

Monte Carlo method

Monte Carlo is a way of approximating expectation value or sample mean of a function of random numbers X_i drawn from pdf $p(x)$

$$\langle f \rangle = \int dx f(x) p(x) \approx \frac{1}{n} \sum_{i=1}^n \{f(X_i)\}$$

Two things are thus needed :

- (i) function or observable $f(x)$ whose expectation to be approximated
- (ii) pdf $p(x)$ that $f(x)$ is expected to obey.

$$0 \leq p(x) \leq 1 \quad \text{and} \quad \sum_{x_i \in D} p(x_i) = 1 \quad \text{or} \quad \int_D p(x) dx = 1$$

$$\langle f(x) \rangle = \sum_{x_i \in D} f(x_i) p(x_i) \quad \text{or} \quad \int_{x \in D} f(x) p(x) dx$$

$$\sigma_f^2 = \langle f(x)^2 \rangle - \langle f(x) \rangle^2$$

From N such independent samples $\langle f \rangle$ and their corresponding σ_f^2 from which we get the MC estimate

$$\langle \langle f(x) \rangle \rangle = \frac{1}{N} \sum_{i=1}^N \langle f(x) \rangle_i \quad \text{and} \quad \sigma_N^2 = \frac{\sigma_f^2}{N} \Rightarrow \sigma_N \sim \frac{1}{\sqrt{N}}$$

MC estimate of $f(x)$ thus decreases as $1/\sqrt{N}$.

If MC is used to estimate $\int_a^b f(x)dx$, then it is certainly at disadvantage when compared to Trapezoidal or Simpson where errors fall as $1/N^2$ and $1/N^4$ resp. But it is true for one or fewer dimensions, in higher dimensions Monte Carlo becomes significantly efficient.

Despite this, it is still advisable to find way(s) to reduce the variance of MC estimators. One basic step would be to choose a good distribution $p(x)$ from which to draw the random variables, the idea known as importance sampling.

Monte Carlo is applied, but not limited to, in the following situations

1. modeling outcomes of systems that involve large number of parameters like rolling of dice, crowd movement, weather phenomena, particle events in collision etc.
2. multi-dimensional integration and minimization problem in high dimensional space
3. classical and quantum mechanical many-particle thermodynamical system, involving Markov Chain.

Consider MC estimate of the integral below

$$\mathcal{I} = \int_{x \in D} f(x) dx = \int_{x \in D} f(x) \frac{p(x)}{p(x)} dx = \int_{x \in D} \frac{f(x)}{p(x)} p(x) dx \equiv \left\langle \frac{f(x)}{p(x)} \right\rangle$$

Let estimator of \mathcal{I} be \mathcal{F} ,

$$\mathcal{F} \equiv \frac{1}{N} \sum_{i=1}^N \frac{f(X_i)}{p(X_i)}$$

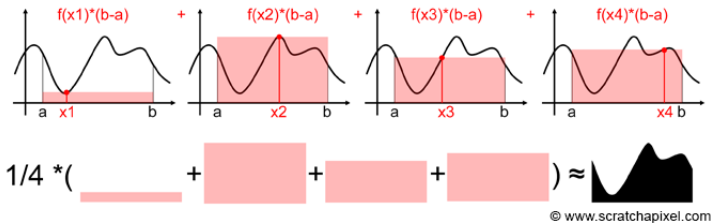
As is the usual practice, if $p(x)$ is uniformly distributed $\in [a, b]$,

$$p(x) = \frac{1}{b-a} \Rightarrow \mathcal{F} = \frac{b-a}{N} \sum_{i=1}^N f(x_i)$$

This looks very similar to [midpoint integration scheme](#), instead of choosing $f(x)$ at interval middle one chooses them at RN coordinate $X_i \in [a, b]$. The steps involved in MC integration are,

1. Choose N to be large, say 500 or 1000 or whatever.
2. Draw $X_i \in [0, 1]$ for N times, each time converting to $Y_i = a + (b-a)X_i$
3. For each Y_i determine $f(Y_i)$ and hence, \mathcal{F}_N and σ_f
4. Keep increasing N till convergence.

Schematically MC integration looks very much like the following



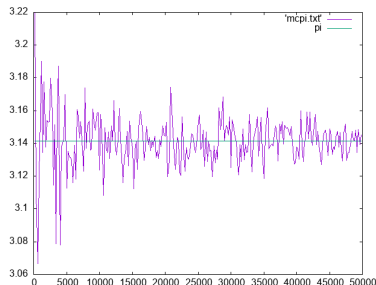
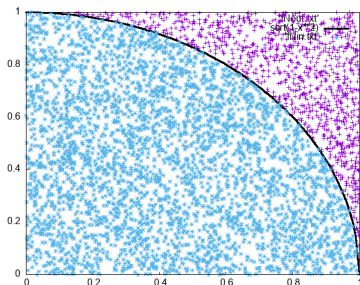
where $N = 4$ RN $\{x_1, x_2, x_3, x_4\}$ and calculated the area $(b-a) * f(Y_i)$ to estimate $\int_a^b f(x) dx$. As N becomes larger and larger, \mathcal{F}_N converges to a value but decrease in σ_f will be rather slow $\sim 1/\sqrt{N}$.

Most popular way to demonstrate the above application of MC is to determine the value of π from the integral

$$\int_0^1 \sqrt{1-x^2} dx = \pi/4$$

1. Put unit circle in a square of side $2r = 2$. Confine to first quadrant
 $\Rightarrow \text{area} = \pi/4$.
2. Randomly generate points (x_i, y_i) and check for inside points
 $x_i^2 + y_i^2 \leq 1$.

A typical plot of the integral, showing RN points inside and those outside.



Then over **LARGE** number of trials

$$\frac{\text{circle area}}{\text{square area}} \approx \frac{\text{total inside}}{\text{total trials}} \Rightarrow \pi \approx 4 \times \frac{\text{total inside}}{\text{total trials}}$$

For more complicated $f(x)$, however, simple uniformly distributed $p(x)$ would inefficiently sample the space and convergence will be slow as would be the decrease in variance. Choice of $p(x)$ closely proportional to $f(x)$ or motivated by physics (way $f(x)$ is likely to be distributed) helps in convergence.

