FYS3150 Computational Physics - Project 5
Variational Monte Carlo

Nils Johannes Mikkelsen
(Dated: December 2, 2018)

I. INTRODUCTION

II. THEORY

A. Quantum Mechanics - Miscellaneous

The reader is assumed to be familiar with most of the basic features of quantum mechanics such as Dirac notation, the notion of kets and Hilbert spaces, quantum spin, composite systems, etc. Nonetheless, the following sections introduce some of the basic aspects of quantum mechanics that are readily applied and studied in this project.

1. Notation

First and foremost, kets and bras are depicted as usual, e.g. $|\psi\rangle$ and $\langle\phi|$. The hermiation conjugate of some object c is denoted by a dagger: $c^\dagger$. In general, operators are denoted by upper case letters and constants by lower case letters, although exceptions such as the canonical operators and ladder operators do exist. Furthermore, operators are denoted by an additional "hat" in the absence of a particular basis. For example, the canonical operators $\hat{\mathbf{r}}$ and $\hat{\mathbf{p}}$ expressed in the position basis $|\mathbf{r}\rangle$, are given by:

$$\mathbf{r} \quad \text{and} \quad \mathbf{p} = -i\hbar\nabla \tag{1}$$

Moreover, the single-system Hamiltonian for a system with mass m in a potential V:

$$\hat{\mathcal{H}} = \frac{\hat{\mathbf{p}}^2}{2m} + V(\hat{\mathbf{r}}, \hat{\mathbf{p}}) \tag{2a}$$

thus takes the following form in the position basis:

$$\mathcal{H} = \frac{-\hbar^2}{2m}\nabla^2 + V(\mathbf{r}, -i\hbar\nabla) \tag{2b}$$

Unless otherwise specified, all operators are expressed in the continuous position basis. Provided a classical operator O, the corresponding quantum operator $\hat{O}$ is found by inserting the canonical operators: $\hat{O} = O(\hat{\mathbf{r}}, \hat{\mathbf{p}})$.

Composite systems are denoted either as a tensor product, say $|\psi\rangle\otimes|\phi\rangle$, or using some form of compact notation

$|\psi, \phi\rangle$, depending on the system in question. This report will favour a compact notation in cases where such notation is natural. Each component of a composite system is considered a degree of freedom, be it different spatial degrees of freedom of the same system or completely unrelated degrees of freedom (e.g. an electron's position and spin). The Hamiltonian of a composite system with N degrees of freedom is equal to the sum of the individual Hamiltonians for each component plus the interaction:

$$\hat{\mathcal{H}}_{\text{composite}} = \sum_{i=1}^{N} \hat{\mathcal{H}}_i + \hat{\mathcal{H}}_{\text{interaction}} \tag{3}$$

where $\hat{\mathcal{H}}_{\text{interaction}}$ encompases all interactions between the N degrees of freedom.

For infinite-dimensional Hilbert spaces, the inner product between $|\psi\rangle$ and $|\phi\rangle$ is

$$\langle\psi|\phi\rangle = \int \mathrm{d}^3\mathbf{r}\, \psi(\mathbf{r})^*\phi(\mathbf{r}) \tag{4}$$

Similarly, the expected value of some operator $\hat{A}$ for a system in a state $|\psi\rangle$ is

$$\langle\psi|\hat{A}|\psi\rangle = \int \mathrm{d}^3\mathbf{r}\, \psi(\mathbf{r})^*\hat{A}\psi(\mathbf{r}) \tag{5}$$

In case $|\psi\rangle$ is an eigenket of $\hat{A}$ with corresponding eigenvalue a, (5) is simplified to

$$\langle\psi|\hat{A}|\psi\rangle = \langle\psi|a|\psi\rangle = a\langle\psi|\psi\rangle = a \tag{6}$$

assuming $|\psi\rangle$ is normalised. Expanding the integral in (6) one finds that the wave functions simplify to $|\psi(\mathbf{r})|^2$: the probability density of finding the system at $\mathbf{r}$.

