comparing the two expressions, we get exactly equation (6)

□

3.2 Dirac and Frenkel Variational principle

We start with discussing the real evolution of a pure state, before generalizing to mixed states and also imaginary time evolution.

Theorem 3.5: Real time evolution

By projecting (6) onto the tangent subspace $\mathbf{Span}\{\frac{\partial\,|\phi(\boldsymbol{\theta}(t))\rangle}{\partial\theta_j}\}$, the evolution of the parameter $\boldsymbol{\theta}(t)$ is obtained by solving the following system of equations:

$$A(t)\boldsymbol{\theta}(t)' = -i\boldsymbol{C}(t) \iff \sum_{j=0}^{k-1} A_{i,j}(t)\theta_j(t)' = -iC_i(t) \quad \forall 0 \le i \le k-1$$

where

$$A_{i,j}(t) = \frac{\partial\,\langle\phi(\boldsymbol{\theta}(t))|}{\partial\theta_i}\frac{\partial\,|\phi(\boldsymbol{\theta}(t))\rangle}{\partial\theta_j} \tag{7}$$

$$C_i(t) = \frac{\partial\,\langle\phi(\boldsymbol{\theta}(t))|}{\partial\theta_i} H\,|\phi(\boldsymbol{\theta}(t))\rangle \tag{8}$$

This is equivalant to the following equality

$$\langle\delta\phi(\boldsymbol{\theta}(t))|\,\frac{d}{dt} + iH\,|\phi(\boldsymbol{\theta}(t))\rangle = 0 \tag{9}$$

where $\langle\delta\phi(\boldsymbol{\theta}(t))|$ is defined as the adjoint of the variation along all direction of the tangent subspace, namely

$$\sum_{i=0}^{k-1} \frac{\partial\,\langle\phi(\boldsymbol{\theta}(t))|}{\partial\theta_i}\delta\theta_i$$

Proof. Consider the projector

$$P = \sum_{i=0}^{k-1} \frac{\partial\,|\phi(\boldsymbol{\theta}(t))\rangle}{\partial\theta_i}\frac{\partial\,\langle\phi(\boldsymbol{\theta}(t))|}{\partial\theta_i}$$

Apply to both side of (6), where

$$\sum_{j=0}^{k-1} P\frac{\partial\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{\partial\theta_j}\delta\theta_j \approx -i\delta t PH\left|\phi(\boldsymbol{\theta}(t))\right\rangle$$

$$\iff \sum_{j=0}^{k-1}\sum_{i=0}^{k-1}(\frac{\partial\left\langle\phi(\boldsymbol{\theta}(t))\right|}{\partial\theta_i}\frac{\partial\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{\partial\theta_j}\delta\theta_j)\frac{\partial\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{\partial\theta_i} \approx -i\delta t\sum_{i=0}^{k-1}\left(\frac{\partial\left\langle\phi(\boldsymbol{\theta}(t))\right|}{\partial\theta_i}H\left|\phi(\boldsymbol{\theta}(t))\right\rangle\right)\frac{\partial\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{\partial\theta_i}$$

$$\iff \sum_{i=0}^{k-1}\left(\sum_{j=0}^{k-1}(\frac{\partial\left\langle\phi(\boldsymbol{\theta}(t))\right|}{\partial\theta_i}\frac{\partial\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{\partial\theta_j}\delta\theta_j)\right)\frac{\partial\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{\partial\theta_i} \approx \sum_{i=0}^{k-1}\left(-i\delta t\frac{\partial\left\langle\phi(\boldsymbol{\theta}(t))\right|}{\partial\theta_i}H\left|\phi(\boldsymbol{\theta}(t))\right\rangle\right)\frac{\partial\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{\partial\theta_i}$$

Then for each $\frac{\partial\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{\partial\theta_j}$, by comparing coefficients, we have

$$\sum_{j=0}^{k-1}\frac{\partial\left\langle\phi(\boldsymbol{\theta}(t))\right|}{\partial\theta_i}\frac{\partial\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{\partial\theta_j}\frac{\delta\theta_j}{\delta t} \approx -i\frac{\partial\left\langle\phi(\boldsymbol{\theta}(t))\right|}{\partial\theta_i}H\left|\phi(\boldsymbol{\theta}(t))\right\rangle \quad \forall 0\le i\le k-1 \tag{10}$$