In multi-variable system where $y = f(x_1, x_2, \dots, x_n)$ and $\{x_i\}$ are set of independent variables or **generalized coordinates** in system's phase space.

A given sequence of such variables or coordinates is called **configuration**.

In classical statistical physics, the weighted sum of all states accessible to the system is called **partition function**.

$$Z = \int \exp[-\beta H(q, p)] dq dp \quad \text{or alternatively} \quad Z = \int DU \exp[-S(U)]$$

where q, p, U are collections of appropriate coordinates. Average of an observable is then given by,

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int DU \mathcal{O}(U) \exp[-S(U)] \quad \text{whose MC gives} \quad \langle \mathcal{O} \rangle \approx \frac{1}{N} \sum_{i=1}^N \mathcal{O}(\{U\}_i)$$

where $\langle \mathcal{O} \rangle$ is called **ensemble average** over a sequence of **independent or uncorrelated** configurations.

Configurations are generated by **Monte Carlo** based on physics contained in the **Action or Hamiltonian**.

From partition function it follows that **configurations** that minimizes **Action / Hamiltonian** contributes most significantly to **ensemble average**.

Importance sampling

Random number generation with **pdf** $f(x)$ using **accept / reject method** gives first glimpse of importance sampling \rightarrow appropriate choice of $g(x)$ such that $f(x)/g(x) = c \approx 1$.

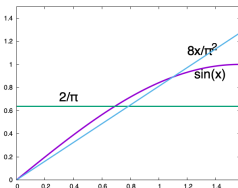
Consider Monte Carlo approximation to

$$\int_0^1 f(x) dx \text{ where } f(x) = 0 \text{ for } 0 > x > 1 \text{ using } \begin{cases} U & \in \text{uniform } [0, 1] \\ W & \in \text{uniform } [0, 4] \end{cases}$$

Monte Carlo approximate using $w_i \in W$ makes little sense since on average 80% of $f(w_i)$ will tell nothing substantial about integral since $f(x) = 0$ for $1 < x < 5$. Consider an example

$$\mathcal{F} = \int_0^{\pi/2} \sin(x) dx \xrightarrow{\text{analytical}} 1 \text{ using } \begin{cases} p_1(x) = 2/\pi \\ p_2(x) = 8x/\pi^2 \end{cases}$$

Which **pdf**, $p_1(x)$ or $p_2(x)$, is expected to reduce variance and converge faster? The plot of the functions $f(x)$, $p_i(x)$ gives a hint



Recall, the MC estimate of integral of $f(x)$,

$$\mathcal{F} = \frac{1}{N} \sum_{i=1}^N \frac{f(X_i)}{p(X_i)} = \frac{1}{N} \sum_{i=1}^N \frac{\sin(X_i)}{p_n(X_i)}$$

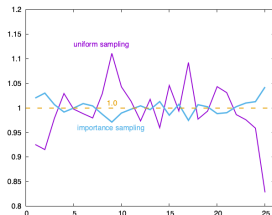
$$p_1(x) = \frac{1}{\pi/2} \Rightarrow \mathcal{F} = \frac{\pi}{2N} \sum_{i=1}^N \sin(X_i) \text{ where } X_i = \frac{\pi}{2} x_i, \quad x_i \in [0, 1]$$

$$p_2(x) = \frac{8x}{\pi^2} \Rightarrow \mathcal{F} = \frac{\pi^2}{8N} \sum_{i=1}^N \frac{\sin(\xi_i)}{\xi_i} \text{ where } \xi_i = \frac{\pi}{2} \sqrt{x_i}, \quad x_i \in [0, 1]$$

RN for **pdf** $p(x) = 8x/\pi^2$ is obtained uniformly distributed RN $\in [0, 1]$ as

$$x = \int_0^\xi \frac{8y}{\pi} dy = \frac{4\xi^2}{\pi^2} \Rightarrow \xi = \frac{\pi}{2} \sqrt{x}$$

Result demonstrates reduction of variance using **importance sampling**



Markov process

One of the most familiar average over weighted configurations is classical many-particle system in canonical ensemble with Boltzman weight

$$p(X) \propto \exp \left[-\frac{E(X)}{k_B T} \right]$$

When temperature is finite, configurations with high energy will appear with smaller probability. An efficient way to generate weighted configurations in MC is **Markov process**.

Markov process generates random states *i.e.* independent configurations with probability distribution depending on previous configuration. For independent or uncorrelated sequence of configurations, the probability of occurrence of a particular sequence is

$$P(C_1, C_2, \dots, C_N) = P(C_1)P(C_2) \cdots P(C_N)$$

A stochastic process where a set of configurations is generated in sequence according to some transition probability $T(C_i \rightarrow C_j)$,

$$P(C_1, C_2, \dots, C_N) = P(C_1) T(C_1 \rightarrow C_2) T(C_2 \rightarrow C_3) \cdots T(C_{N-1} \rightarrow C_N)$$

This defines a **Markov chain**, whose elements are random variables because of built-in probability in generating configurations.

Configuration average in MC is taken over elements of **Markov chain**,

$$\langle \mathcal{O} \rangle = \frac{1}{N} \sum_{i=1}^N \mathcal{O}(C_i)$$

Markov chain should have the following properties

1. Chain is **irreducible**. Starting from an arbitrary configuration C_i there exists a finite probability of reaching any other configuration C_k in a finite number of Markov steps.
2. Markov chain is **aperiodic**. Meaning after generating C_i , Markov step cannot bring it back to the same C_i except after finite mean recurrence time. Such a state is called **positive**.

Markov chain that satisfies these conditions is called **ergodic**. Condition satisfied by positive, irreducible Markov chain is **detailed balance**,

$$P(C) T(C \rightarrow C') = P(C') T(C' \rightarrow C)$$

Note $T(C \rightarrow C')$ is not a function of variables labeling configurations C , C' but stands for a rule to select C' from C in **Markov chain**.