2. The variational principle & method

The variational principle (not to be confused with the variational method, see below) is a simple statement that is true for all quantum systems. It states that the expected energy of a quantum system in a state $|\psi\rangle$ is greater than or equal to the ground state energy:

$$E_g \leq \frac{\langle\psi|\hat{\mathcal{H}}|\psi\rangle}{\langle\psi|\psi\rangle} \tag{7}$$

Here $|\psi\rangle$ is not necessarily normalized (which is way the expected energy is divided by the norm). The equation becomes an equality only when $|\psi\rangle$ is the ground state.

The variational principle is the basis for the *variational method*, which exploits the simplicity of the varitional principle in order to approximate the ground state energy. The primary problem is to study some Hamiltonian $\hat{\mathcal{H}}$ whose energy eigenkets are unknown. The idea is to introduce a so-called *trial state* $|\psi_T(\alpha)\rangle$, which is dependent on some parameter α (where α is not necessarily scalar). If chosen correctly, the trial state may be able to reproduce the functional shape of the ground state, thus providing an approximate upper bound E_T of the ground state energy. The variational method is actually an umbrella term for several methods, however, their underlying idea is more or less the same. The process is most easily shown as an algorithm:

Algorithm 1 The Varitational Method

1: Suggest a trial state $|\psi_T(\alpha)\rangle$.
2: Compute the trial energy E_T as a function of α:

$$E_T(\alpha) = \frac{\langle\psi_T(\alpha)|\,\hat{\mathcal{H}}\,|\psi_T(\alpha)\rangle}{\langle\psi_T(\alpha)|\psi_T(\alpha)\rangle}$$

3: Minimise E_T with respect to α.
4: Choose the optimal $|\psi_T(\alpha)\rangle$ based on the α that minimises E_T.
  The resulting optimal E_T is an upper bound on the ground state energy, while the corresponding trial state is the optimal approximation of the ground state.

Due to computational aspects, a common variant of the algorithm is to have the process of minimsation of E_T be integrated in a loop over different parameters α.

3. The Coulomb interaction

An interesting, although somewhat discouraging, reality is that one of the fundamental interactions of nature, the Coulomb interaction, is more or less analytically unsolvable for everything but the simplest of systems. The Hydrogen is one such system in which the interaction is perfectly solvable. On the other hand, the second simplest element, Helium, is not. The problem arises from the electron-electron interaction:

$$V_{e^-e^-} = \frac{e^2}{4\pi\varepsilon_0}\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} = \frac{k_e e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} \tag{8}$$

where ε_0 is the permittivity of free space, $k_e = 1/4\pi\varepsilon_0$ is Coulomb's constant, and $\mathbf{r}_1$ and $\mathbf{r}_2$ are the positions of electron 1 and 2. In general, given N particles of change q_i, the total electric potential energy is

$$V_C = \frac{k_e}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{9}$$

where the 1/2 factor ensures that the potential is equally shared between the symmetric q_i-q_j and q_j-q_i interactions. An equivalent and more computationally friendly

way of expressing (9) is

$$V_C = k_e\sum_{j=2}^{N}\sum_{i=1}^{j-1}\frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{10}$$

B. Quantum Mechanics - The Harmonic Oscillator

The harmonic oscillator is one of the finest achivements of quantum mechanics with applications in many areas of physics and other sciences. The following section will present the algebraic solution to the harmonic oscillator Hamiltonian, and generalise the results to an N-dimensional system of oscillators.

1. The single harmonic oscillator

The potential energy of a single-degree classical harmonic oscillator with oscillation frequency ω is $V = \frac{1}{2}\omega^2 x^2$, it follows that the Hamiltonian for a single-degree quantum harmonic oscillator with mass m is

$$\hat{\mathcal{H}}_{\text{HO}} = \frac{\hat{p}^2}{2m} + \frac{1}{2}\omega^2\hat{x} \tag{11}$$

where $\hat{p} = \hat{\mathbf{p}}\cdot\mathbf{e}_x$ is the system's momentum operator. The problem at hand is to find the energy eigenkets of the Hamiltonian. This will be done using the so-called *ladder operators*:[1]

$$\hat{a} = \frac{1}{\sqrt{2\hbar m\omega}}\big(+i\hat{p} + m\omega\hat{x}\big) \tag{12a}$$

$$\hat{a}^\dagger = \frac{1}{\sqrt{2\hbar m\omega}}\big(-i\hat{p} + m\omega\hat{x}\big) \tag{12b}$$