And take $\lim\delta t\to 0$ of (10), to get $\frac{\delta\theta_j}{\delta t}\to\frac{d\theta_j}{dt}$ we arrive precisely at

$$A(t)\boldsymbol{\theta}(t)' = -i\boldsymbol{C}(t) \iff \sum_{j=0}^{k-1}A_{i,j}(t)\theta_j(t)' = -iC_i(t) \quad \forall 0\le i\le k-1$$

After taking the limit, also notice by chain rule, we have

$$\sum_{j=0}^{k-1}\frac{\partial\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{\partial\theta_j}\frac{d\theta_j}{dt} = \frac{d\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{dt}$$

Which again from (10), gives us

$$\frac{\partial\left\langle\phi(\boldsymbol{\theta}(t))\right|}{\partial\theta_i}\frac{d\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{dt} = -i\frac{\partial\left\langle\phi(\boldsymbol{\theta}(t))\right|}{\partial\theta_i}H\left|\phi(\boldsymbol{\theta}(t))\right\rangle \quad \forall 0\le i\le k-1$$

$$\iff \left\langle\delta\phi(\boldsymbol{\theta}(t))\right|\frac{d\left|\phi(\boldsymbol{\theta}(t))\right\rangle}{dt} = -i\left\langle\delta\phi(\boldsymbol{\theta}(t))\right|H\left|\phi(\boldsymbol{\theta}(t))\right\rangle \iff \left\langle\delta\phi(\boldsymbol{\theta}(t))\right|\frac{d}{dt}+iH\left|\phi(\boldsymbol{\theta}(t))\right\rangle = 0$$

□

To implement the evolution, we essentially consider $\boldsymbol{\theta}(t)'$ as $\frac{\delta\boldsymbol{\theta}(t)}{\delta t}$, which gives us the following iteration:

$$\boldsymbol{\theta}(t+\delta t)\approx\boldsymbol{\theta}(t)+\delta\boldsymbol{\theta}(t) = \boldsymbol{\theta}(t)-i\delta tA^{-1}(t)C(t) \tag{11}$$

Also notice that from definitions (7) and (8), the matrix $A(t)$ only depends on the choice of $\boldsymbol{\theta}(t)$, which we call "ansatz-dependent", since ansatz refers to our choice of parameterization. On the other hand, the vector $C(t)$ further depends on the Hamiltonian H, which we call "generator-dependent", since the Hamiltonian is considered the generator of time translations.

Another interesting observation we can make, is that for the Schrodinger Equation with respect to the trial state vector, the Lagrangian of the same system is described as:

$$\mathcal{L} = \left\langle\phi(\boldsymbol{\theta}(t))\right|\frac{d}{dt}+iH\left|\phi(\boldsymbol{\theta}(t))\right\rangle$$

This makes it much easier to interpret (9), which essentially means the variation in the trial space must be physical, in the sense that they don't change the value of the Lagrangian.

If we compare the iteration formula (11) to the definition of time evolution in Section 2, we find a striking similarity.

3.3 McLachlan's variational principle

Theorem 3.6: Real time evolution

Minimizing the distance between LHS and RHS of (6), the target problem is reformulated as

$$\delta I[\boldsymbol{\theta}(\boldsymbol{t})] = \delta \left\| (\frac{d}{dt} + iH) \left| \phi(\boldsymbol{\theta}(t)) \right\rangle \right\| = 0$$

Where the δ means the first variation of the functional. Then, the evolution of the parameter $\boldsymbol{\theta}(t)$ is obtained by solving the following system of equations:

$$\mathfrak{Re}(A)\boldsymbol{\theta}' = \mathfrak{Im}(\boldsymbol{C}) \iff \sum_{j=0}^{k-1} \mathfrak{Re}(A_{i,j})\theta'_j = \mathfrak{Im}(C_i) \quad \forall 0 \leq i \leq k-1$$

Where $A_{i,j}$ and C_i are as defined in (7).