Metropolis algorithm

In context of classical thermal equilibrium, **detailed balance** gives

$$\frac{T(C \rightarrow C')}{T(C' \rightarrow C)} = \frac{P(C')}{P(C)} = \frac{\exp[-\beta E']}{\exp[-\beta E]} = \exp[-\beta(E' - E)]$$

Detailed balance does not determine transition probability uniquely, therefore gives some freedom to choose **optimal** algorithm specific to problems. Using this freedom, we model it as

$$T(C \rightarrow C') = w_{CC'} A_{CC'} \quad \text{where} \quad w_{CC'} = w_{C'C}$$

$w_{CC'}$ is selection or trial state probability, $A_{CC'}$ is **acceptance probability**.

$$\frac{A_{CC'}}{A_{C'C}} = \exp[-\beta(E' - E)] = \exp[-\beta \Delta E]$$

This methodology is basis of **Metropolis algorithm** – a two step process

- (i) propose C' with probability $w_{CC'}$ given initial configuration is C
- (ii) Accept or reject the proposed C'

Way to propose a state C' is what **molecular dynamics** all about.

Metropolis algorithm satisfying detailed balance and ergodicity is

1. Given C , generate C' with probability $w_{CC'}$ by molecular dynamics.
2. Calculate $P(C')/P(C) = \exp[-\beta\Delta E]$.
3. If $\Delta E < 0$ i.e. $P(C') > P(C)$, always accept C' as new configuration since we move to a state with lower energy. This implies acceptance probability $A_{CC'} = 1$.
4. If $\Delta E > 0$, meaning energy is increasing, choose with acceptance probability $A_{CC'} = \exp[-\beta\Delta E]$. To accept configuration with probability $A_{CC'}$ and reject it with probability $1 - A_{CC'}$.
5. Choose a random number $X \in [0, 1]$. If $X < A_{CC'}$ accept the state else reject it and keep the old one. Standard way to express this step is

$$X = \min \left\{ 1, \frac{\exp[-\beta E']}{\exp[-\beta E]} \right\}$$

Metropolis accept / reject step is important as it allows thermal fluctuations around equilibrium state and thermal fluctuations gives rise to phase transition. Whenever an update algorithm is followed by Metropolis accept/reject, we call the scheme **exact**.

Molecular dynamics

Question : how to make change $C \rightarrow C'$ maintaining detailed balance?

Statistical system, classical or quantum, is described by **partition function** featuring either **Hamiltonian** or **Action**.

Suggests deterministic way to evolve is to use **classical EoM** using either **Newton's EoM** (velocity Verlet) or **energy conserving Hamiltonian EoM**.

Evolution in classical mechanics using **Liouville operator** $i\mathcal{L}$,

$$i\mathcal{L} = \sum_{\alpha=1}^N \left[\dot{q}_{\alpha} \frac{\partial}{\partial q_{\alpha}} + \dot{p}_{\alpha} \frac{\partial}{\partial p_{\alpha}} \right]$$

Time dependence of an arbitrary phase-space function $\mathcal{A}(t)$ is

$$\frac{d\mathcal{A}}{dt} = i\mathcal{L} \mathcal{A} \Rightarrow \mathcal{A}(t) = e^{i\mathcal{L}t} \mathcal{A}(0)$$

Operator $\exp(i\mathcal{L}t)$ is classical propagator and closely resembles quantum propagator $\exp(-i\hat{H}t/\hbar)$. Time evolution of phase space vector $\mathbf{x}(t)$ then is

$$\mathbf{x}(t) = e^{i\mathcal{L}t} \mathbf{x}(0)$$

This equation is central in molecular dynamics to obtain time dependent trajectory $\mathbf{x}(t)$ in phase space given an initial condition $\mathbf{x}(0)$.

Verlet algorithm

It solves Newton's EoM with conservative force numerically starting from

$$\frac{d\mathbf{x}}{dt} = \mathbf{v} \quad \text{and} \quad \frac{d\mathbf{v}}{dt} = \frac{\mathbf{F}}{m}$$

Simple forward discretization with time interval Δt to $\mathcal{O}(\Delta t)$ yields,

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t) \Delta t \quad \text{and} \quad \mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{\mathbf{F}(t)}{m} \Delta t$$

A more time symmetric approach, known as **leapfrog Verlet** gives better accuracy over long simulations

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}(t + \Delta t/2) \Delta t \quad \text{and} \quad \mathbf{v}(t + \Delta t/2) = \mathbf{v}(t - \Delta t/2) + \frac{\mathbf{F}(t)}{m} \Delta t$$

In actual practice, a more practical form of **leapfrog Verlet** is used which is corrected to $\mathcal{O}(\Delta t^2)$. Starting with evaluating force $\mathbf{F}(t) = -\nabla U(\mathbf{x}(t))$, the steps involved are

$$\begin{aligned} \mathbf{v}(t + \Delta t/2) &\equiv \bar{\mathbf{v}}(t) = \mathbf{v}(t) + \mathbf{F}(t)\Delta t/2 & \mathbf{x}(t + \Delta t) &= \mathbf{x}(t) + \bar{\mathbf{v}}(t)\Delta t \\ \mathbf{v}(t + \Delta t) &= \bar{\mathbf{v}}(t) + \mathbf{F}(t + \Delta t)\Delta t/2 & \mathbf{F}(t + \Delta t) &= -\nabla U(\mathbf{x}(t + \Delta t)) \end{aligned}$$

Another integrator used for solving Newton's EoM is **Gear's 6th order algorithm** that returns global accuracy of $\mathcal{O}(\Delta t^6)$. Steps involved are

$$\mathbf{r}_0(t) \equiv \mathbf{r}(t), \quad \mathbf{r}_1(t) \equiv \Delta t \frac{d\mathbf{r}(t)}{dt}, \quad \mathbf{r}_2 \equiv \frac{\Delta t^2}{2!} \frac{d^2\mathbf{r}(t)}{dt^2}, \quad \mathbf{r}_3 \equiv \frac{\Delta t^3}{3!} \frac{d^3\mathbf{r}(t)}{dt^3} \dots$$

$$\mathbf{r}_n(t + \Delta t) = \mathbf{r}_n^p(t + \Delta t) + C_n \left[\mathbf{r}_2^c(t + \Delta t) - \mathbf{r}_2^p(t + \Delta t) \right]$$

$$C_0 = \frac{3}{16}, \quad C_1 = \frac{251}{360}, \quad C_2 = 1, \quad C_3 = \frac{11}{18}, \quad C_4 = \frac{1}{6}, \quad C_5 = \frac{1}{60}$$

where the superscripts **p**, **c** are predicted and corrected positions.