One can easily show that the ladder operators satisfy the commutation relation $[\hat{a}, \hat{a}^\dagger] = 1$. Rewriting $\hat{x}$ and $\hat{p}$ in terms of $\hat{a}$ and $\hat{a}^\dagger$, one finds that the Hamiltonian can be written as

$$\hat{\mathcal{H}}_{\text{HO}} = \hbar\omega\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right) = \hbar\omega\left(\hat{a}\hat{a}^\dagger - \frac{1}{2}\right) \tag{13}$$

It turns out that the energy eigenkets can be uniquely labelled using a single non-negative integer index n: $|n\rangle$. Furthermore, the action of ladder operators on the energy eigenkets raises or lowers n as follows:

$$\hat{a}\,|n\rangle = \sqrt{n}\,|n-1\rangle \quad\text{and}\quad \hat{a}^\dagger\,|n\rangle = \sqrt{n+1}\,|n+1\rangle \tag{14}$$

---

[1] These operators have several names, e.g. *raising/lowering* operators, *creation/annahilation* operators, etc.

Note that $n < 0$ is avoided as $\hat{a}\ket{0} = 0$.[2] It follows that $\hat{a}^\dagger \hat{a}\ket{n} = n\ket{n}$ such that

$$\hat{\mathcal{H}}_{\text{HO}}\ket{n} = \hbar\omega\left(n + \frac{1}{2}\right)\ket{n} = E_n\ket{n} \qquad (15)$$

The energy eigenkets can be uniquely described in terms of $\ket{0}$ and $\hat{a}^\dagger$ via

$$\ket{n} = \frac{\left(\hat{a}^\dagger\right)^n}{\sqrt{n!}}\ket{0} \qquad (16)$$

where the factorial accounts for the $\sqrt{n+1}$ scaling in (14).

2. The virial theorem

The virial theorem is an important result in both classical and quantum mechanics, a simple version of the virial theorem states that the expected total kinetic energy $\langle K \rangle$ is proportional to the expected potential energy $\langle V \rangle$. Peculiar to the harmonic oscillator, provided an energy eigenstate, the expected total kinetic energy is actually equal to the expected potential energy.

Suppose a quantum system subject to a harmonic potential is in the energy eigenket $\ket{n}$. The expected kinetic and potential energies are thus

$$\langle K \rangle_n = \bra{n}\hat{K}\ket{n} = \bra{n}\frac{\hat{p}^2}{2m}\ket{n} \qquad (17a)$$

$$\langle V \rangle_n = \bra{n}\hat{V}\ket{n} = \bra{n}\frac{1}{2}m\omega^2\hat{x}^2\ket{n} \qquad (17b)$$

Rewriting $\hat{x}$ and $\hat{p}$ in terms of $\hat{a}$ and $\hat{a}^\dagger$ results in:

$$\hat{x}^2 = \frac{\hbar}{2m\omega}\left(\hat{a}^\dagger + \hat{a}\right)^2 \quad \text{and} \quad \hat{p}^2 = -\frac{m\hbar\omega}{2}\left(\hat{a}^\dagger - \hat{a}\right)^2$$

The ladder operator polynomials are:

$$\left(\hat{a}^\dagger + \hat{a}\right)^2 = \hat{a}^\dagger\hat{a}^\dagger + \hat{a}^\dagger\hat{a} + \hat{a}\hat{a}^\dagger + \hat{a}\hat{a}$$
$$= \hat{a}^\dagger\hat{a}^\dagger + \hat{a}\hat{a} + 2\hat{a}^\dagger\hat{a} + 1$$

$$\left(\hat{a}^\dagger - \hat{a}\right)^2 = \hat{a}^\dagger\hat{a}^\dagger - \hat{a}^\dagger\hat{a} - \hat{a}\hat{a}^\dagger + \hat{a}\hat{a}$$
$$= \hat{a}^\dagger\hat{a}^\dagger + \hat{a}\hat{a} - 2\hat{a}^\dagger\hat{a} - 1$$