Proof. We first express the norm in terms of the inner product in terms of the basis of the trial space, where

$$\left\| (\frac{d}{dt} + iH) \left| \phi(\boldsymbol{\theta}(t)) \right\rangle \right\| = \left((\frac{d}{dt} + iH) \left| \phi(\boldsymbol{\theta}(t)) \right\rangle \right)^\dagger \left((\frac{d}{dt} + iH) \left| \phi(\boldsymbol{\theta}(t)) \right\rangle \right)$$

Since H is self-adjoint because it is an observable, this means $(iH)^\dagger = (-iH)$. Furthermore, the differential operator $\dfrac{d}{dt}$ acting on $|\phi(\boldsymbol{\theta}(t))\rangle$ can be expressed in terms of the partial derivatives, where

$$\frac{d}{dt} \left| \phi(\boldsymbol{\theta}(t)) \right\rangle = \sum_{i=0}^{k-1} \frac{\partial \left| \phi(\boldsymbol{\theta}(t)) \right\rangle}{\partial \theta_i} \theta'_i$$

□

The difference is, in the first method we allow some difference Δ between the LHS and RHS, i.e.

$$\sum_{j=0}^{k-1} \frac{\partial \left| \phi(\boldsymbol{\theta}(t)) \right\rangle}{\partial \theta_j} \delta\theta_j = -i\delta t H \left| \phi(\boldsymbol{\theta}(t)) \right\rangle + \Delta$$

and projects that Δ onto the trial space. This method differs in the way that, we are actively try to minimize that difference Δ during our evolution.

Equivalence of anti-Hermitian Hamiltonian and evolution in imaginary time (non unitary evolution).

3.4 Ansatz design from Hamiltonian Symmetry

The Black-Scholes Equation can be expressed in the form of the Schrodinger equation, with the following Hamiltonian:

$$\hat{H}_{BS} = i(\frac{1}{2}(\sigma x)^2 \frac{\partial^2}{\partial x^2} + rx\frac{\partial}{\partial x} - r)$$

It is easy to verify that this Hamiltonian is non-Hermitian, which induces non-unitary time evolution. Hence the approach of a variational algorithm.

As an example, consider the following non-linear Black Scholes equation, taking into account of market liquidity:

$$\frac{\partial u}{\partial t} + \frac{\sigma^2 S^2}{2}\left(\frac{\frac{\partial^2 u}{\partial S^2}}{1 - \rho\lambda(S)\frac{\partial^2 u}{\partial S^2}}\right) = 0$$

Consider the two dimensional space X of the independent variables $(S, t) \in X$ and a one dimensional space of dependent variables $u \in U$. We denote by $M = X \times U$ the base space of the differential equation we are investigating.

We further consider $U_{(1)}$ the space of first order derivatives, where

$$(\frac{\partial u}{\partial S}, \frac{\partial u}{\partial t}) \in U_{(1)}$$

and $U_{(2)}$ the space of second order derivatives, where

$$(\frac{\partial^2 u}{\partial S^2}, \frac{\partial^2 u}{\partial t^2}, \frac{\partial^2 u}{\partial S \partial t}) \in U_{(2)}$$

To represent the differential equation, we introduce the following space, called a second order jet bundle $M^{(2)}$ of the base space M:

$$M^{(2)} = M \times U_{(1)} \times U_{(2)}$$

With a natural contact structure. We label the coordinates in this jet bundle by $w = (S, t, u, u_S, u_t, u_{SS}, u_{tt}, u_{st})$.

Then, the Black-Scholes equation has a solution manifold L_Δ living in $M^{(2)}$, where

$$L_\Delta = \{w \in M^{(2)} \mid \Delta(w) = 0\}$$

with

$$\Delta(S, t, u, u_S, u_t, u_{SS}, u_{tt}, u_{st}) = u_t + \frac{\sigma^2 S^2}{2}\left(\frac{u_{SS}}{1 - \rho\lambda(S)u_{SS}}\right)$$