What about **RK4**? It is good but computational cost is high, force being evaluated in each of 4 steps in every iterations. Besides, **detailed balance** i.e. **time reversibility** is often tend to get spoiled after every few steps.

Hybrid Monte Carlo

It is combination of molecular dynamics followed by Metropolis accept / reject. Advantage is that HMC is a global algorithm that uses MD to update coordinates or fields at all sites in a single trajectory. The accept / reject decision renders the algorithm **exact**. By **exact** it means that HMC contains no discretization errors arising out of MD part.

Langevin proposed a stochastic differential equation to evolve a system subjected to both **Newton's EoM** and **random forces**, which help attaining **ergodicity** in phase space. One of the applications of it was describing **Brownian motion** and systems related to similar physics.

$$m \frac{d\mathbf{v}}{dt} = -\lambda \mathbf{v} + \boldsymbol{\eta}(t), \quad \text{where, } \langle \eta_i(t) \eta_j(t') \rangle \propto \delta_{ij} \delta(t - t')$$

where (viscous) force is proportional to velocity and $\boldsymbol{\eta}$ is Gaussian distributed noise. When applied to Monte Carlo, a **time variable** τ is introduced, which when discretized **labels the elements of Markov chain**.

If \mathbf{q} denotes generalized coordinates of a system with action $S[\mathbf{q}]$, then **Langevin algorithm** says ($\boldsymbol{\eta}$ being Gaussian noise),

$$q_i(\tau_{n+1}) = q_i(\tau_n) + \epsilon_L \left(-\frac{\partial S[\mathbf{q}]}{\partial q_i(\tau_n)} + \eta_i(\tau_n) \right) \xrightarrow{\epsilon_L \rightarrow 0} \frac{dq_i}{d\tau} = -\frac{\partial S[\mathbf{q}]}{\partial q_i} + \eta_i(\tau)$$

Langevin algorithm satisfies both ergodicity and detail balance. For updating field variables φ , simply replaced $S[q] \rightarrow S[\varphi]$. Choice of ϵ_L can in principle be large to cover system's phase (actually configuration) space in reasonable number of updating steps or trajectories. But absence of $\mathcal{O}(\epsilon_L^2)$ terms, systematic error is rather large and uncontrolled.

Next best thing is **modified molecular dynamics**. Idea is to make the following transformation

$$e^{-S[\varphi]} \rightarrow e^{-H[\varphi, \pi]} \quad \text{through} \quad \int d\pi e^{-\pi^2/2} e^{-S[\varphi]}$$

implying adding **Gaussian noise** π to action $S[\varphi]$ similar to Langevin and calling the combination as **Hamiltonian** $H[\varphi, \pi]$.

This Hamiltonian drives the system in phase space or space of $[\varphi]$ configurations through **fictitious / computer time**,

$$H[\varphi, \pi] = \sum_i \frac{\pi_i^2}{2} + S[\varphi]$$

and the corresponding change in partition function is

$$Z = \int D\varphi e^{-S[\varphi]} \rightarrow Z = \int D\varphi D\pi e^{-H[\varphi, \pi]}$$

NB : $S[\varphi]$, $H[\varphi, \pi]$, π are all dimensionless variables.

If **motion** generated by Hamiltonian is ergodic then Hamiltonian EoM,

$$\dot{\varphi}_i = \frac{\partial H[\varphi, \pi]}{\partial \pi_i} \quad \text{and} \quad \dot{\pi}_i = -\frac{\partial H[\varphi, \pi]}{\partial \varphi_i}$$

generate $\{\phi\}$ configurations with constant energy.

Momenta π induces dynamics in phase space but does not change physics since $\int D\pi \exp(-\pi^2/2)$ is a Gaussian integration contributing to overall normalization. Taylor expand φ, π ,

$$\begin{aligned}\varphi_i(\tau + d\tau) &= \varphi(\tau) + d\tau \dot{\varphi}_i(\tau) + \frac{d\tau^2}{2} \ddot{\varphi}_i(\tau) + \dots \\ \pi_i(\tau + d\tau) &= \pi_i(\tau) + d\tau \dot{\pi}_i(\tau) + \frac{d\tau^2}{2} \ddot{\pi}_i(\tau) + \dots\end{aligned}$$

From Hamiltonian EoM, various **time** derivatives are

$$\dot{\varphi}_i(\tau) = \pi_i(\tau)$$

$$\ddot{\varphi}_i(\tau) = \dot{\pi}_i(\tau) = -\frac{\partial S}{\partial \varphi_i}$$

$$\ddot{\pi}_i(\tau) = -\sum_j \frac{\partial^2 S}{\partial \varphi_i \partial \varphi_j} \pi_j(\tau) = \frac{1}{d\tau} \left(\frac{\partial S}{\partial \varphi_i(\tau + d\tau)} - \frac{\partial S}{\partial \varphi_i(\tau)} \right) + \mathcal{O}(d\tau)$$