Hence,

$$\hat{K}\ket{n} = -\frac{\hbar\omega}{4}\left[\sqrt{(n+1)(n+2)}\ket{n+2} + \sqrt{n(n-1)}\ket{n-2} - 2n\ket{n} - \ket{n}\right] \qquad (18a)$$

$$\hat{V}\ket{n} = +\frac{\hbar\omega}{4}\left[\sqrt{(n+1)(n+2)}\ket{n+2} + \sqrt{n(n-1)}\ket{n-2} + 2n\ket{n} + \ket{n}\right] \qquad (18b)$$

It follows then that

$$\langle K \rangle_n = \langle V \rangle_n = \frac{E_n}{2} = \frac{\hbar\omega}{2}\left(n + \frac{1}{2}\right) \qquad (19)$$

That is, for a single harmonic oscillator with oscillation frequency ω in an energy eigenstate, the expected kinetic and potential energy is the same.

In case the harmonic oscillator is not in an energy eigenstate, but in some arbitrary state $\ket{\psi}$, it is not necessarily the case that $\langle K \rangle = \langle V \rangle$. Consider the fact that the energy eigenkets span the Hilbert space, thereby implying that

$$\ket{\psi} = \sum_n \alpha_n\ket{n} \qquad (20)$$

for an appropriate choice of α_n. Furthermore, the expected kinetic energy and potential energy is therefore

$$\bra{\psi}\hat{K}\ket{\psi} = \sum_n\sum_m \alpha_n^*\alpha_m\bra{n}\hat{K}\ket{m} \qquad (21a)$$

$$\bra{\psi}\hat{V}\ket{\psi} = \sum_n\sum_m \alpha_n^*\alpha_m\bra{n}\hat{V}\ket{m} \qquad (21b)$$

It is not immediately obvious from (21) whether $\langle K \rangle$ is equal to $\langle V \rangle$ or not. However if $\langle K \rangle = \langle V \rangle$, then

$$\bra{\psi}\hat{K}\ket{\psi} - \bra{\psi}\hat{V}\ket{\psi} = \bra{\psi}\left(\hat{K} - \hat{V}\right)\ket{\psi}$$
$$= \sum_n\sum_m \alpha_n^*\alpha_m\bra{n}\left(\hat{K} - \hat{V}\right)\ket{m}$$

must be equal to zero. The braket can be simplified by inserting (18):

$$-\frac{\hbar\omega}{2}\left[\sqrt{(m+1)(m+2)}\delta_{n,m+2} + \sqrt{m(m-1)}\delta_{n,m-2}\right]$$

---

[2] This is actually somewhat missleading, but it originates in the derivation of the harmonic oscillator energies. By requiring non-negative energies, one finds that the "ladder of energy eigenkets" must be truncated in order to avoid $\langle n|n \rangle < 0$. This truncation is also the origin of the index n.

which in turn yields:

$$\langle\psi|\left(\hat{K}-\hat{V}\right)|\psi\rangle =$$
$$-\frac{\hbar\omega}{2}\sum_n \alpha_n^*\left(\alpha_{n-2}\sqrt{(n-1)n}+\alpha_{n+2}\sqrt{(n+2)(n+1)}\right)$$

As this expression is not necessarily 0 for any state $|\psi\rangle$, it is not necessarily true that $\langle K\rangle$ is equal to $\langle V\rangle$.

3.   The harmonic oscillator eigenfunctions

Having found the algebraic solution to the harmonic oscillator problem, the next step is to express the eigenkets in the position basis. The simplest eigenstate is the ground state, for which $E_0 = \hbar\omega/2$. Inserting $\hat{x}=x$ and $\hat{p}=-i\hbar\partial_x$ into (11) yields:

$$\mathcal{H}_{\text{HO}} = \frac{-\hbar^2}{2m}\partial_x{}^2 + \frac{1}{2}\omega^2 x^2$$