By the prolongation formula and infinitesimal symmetry criterion, we are able to explicitly compute the symmtry group of systems of differential equations. Using the package of DifferentialGeometry and JetCalculus in Maple, I would like to extract the symmetry of the Black Scholes equation with stochastic volatility, under the Hull-White model. For an European Option with Volatility $\sqrt{y_t}$, and an underlying stock x, we have two stochastic processes:

$$dx_t = \mu x_t dt + \sqrt{y_t} x_t dW_1$$

$$dy_t = aydt + bydW_2$$

Where the correlation between the two Wiener process dW_1 and dW_2 is given by ρ. Let $V(x, y, t)$ be the value of the option on the underlying x at time t. Then, using portfolio replication with Ito's lemma, coupled with non-arbitrage arguments gives the following PDE:

$$\frac{\partial u}{\partial t} + \left(rx\frac{\partial}{\partial x} + \frac{1}{2}x^2 y\frac{\partial^2}{\partial x^2} + b\rho x y^{\frac{3}{2}}\frac{\partial^2}{\partial x \partial y} + \frac{1}{2}b^2 y^2\frac{\partial^2}{\partial y^2} + ay\frac{\partial}{\partial y} - r \right) u = 0$$

To explore the symmetry of the PDE, we first inspect it's solution manifold.

To find the solution of the PDE using VarQITE, we can express this in a Schrodinger-type equation. First, we write

$$\frac{\partial u}{\partial t} = \mathfrak{G} u$$

With the infinitesimal generator

$$\mathfrak{G} = -\left(rx\frac{\partial}{\partial x} + \frac{1}{2}x^2 y\frac{\partial^2}{\partial x^2} + b\rho x y^{\frac{3}{2}}\frac{\partial^2}{\partial x \partial y} + \frac{1}{2}b^2 y^2\frac{\partial^2}{\partial y^2} + ay\frac{\partial}{\partial y} - r \right)$$

We consider the simplist case where a, b, r have no dependence on x, y, t and treat them as constants. For a numerical experiment, we set $a = b = r = 1$ where $\mathfrak{G}$ becomes

$$\mathfrak{G} = -\left(x\frac{\partial}{\partial x} + \frac{1}{2}x^2 y\frac{\partial^2}{\partial x^2} + \rho x y^{\frac{3}{2}}\frac{\partial^2}{\partial x \partial y} + \frac{1}{2}y^2\frac{\partial^2}{\partial y^2} + y\frac{\partial}{\partial y} - 1 \right)$$

The first step is to apply a change of variable $u = e^{g(x,y,t)v(x,y,t)}$ to simplify the generator, and also to apply a wick rotation.

With the substitution, we get

The next step is to discretize $\mathfrak{G}$ with periodic boundary conditions, and decompose $\mathfrak{G}$ into unitary operations.

Consider a 16 by 16 grid of qubit states, achieved by using 8 qubits. We embed the solution u_{ij} into this grid, where the value of u_{ji} corresponds to the probability amplitude of the qubit state it is located at.

To discretize the generator, we first express the first order derivatives via the following approximation:

$$\frac{\partial u}{\partial x} \approx \frac{u_{i+1,j} - u_{ij}}{\Delta x}$$
$$\frac{\partial u}{\partial x} \approx \frac{u_{i,j+1} - u_{ij}}{\Delta y}$$

And for second order derivatives, we have

$$\frac{\partial^2 u}{\partial x^2} \approx \frac{u_{i+1,j} - 2u_{ij} + u_{i-1,j}}{\Delta x^2}$$
$$\frac{\partial^2 u}{\partial x^2} \approx \frac{u_{i,j+1} - 2u_{ij} + u_{i,j-1}}{\Delta y^2}$$
$$\frac{\partial^2 u}{\partial x \partial y} \approx \frac{u_{i+1,j+1} - u_{i+1,j-1} - u_{i-1,j+1} + u_{i-1,j-1}}{4\Delta x \Delta y}$$

Then, $\mathfrak{G}$ can be discretised as the following:

$$[\mathfrak{G}]_{i,j} =$$

For the ansatz $|\phi(\boldsymbol{\theta})\rangle$, we initialise it by finding the parameter θ_0 that brings this trial wavefunction as close to the original wavefunction we are starting with, where

$$\boldsymbol{\theta_0} = \arg\min_{\theta \in \mathbf{R}^n} \| |\phi(\boldsymbol{\theta})\rangle - |\psi(0)\rangle \|$$

This process is done classically, via some variant of gradient descent.

The ansatz $|\phi(\boldsymbol{\theta})\rangle$ is prep Difficulty I am encountering: - How to translate the information of the lie symmetry to the ansatz? The ansatz is only composed of rotation and controlled rotation on each qubit, while the qubit themselves are encoding the information of the differential equation with a low resolution.