Using these **time** derivatives in Taylor expansion,

$$\begin{aligned}\varphi_i(\tau + d\tau) &= \varphi(\tau) + d\tau \left(\pi_i(\tau) - \frac{d\tau}{2} \frac{\partial S}{\partial \varphi_i(\tau)} \right) + \mathcal{O}(d\tau^3) \\ \left(\pi_i(\tau + d\tau) - \frac{d\tau}{2} \frac{\partial S}{\partial \varphi_i(\tau + d\tau)} \right) &= \left(\pi_i(\tau) - \frac{d\tau}{2} \frac{\partial S}{\partial \varphi_i(\tau)} \right) - d\tau \frac{\partial S}{\partial \varphi_i(\tau + d\tau)} + \mathcal{O}(d\tau^3)\end{aligned}$$

Hamiltonian EoM used in **molecular dynamics** is called **leapfrog integrator**,

$$\begin{aligned}\varphi_i(\tau + d\tau) &= \varphi(\tau) + d\tau \pi_i \left(\tau + \frac{d\tau}{2} \right) \\ \pi_i \left(\tau + \frac{3}{2}d\tau \right) &= \pi_i \left(\tau + \frac{d\tau}{2} \right) - d\tau \frac{\partial S}{\partial \varphi_i(\tau + d\tau)}\end{aligned}$$

After a suitably chosen **molecular dynamics** time step, follows **Metropolis accept / reject**. This scheme of generating elements of **Markov chain** or new field configurations is called **Hybrid Monte Carlo**.

Quantum Monte Carlo

Quantum mechanical systems are described by Schrödinger equation, but in only handful of cases it can be solved analytically.

Solving becomes more challenging in many-particle systems like He, molecules, nucleus, electronic structures in solids etc. This is particularly because such systems consist of large number of interacting electrons, ions or nucleons.

Often important physics can be learnt from the ground state properties like energy etc. of such systems. In that case approximate methods like variational method, diffusion monte carlo, biased random walks are used.

Variational MC

Hamiltonian of a quantum mechanical system is of the form

$$\hat{H} = \sum_i^N -\frac{\hbar^2}{2m} \nabla_i^2 + \sum_{i>j}^N V(|r_i - r_j|)$$

where N can be fairly large. Ground state energy of the system comes from solving the above differential equation. For $N > 2$ it requires approximate method(s).

Variational principle makes educated guess of ground state trial wavefunction $\psi_T(\mathbf{r}, \alpha)$

(i) usually not eigenstate of \hat{H}

(ii) assumed to be real

Expectation value of \hat{H} w.r.t. ψ_T ,

$$\langle \hat{H} \rangle = \frac{\int d\mathbf{r} \psi_T^*(\mathbf{r}, \alpha) \hat{H} \psi_T(\mathbf{r}, \alpha)}{\int d\mathbf{r} \psi_T^*(\mathbf{r}, \alpha) \psi_T(\mathbf{r}, \alpha)} \geq E_0$$

$\langle \hat{H} \rangle$ is upper bound of E_0 that can be minimized with respect to parameter set α

$$\frac{\partial \langle \hat{H} \rangle}{\partial \alpha_i} = 0 \quad \text{where } i = 1, 2, \dots, N$$

Typically, wavefunction has values in small part of a larger configuration space \Rightarrow sample regions of configuration space where wavefunction takes appreciable values = Importance sampling.

Importance sampling is implemented by defining appropriate probability distribution function

$$p(\mathbf{r}) = \frac{\psi_T^*(\mathbf{r}, \alpha) \psi_T(\mathbf{r}, \alpha)}{\int d\mathbf{r} \psi_T^*(\mathbf{r}, \alpha) \psi_T(\mathbf{r}, \alpha)}$$

Importance sampling with $p(\mathbf{r})$ leads to

$$\begin{aligned}\langle \hat{H} \rangle &= \int d\mathbf{r} \frac{1}{\psi_T(\mathbf{r}, \alpha)} \frac{\psi_T^*(\mathbf{r}, \alpha) \psi_T(\mathbf{r}, \alpha)}{\int d\mathbf{r} \psi_T^*(\mathbf{r}, \alpha) \psi_T(\mathbf{r}, \alpha)} \hat{H} \psi_T(\mathbf{r}, \alpha) \\ &\equiv \int d\mathbf{r} p(\mathbf{r}) \hat{\mathcal{E}}(\mathbf{r}) = \langle \hat{\mathcal{E}}(\mathbf{r}) \rangle, \quad \text{where} \quad \hat{\mathcal{E}}(\mathbf{r}) = \frac{1}{\psi_T(\mathbf{r}, \alpha)} \hat{H} \psi_T(\mathbf{r}, \alpha)\end{aligned}$$

New operator $\hat{\mathcal{E}}(\mathbf{r})$ is called **local energy density**. This equation expresses **variational MC** approach to quantum system. It follows,

$$\langle \hat{H} \rangle = \langle \hat{\mathcal{E}}(\mathbf{r}) \rangle = \int d\mathbf{r} p(\mathbf{r}) \hat{\mathcal{E}}(\mathbf{r}) = \frac{1}{N_s} \sum_i \mathcal{E}(\mathbf{r}_i)$$

where \mathbf{r}_i are to be distributed according to the p.d.f. $p(\mathbf{r})$ and N_s are number of sample points. Steps involved are

1. MC starts with initial choices of $\{\mathbf{r}_0\}$ and variational parameters $\{\alpha_0\}$. Compute $p(\mathbf{r}) \propto |\psi_T(\mathbf{r}, \alpha)|^2$ and hence initial energy $\langle \hat{H} \rangle = \langle \hat{\mathcal{E}} \rangle$
2. Change $\mathbf{r} \rightarrow \mathbf{r}' = \mathbf{r} + \boldsymbol{\rho} \star \epsilon$, where ϵ is MC stepsize and $\boldsymbol{\rho} \in [0, 1]$.
3. Perform metropolis accept / reject $\{1, p(\mathbf{r}')/p(\mathbf{r})\}$. If accepted update $\langle \hat{\mathcal{E}} \rangle$. Either way, return to step 2.