It is difficult to work with a plethora of constants and variables, thus consider the dimensionless variable $\xi = \sqrt{m\omega/\hbar}x$ with correponding derivative $\partial_\xi = \sqrt{m\omega/\hbar}\partial_x$. Written in terms of ξ, the equation $\hat{\mathcal{H}}|0\rangle = \frac{1}{2}\hbar\omega|0\rangle$ may be written as

$$\frac{-\hbar\omega}{2}\partial_\xi{}^2\psi_0(\xi) + \frac{\hbar\omega}{2}\xi^2\psi_0(\xi) = \frac{\hbar\omega}{2}\psi_0(\xi)$$

where $\psi_0(\xi) = \langle\xi|0\rangle$ is the ground state wave function. The solution is

$$\psi_0(\xi) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4}e^{-\xi^2/2} = Ce^{-\xi^2/2} \qquad (22)$$

Because the ground state is known, equation (16) implies that the excited states may be found by successive application of $\hat{a}^\dagger$. In terms of ξ and ∂_ξ, $\hat{a}^\dagger$ is:

$$a^\dagger = \frac{1}{\sqrt{2m\hbar\omega}}\left(-\hbar\partial_x + m\omega x\right) = \frac{1}{\sqrt{2}}\left(\xi - \partial_\xi\right)$$

Computing $(a^\dagger)^n\psi_0/\sqrt{n!}$ becomes incredibly inefficient for large n. However, it is actually unecessary because there exists exact solutions in terms of the so-called *Hermite polynomials*. These polynomials, commonly denoted by H, are solutions to the differential equation:

$$\left[\partial_\xi{}^2 - 2\xi\partial_\xi + \left(\lambda-1\right)\right]H(\xi) = 0 \qquad (23)$$

The differential equation has been generalised for any complex λ, but the solutions interesting to the harmonic oscillator have $\lambda = 2n+1$, where $n = 0, 1, \dots$ is the same index as used by the energy eigenkets. The Hermite polynomials can be explicitly expressed by Rodrigues' formula:

$$H_n(\xi) = (-1)^n e^{\xi^2}\left(\partial_\xi\right)^n e^{-\xi^2} \qquad (24)$$

from which one can show that the Hermite polynomials satisfy the following recurrence relations:

$$H_{n+1}(\xi) = 2\xi H_n(\xi) - \partial_\xi H_n \qquad (25a)$$
$$\partial_\xi H_n = 2n H_{n-1}(\xi) \qquad (25b)$$

Moreover, the Hermite polynomials satisfy the orthogonality condition:

$$\int_{-\infty}^{\infty}\mathrm{d}\xi\, e^{-\xi^2}H_n(\xi)H_m(\xi) = \delta_{nm}2^n n!\sqrt{\pi} \qquad (26)$$

The harmonic oscillator energy eigenfunctions are expressed in terms of the Hermite polynomails as:

$$\psi_n(x) = \frac{C}{\sqrt{2^n n!}}H_n(\xi)e^{-\xi^2/2}, \quad \xi = \sqrt{\frac{m\omega}{\hbar}}x \qquad (27)$$

(where C is the same as before.) While this equation may seem completely unmotivated, it can easily be shown from an argument of induction. The $n=0$ case is easy:

$$\langle\xi|0\rangle = \frac{C}{\sqrt{2^0 0!}}H_0(\xi)e^{-\xi^2/2} = Ce^{-\xi^2/2} = \psi_0(\xi)$$

Now let $n=k$ and consider $\langle\xi|k+1\rangle$:

$$\langle\xi|\frac{\hat{a}^\dagger}{\sqrt{k+1}}|k\rangle = C\frac{\xi-\partial_\xi}{\sqrt{2(k+1)}}\frac{1}{\sqrt{2^k k!}}H_k(\xi)e^{-\xi^2/2}$$

It follows from (25) that $\left(\xi-\partial_\xi\right)H_k(\xi)e^{-\xi^2/2}$ is equal to $H_{k+1}(\xi)e^{-\xi^2/2}$, which completes the induction proof:

$$\langle\xi|k+1\rangle = C\frac{1}{\sqrt{2^{k+1}(k+1)!}}H_{k+1}(\xi)e^{-\xi^2/2} = \psi_{k+1}(\xi)$$

Note that the question of orthonormality is ensured by the orthonormality of the Hermite polynomials.

4.   Composite systems of harmonic oscillators

Having studied the single harmonic oscillator, the final piece of the puzzle is to address a system of harmonic oscillators. Such a composite system could consist of different spatial degrees of freedom of the same particle, different particles or a mixture. Each individual harmonic oscillator behaves according to their Hamiltonian $\hat{\mathcal{H}}_{\text{HO}}$, and the system as a whole obeys equation (3).

Mathematically speaking, the Hamiltonian of a composite system with N harmonic oscillators is

$$\hat{\mathcal{H}}^{\text{sys}} = \hat{\mathcal{H}}_{\text{HO}}\otimes\cdots\otimes\hat{I} + \cdots + \hat{I}\otimes\cdots\otimes\hat{\mathcal{H}}_{\text{HO}} \qquad (28)$$

But this is tedius to write, thus let the ith term (i.e., the term where the ith component of the tensor product is $\hat{\mathcal{H}}_{HO}$) be denoted by $\hat{\mathcal{H}}_{\text{HO}}^i$ such that the above expression simplifies to equation (3). The system's energy eigenkets are

$$|\psi\rangle = |n_1,\dots,n_N\rangle = |n_1\rangle\otimes\cdots\otimes|n_N\rangle \qquad (29)$$

where each $|n\rangle$ is the regular harmonic oscillator energy eigenkets. Because $\hat{\mathcal{H}}_{\mathrm{HO}}^i$ only acts on the ith component of the tensor product, only the energy index n_i is affected:

$$\hat{\mathcal{H}}_{\mathrm{HO}}^i |\psi\rangle = \hbar\omega_i \left(n_i + \frac{1}{2} \right) |\psi\rangle \qquad (30)$$

where ω_i is the oscillation frequency of the ith harmonic oscillator. It follows that the composite system's energy eigenvalues are

$$\hat{\mathcal{H}}^{\mathrm{sys}} |\psi\rangle = \hbar \left[\omega_1 \left(n_1 + \frac{1}{2} \right) + \cdots + \omega_N \left(n_N + \frac{1}{2} \right) \right] |\psi\rangle \ (31)$$

This expression greatly simplifies when the oscillation frequency is uniform (i.e. $\omega_i = \omega$):

$$\hat{\mathcal{H}}^{\mathrm{sys}} |\psi\rangle = \hbar\omega \left(n_1 + \cdots + n_N + \frac{N}{2} \right) |\psi\rangle \qquad (32)$$

The ground state $(n_i = 0)$ energy of the composite harmonic oscillator with uniform frequency is therefore

$$E_g = \hbar\omega \frac{N}{2} \qquad (33)$$

If all oscillators are in an energy eigenstate, then $\langle K \rangle_n = \langle V \rangle_n$ implies that the total expected kinetic energy is equal to the total expected potential energy.

C. Markov Chain Monte Carlo Methods

1. Monte Carlo integration

One of the most versatile approaches to numerical integration is Monte Carlo integration. Its clever exploit of stochastic variables allows it to tackle high-dimensional integrals with little extra effort.

The goal of Monte Carlo integration is to evaluate integrals on the form:

$$I = \int_D \mathrm{d}x\, f(x), \quad x \in D \qquad (34)$$

where x is not necessarily one-dimensional. Let $p(x)$ denote a known distribution and consider the change of variables from x to $p(x)$:

$$\int_D \mathrm{d}x\, f(x) = \int_{x \in D} \mathrm{d}p(x)\, \frac{f(x)}{p(x)} = \left\langle \frac{f(x)}{p(x)} \right\rangle_{p(x)} \qquad (35)$$

That is, the integral I is equal to the expected value of $f(x)/p(x)$ with respect to the probability distribution $p(x)$. The idea is to draw N_{MC} random numbers $X \sim p(x)$ and estimate the expected value with the sample expected value:

$$\left\langle \frac{f(x)}{p(x)} \right\rangle_X \approx \sum_{i=1}^{N_{\mathrm{MC}}} \frac{f(x_i)}{p(x_i)} p(x_i) = \sum_{i=1}^{N_{\mathrm{MC}}} f(x_i) \qquad (36)$$

The basic Monte Carlo integration algorithm is summarised below:

Algorithm 2 Standard Monte Carlo Integration

1: Define N_{MC}.
2: Initialise SUM = 0.
3: **for** $i = 1, \ldots, N_{\mathrm{MC}}$:
4:     Draw x_i at random according to $p(x)$.
5:     Evaluate $f(x_i)$.
6:     Add $f(x_i)$ to SUM.
7: **end for**
8: Normalise final value: $I = \mathrm{SUM}/N_{\mathrm{MC}}$.