Simple Harmonic Oscillator

Consider 1-dim SHO, with $m = \hbar = k = 1$ for simplicity, for which analytical results are available

$$\hat{H} = \frac{1}{2} \frac{d^2}{dx^2} + \frac{1}{2} x^2 \Rightarrow E_0 = \frac{1}{2} \text{ and } \psi_0(x) = \mathcal{N} e^{-x^2/2}$$

Trial wavefunction is decided by the property $\psi_T \rightarrow 0$ as $x \rightarrow \pm\infty$.

$$\psi_T(x, \alpha) = \mathcal{N} e^{-\alpha x^2}$$

$$\mathcal{E} = \frac{1}{\psi_T} \hat{H} \psi_T = e^{\alpha x^2} \left[-\frac{1}{2} \frac{d^2}{dx^2} + \frac{1}{2} x^2 \right] e^{-\alpha x^2} = \alpha + x^2 \left(\frac{1}{2} - 2\alpha^2 \right)$$

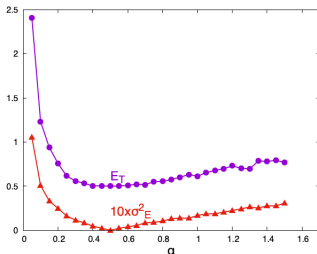
Obviously, for $\alpha = 1/2$ the energy $\mathcal{E} = 1/2$ because trial and true ground state wavefunctions are identical. Using Gaussian pdf $p(x) = |\psi_T|^2$, energy expectation value as a function of α is,

$$|\psi_T(x, \alpha)|^2 = e^{-2\alpha x^2} \text{ where variance } \sigma^2 = 1/4\alpha$$

$$\langle \mathcal{E} \rangle = \int_{-\infty}^{\infty} dx \psi_T^2 \mathcal{E} = \frac{\alpha}{2} + \frac{1}{8\alpha} \text{ and } \sigma_{\mathcal{E}}^2 = \frac{(1 - 4\alpha^2)^2}{32\alpha^2}$$

For $\alpha = 1/2$, variance $\sigma_{\mathcal{E}}^2 = 0$, $\langle \mathcal{E} \rangle_{\min} = 1/2$ certainly.

Without prior knowledge of **variational parameter** α for which $\langle \mathcal{E} \rangle_{\min}$, MC integration is performed over a range of α .



For potential other than SHO, say for instance x^4 can similarly be addressed. Using the above trial function ψ_T will give us a similar estimate of ground state energy but variance wouldn't be 0 at minimum.

Hydrogen atom

For hydrogen atom, overall wavefunction is complex but it comes from θ, ϕ angular dependence. The radial part is purely real. The radial part of hydrogen atom, $u(r) = r R(r)$ being the wavefunction,

$$-\frac{\hbar^2}{2m} \frac{\partial^2 u(r)}{\partial r^2} - \left(\frac{ke^2}{r} - \frac{\hbar^2 l(l+1)}{2mr^2} \right) u(r) = E u(r)$$

Solving the hydrogen atom Schrödinger equation above, it is imperative that it is written in terms of dimensionless variables because large variation of several constants result in loss of numerical precision.

$$\left[-\frac{1}{2} \frac{\partial^2}{\partial \rho^2} - \frac{1}{\rho} + \frac{l(l+1)}{2\rho^2} \right] u(\rho) = \lambda u(\rho)$$

Analytical ground state solution is $\lambda = -1/2$ and $u(\rho) = \rho e^{-\rho}$ where at $\rho = 0$, $u(\rho) = 0$ and $\rho \rightarrow \infty$, $u(\rho) = 0$. Using variational parameter α , (simplest) trial wavefunction is taken as

$$u_T(\rho, \alpha) = \alpha \rho e^{-\alpha \rho}$$

Local energy and pdf are

$$\mathcal{E}_L = \frac{1}{u_T} \left[-\frac{1}{2} \frac{\partial^2}{\partial \rho^2} - \frac{1}{\rho} + \frac{l(l+1)}{2\rho^2} \right] u_T = -\frac{1}{\rho} - \frac{\alpha}{2} \left(\alpha - \frac{2}{\rho} \right)$$
$$p(\rho) = \alpha^2 \rho^2 e^{-2\alpha \rho} \quad \text{where } \rho \in [0, \infty]$$

Helium atom

Generally Schrödinger equations for atomic, molecular and solid state systems, which involve interacting electrons and ions, can only be solved through various approximations. Simplest representative of this class is **helium atom**. Common approximation used to solve Schrödinger equation for systems of interacting electrons is **Born-Oppenheimer approximation**.

It involves separating **nuclear** and **electronic** degrees of freedom since nucleus is much heavier than electrons and very little momentum will be transferred between them. Consider $\mathbf{r}_{1,2}$ are position vectors of two electrons, the Hamiltonian is

$$H = -\frac{1}{2} \nabla_1^2 - \frac{1}{2} \nabla_2^2 - \frac{Z}{r_1} - \frac{Z}{r_2} + \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}, \quad \text{where } Z = 2$$

Because of the last term, Helium atom cannot be solved exactly. **Workable but questionable** approximation is neglect the last term i.e. **no repulsion between the electrons!** \Rightarrow in completely separated Hamiltonian,

$$H_1 = -\frac{1}{2} \nabla_1^2 - \frac{Z}{r_1} \quad \rightarrow \quad H_1 \psi_1 = E_1 \psi_1$$

$$H_2 = -\frac{1}{2} \nabla_2^2 - \frac{Z}{r_2} \quad \rightarrow \quad H_2 \psi_2 = E_1 \psi_2$$

$$\psi = \psi_1 \psi_2 \quad \text{and} \quad E = E_1 + E_2$$

A problem : Trial functions for variational MC cannot be $\exp(-\alpha r)$ since it cannot prevent two electrons coming close for no repulsion condition to hold. Cusp condition put constraints on wavefunction and its derivatives when $r_{1,2} \text{ \& } r_{12} \rightarrow 0$.

Consider electron 1 is close to nucleus $r_1 \rightarrow 0$ but two electrons are far apart from each other $r_2 \neq 0$. Local energy when $l = 0$ is then

$$\begin{aligned}\mathcal{E}_L(r) &= \frac{1}{\psi_T(r)} H \psi_T(r) = \frac{1}{\psi_T(r)} \left(-\frac{1}{2} \nabla_1^2 - \frac{Z}{r_1} \right) \psi_T(r) + \text{finite terms} \\ &= \frac{1}{\mathcal{R}_T(r_1)} \left(-\frac{1}{2} \frac{d^2}{dr_1^2} - \frac{1}{r_1} \frac{d}{dr_1} - \frac{Z}{r_1} \right) \mathcal{R}_T(r_1) + \text{finite terms} \\ \lim_{r_1 \rightarrow 0} \mathcal{E}_L(r) &= \frac{1}{\mathcal{R}_T(r_1)} \left(-\frac{1}{r_1} \frac{d}{dr_1} - \frac{Z}{r_1} \right) \mathcal{R}_T(r_1)\end{aligned}$$

d^2/dr_1^2 does not diverge because $\psi(r_1)$ is finite at origin. This indicates that in order to remove diverging term in the limit $r_1 \rightarrow 0$,

$$\frac{1}{\mathcal{R}_T(r_1)} \frac{d\mathcal{R}_T(r_1)}{dr_1} = -Z \Rightarrow \mathcal{R}_T(r_1) \propto e^{-Zr_1}$$

Similar cusp condition applies to electron 2. For $l \neq 0$, $Z \rightarrow Z/(l+1)$

Other constraint on wavefunction corresponds to stopping electron coming close to each other. When $r_{12} \rightarrow 0$, resulting equation for r_{12} dependence is similar to electron-nucleus. Except attractive Coulomb interaction between nucleus and electron is replaced by repulsive interaction and kinetic energy term is twice as large because reduced mass of two electrons is the half the electron mass. Replacing $-Z/r$ by $1/r_{12}$,

$$\left(-2 \frac{d^2}{dr_{12}^2} - \frac{4}{r_{12}} \frac{d}{dr_{12}} + \frac{2}{r_{12}} + \frac{l(l+1)}{r_{12}^2} \right) \mathcal{R}_T(r_{12}) = 0$$

The local energy is \mathcal{E}_L ,

$$\lim_{r_1 \rightarrow 0} \mathcal{E}_L(r) = \frac{1}{\mathcal{R}_T(r_{12})} \left(-\frac{4}{r_{12}} \frac{d}{dr_{12}} - \frac{Z}{r_{12}} \right) \mathcal{R}_T(r_{12})$$

Changing variable $\rho(r_{12}) = r_{12}^{-l} \mathcal{R}_T(r_{12})$ yields cusp condition as

$$\frac{1}{\mathcal{R}_T(r_{12})} \frac{d\mathcal{R}_T(r_{12})}{dr_{12}} = \frac{1}{2(l+1)}$$

For general system containing more than two electrons, there is one cusp condition for each ij pair of electrons.

Therefore, a possible ground state trial function can be

$$\psi_T(r) = e^{-\alpha(r_1+r_2)+r_{12}/2}$$

Both the electrons in Helium atom are in hydrogen atom ground state **1s**. Since the spins of electrons are not considered, there is no **Pauli exclusion principle**. But this approach cannot be taken for atoms beyond Helium.

The first step is to construct **pdf** according to

$$\rho(r_1, r_2) \sim |\psi_T(r_1, r_2)|^2 = e^{-2\alpha(r_1+r_2)+r_{12}}$$

In the second, calculated the local energy \mathcal{E}_L .

Finally, calculated ground state energy and its variance over changing α .

Maximum Entropy Method

MEM is an inverse problem – trying to find out **most likely pdf** given some known moments *i.e.* average properties.

Typically, the averages or moments are of the forms

$$\int_{\Omega} x^k p(x) dx = \mu_k \text{ for } k = 0, 1, 2, \dots, N \text{ or say, } C(x, x_0) = \int_{\Omega} \rho(x) p(x) dx$$

Estimation of function **pdf** $p(x)$ is obtained by maximizing **Shannon entropy** (or priors)

$$S = - \int_{\Omega} p(x) \ln [p(x)] dx$$

Entropy is defined as average rate at which information is produced by a system. Alternatively said, **S** measures how much of the system is unknown. **Maximum Entropy method** is based on the idea that – **distribution that maximizes information entropy is the one that is statistically most likely to occur.**

Introducing **Lagrangian multipliers** λ_k , entropy functional is defined as

$$H \equiv S + \sum_k \lambda_k \left(\int_{\Omega} x^k p(x) dx - \mu_k \right)$$

For discrete cases, above integrals are replaced by sum, for instance

$$S = \sum_i p_i \ln p_i, \quad \mu_k = \sum_i x_i^k p_i \quad \text{etc.}$$

Functional H is maximum when functional derivatives with respect to $p(x)$ and λ_k are zero

$$\frac{\delta H}{\delta \lambda_k} = 0 \quad \text{and} \quad \frac{\delta H}{\delta p(x)} = 0 \quad \Rightarrow \quad \ln p(x) = -1 + \sum_{k=0}^N \lambda_k x^k$$

The general solution is more conveniently written as

$$p(x) = e^{-1 + \sum_k \lambda_k x^k} = e^{\sum_k \lambda_k x^k} \quad \text{by using } \lambda_0 - 1 \rightarrow \lambda_0$$

To find maximum entropy solution $p(x)$, the **Lagrangian multipliers** λ_k are required to be solved from a system of nonlinear equations

$$\mu_0 = \exp(\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \cdots \lambda_k x^k)$$

$$\mu_1 = x \exp(\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \cdots \lambda_k x^k)$$

...

$$\mu_k = x^k \exp(\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \cdots \lambda_k x^k)$$

Newton solver may be used to calculate λ_k .