---

2. Markov chains

A Markov chain is a stochastic model for the discrete evolution of a so-called "memoryless system",[3] i.e. a system whose current state only depends on the previous state of the system. Formally, such a memoryless system satisfies the *Markov property*, which requires that the probability of moving from a state X_n to state X_{n+1} is only dependent on X_n. The state space may either be continuous or discrete, each of which introduce a similar albeit distinct formalism. This project will only use a continuous state space, thus the discrete formalism will be ignored here.

Say the state space for a particular Markov chain is $\mathcal{S}$, then a *stochastic kernel* on $\mathcal{S}$ is a function $p : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ that satisfies

$$p(x,y) \leq 0, \ \forall x,y \in \mathcal{S} \quad \text{and} \quad \int_{\mathcal{S}} \mathrm{d}y\, p(x,y) = 1, \ \forall x \in \mathcal{S}$$

Each Markov chain is parametrised by its corresponding stochastic kernel, which in this sense is usually referred to as the transition probability density. Now for every $p(x,y)$ there exists a so-called *Markov operator P* that is such that

$$\left[(\cdot) P \right](y) = \int_{\mathcal{S}} \mathrm{d}x\, p(x,y)(\cdot)$$

Note in particular that P is a left-acting operator, as per usual in the literature.

Furthermore, say the Markov chain has continued for n steps with previous measurements $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, then the next measurement $X_{n+1} = y$ is drawn according to the distribution $p^{(n+1)}(y)$ whose evolution from the current distribution $p^{(n)}(x)$ is given by the Markov operator of the Markov chain:

$$p^{(n+1)}(y) = \left[p^{(n)}(x) P \right](y) = \int_{\mathcal{S}} \mathrm{d}x\, p(x,y) p^{(n)}(x) \quad (37)$$

---

[3] A Markov chain may be generalised to continuous evolution, but this will not be used in this project.

A special case of (37) is when the distribution is invariant of P, that is, when $p^{(n+1)} = p^{(n)}$. This behaviour is called a stationary distribution and is usually denoted by $\pi(x)$:[4]

$$\pi(y) = \int_{\mathcal{S}} \mathrm{d}x \, p(x, y)\pi(x) \tag{38}$$

Such a state is guaranteed if the Markov chain obeys the so-called *detailed balance* condition. Provided a transition probability density $p(x, y)$ and a particular distribution $\pi(x)$, then the detailed balance condition states that the Markov chain corresponding to $p(x, y)$ is *reversible* with respect to $\pi(x)$ if

$$\pi(x)p(x, y) = \pi(y)p(y, x), \; \forall x, y \in \mathcal{S} \tag{39}$$

It follows that

$$\int_{\mathcal{S}} \mathrm{d}x \, p(x, y)\pi(x) = \pi(y) \int \mathrm{d}x \, p(y, x) = \pi(y)$$

because $p(x, y)$ is normalised for all $x, y \in \mathcal{S}$. In conclusion, in case a Markov chain is "well-behaved", meaning it satisfies detailed balance, then the so-called *limiting distribution* of the Markov chain approaches π as $n \to \infty$:

$$\lim_{n \to \infty} \left[p^{(0)}(x)P^n \right](y) = \pi(y) \tag{40}$$

where P^n implies repeated operations on $p^{(0)}$, $p^{(1)}$, etc.

3. Markov Chain Monte Carlo: The Metropolis algorithm

As the name implies, Markov Chain Monte Carlo (MCMC) methods combine the concept of a Markov chain with Monte Carlo methods. There exists several MCMC methods, each with unique properties that may or may not be benefitial. This project will focus on the so-called *Metropolis* algorithm, which is a special case of the *Metropolis-Hastings* algorithm.

The basic idea behind the MCMC methods stems from equation (36). Ideally one would be able to draw a sufficiently large number of states x from $p(x)$ and follow algorithm 2 religiously. However, this is incredibly computation-ineffective and essentially infeasible for a standard computer. Enter Markov chains: A Markov chain with limiting distribution $\pi(x) = p(x)$ would necessarily generate states that are approximately distributed according to $p(x)$. It turns out that this may be exploited in order to avoid large computations.

The mathematics of MCMC methods are based on the detailed balance principle, consider the following rearranging of equation (39):

$$\frac{p(x, y)}{p(y, x)} = \frac{\pi(y)}{\pi(x)}$$

———

[4] Becuase choosing π to represent something else other than π seemed like a great idea, obviously.

Now, suppose the current state of the Markov chain governed by $p(x, y)$ is X_n, i.e., the nth state in the chain. The idea is to separate the transition process into two steps: an initial *proposal*, and an *acceptance/rejection* step. These steps should be independent so that probability is independent, meaning the probability density of proposing a state y, $g(y|x)$, is independent of the probability of accepting the proposal, $\alpha(x, y)$. Because they are independent, it follows that

$$p(x, y) = g(y|x)\alpha(x, y) \tag{41}$$

meaning the above criteria may be rewritten as

$$\frac{\alpha(y, x)}{\alpha(x, y)} = \frac{p(y)}{p(x)} \frac{g(x|y)}{g(y|x)} \tag{42}$$

where $p(x) = \pi(x)$ is the assertion of MCMC methods that the limiting distribution of the Markov chain governed by $p(x, y)$ is the integrand of equation (34). This expression is particularly easy to implement in case the normalisation of p and g are independent on x and y, or if they happen to cancel each other (although this is very unlikely).

The last step it to choose an acceptance probability $\alpha(y, x)$ that satisfies 42, the *Metropolis choice* is

$$\alpha(y, x) = \min \left\{ 1, \frac{p(y)}{p(x)} \frac{g(x|y)}{g(y|x)} \right\} \tag{43}$$

The Metropolis-Hastings algorithm may now be stated:

Algorithm 3 The Metropolis-Hastings Algorithm

1: Initialise the first state X_0.
2: **for** $n = 1, \ldots, N_{\text{MH}}$:
3:      Generate a proposal y according to $g(y|x)$.
4:      Evaluate acceptance probability $\alpha(y, x)$.
5:      **if** $\alpha(y, x) = 1$:
6:          Accept y.
7:      **else**
8:          Draw a uniformly distributed number $a \in [0, 1]$.
9:          **if** $a \leq \alpha(y, x)$:
10:             Accept y.
11:          **else**
12:             Reject y and set $X_{n+1} = X_n$.
13:          **end if**
14:      **end if**
15: **end for**

---

where N_{MH} is the number of steps in the Markov chain. Note that the algorithm above does not factor in the Monte Carlo update, this is because the central purpose of the Metropolis-Hastings algorithm is namely just to sample x values according to $p(x)$. Nonethelesss, implementations of the algorithm usually include an intermediate step inside this loop in order to avoid multiple loops.

Furthermore, although the distribution is guaranteed to converge, it does not necessarily converge within the first few steps. Accordingly, a common practice is to

perform N_{prep} preparation updates to X_0 in a process known as "burn-in", before invoking the Monte Carlo steps. When a step in the Markov chain represents a forward step in time, the amount of necessary burn-in is commonly referred to as the equilibration time of the system.

As previously mentioned, this project will implement the Metropolis algorithm, which is a special case of the Metropolis-Hastings algorithm. The Metropolis assumes that g is symmetric in x and y, meaning $g(x|y) = g(y|x)$. This implies that the acceptance probability may be simplified to

$$\alpha_{\text{M}}(y, x) = \min\left\{1, \frac{p(y)}{p(x)}\right\} \qquad (44)$$

III. METHOD

IV. RESULTS

V. DISCUSSION

VI. CONCLUSION

Appendix A: Miscellaneous Material