Hartree / Hartree-Fock

Variational technique to approximate ground state wavefunction and energy of quantum multi-electron system in stationary state. Generic form of Hamiltonian of a system consisting of N electrons and K nuclei with charges Z is

$$H = - \sum_{i=1}^N \frac{1}{2m_e} \nabla_i^2 - \sum_{k=1}^K \frac{1}{2M_n} \nabla_k^2 + \sum_{\substack{i,j \\ i>j}} \frac{e^2}{r_{ij}} - \sum_{i,k}^{N,K} \frac{Ze^2}{r_{ik}} + \sum_{\substack{k,k' \\ k>k'}} \frac{Z^2e^2}{r_{kk'}}$$

where terms respectively are kinetic energies of electrons followed by nucleus, electron-electron repulsion, nucleus-electron Coulomb attraction and finally nucleus-nucleus Coulombic repulsion.

For Schrödinger equation to be solvable (even approximately), variables of the electrons and nuclei have to be separated – **Born Oppenheimer approx.**

This approximation \Rightarrow set kinetic term for nuclei to zero as $M_n \gg m_e$ and the last or fifth term is constant. Wavefunction of the approximated Schrödinger equation is

$$|\psi(r_1, r_2, \dots, r_N)\rangle \approx \psi_1(r_1) \psi_2(r_2) \cdots \psi_N(r_N)$$

where $\psi_i(r_i)$'s are solutions of **one-electron Schrödinger eqns.**

One-electron Schrödinger eqn. is of the form

$$H = - \sum_{i=1}^N \frac{1}{2m_e} \nabla_i^2 + \mathcal{V}(r) \quad \text{where} \quad \mathcal{V}(r) = - \sum_k \frac{Ze^2}{r_{ik}} + V_{\text{hf}}(r)$$

$$V_{\text{hf}}(r) = - \int dr' \frac{\rho(r')}{|r - r'|} \quad \text{where} \quad \rho(r') = e |\psi_i(r_i)|^2$$

V_{hf} is **mean field** arising from $N - 1$ other electrons assumed to be distributed with charge density $\rho(r')$. Thus electrons are uncorrelated that helped building separable N one-electron wavefunctions.

So far two important quantum properties are completely neglected – (i) **indistinguishability** (ii) **spin**.

$$\psi_1(r_1) \psi_2(r_2), \dots, \psi_N(r_N) \neq \psi_1(r_2) \psi_2(r_1), \dots, \psi_N(r_N) \\ \psi(r_1, r_2, \dots, r_N) \neq -\psi(r_2, r_1, \dots, r_N)$$

Both these problems are addressed by constructing **anti-symmetric** trial wavefunction from above $\psi_i(r_i)$ s. Considering a simple two-electron system as in Helium atom,

$$\Psi(r_1, r_2) = \frac{1}{\sqrt{2}} [\psi_1(r_1)\psi_2(r_2) - \psi_1(r_2)\psi_2(r_1)] \equiv \frac{1}{\sqrt{2}} \begin{vmatrix} \psi_1(r_1) & \psi_2(r_1) \\ \psi_1(r_2) & \psi_2(r_2) \end{vmatrix}$$

Energy equation in terms of Hartree approximation is

$$E = \langle \Psi | H | \Psi \rangle = \left\langle \Psi \left| - \sum_{i=1}^N \frac{1}{2m_e} \nabla_i^2 - \sum_k \frac{Ze^2}{r_{ik}} - \int dr' \frac{\rho(r')}{|r - r'|} \right| \Psi \right\rangle$$

Expectation value of first two energy integrals is **one-electron integral**

$$\begin{aligned} \left\langle \Psi \left| - \sum_{i=1}^N \frac{1}{2m_e} \nabla_i^2 - \sum_k \frac{Ze^2}{r_{ik}} \right| \Psi \right\rangle &= - \sum_{i=1}^N \int dr \psi_i^*(r) \frac{1}{2m_e} \nabla_i^2 \psi_i(r) \\ &\quad - Ze^2 \sum_k \int dr \psi_i^*(r) \frac{1}{r_{ik}} \psi_i(r) \end{aligned}$$

These terms depend only on position or momentum of one electron and its interaction with nuclei. Last term is the sum of repulsion between all electron pairs, hence they constitute **two-electron integrals**

$$\begin{aligned} \left\langle \Psi \left| - \int dr' \frac{\rho(r')}{|r - r'|} \right| \Psi \right\rangle &= \left\langle \Psi \left| - \int dr' \frac{|\psi(r')|^2}{|r - r'|} \right| \Psi \right\rangle \\ &= e^2 \sum_{\substack{i,j \\ i>j}} \int dr dr' \psi_i^*(r) \psi_j^*(r') \frac{1}{|r - r'|} \psi_i(r) \psi_j(r') \end{aligned}$$

In this **two-electron integral**, anti-symmetrization i.e. **Slater determinant** plays a role.

How computer solves Hartree-Fock

The coordinates of each of $\Psi(r)$ comprises two parts – radial and spin.

$$\Psi(r) \equiv \phi(r) \chi(s) \Rightarrow \Psi(r_1, r_2) = \phi(r_1)\phi(r_2) \frac{1}{\sqrt{2}} [\chi_1(s_1)\chi_2(s_2) - \chi(s_2)\chi_2(s_1)]$$

where Hamiltonian operator does not act on spin coordinate. The Schrödinger equation now becomes,

$$\left[-\frac{1}{2m_e} \nabla_1^2 - \frac{1}{2m_e} \nabla_2^2 - \frac{2e^2}{r_1} - \frac{2e^2}{r_2} + \frac{e^2}{|r_1 - r_2|} \right] \phi(r_1)\phi(r_2) = E\phi(r_1)\phi(r_2)$$

Remove r_2 dependence by multiplying both sides from the left by $\phi^*(r_2)$ and integrating over it,

$$\left[-\frac{1}{2m_e} \nabla_1^2 - \frac{2e^2}{r_1} + \int d^3r_2 \frac{|\phi(r_2)|^2}{|r_1 - r_2|} \right] \phi(r_1) = E' \phi(r_1)$$

This has the form of self-consistency problem – ϕ is solution of Schrödinger equation but the latter is determined by ϕ itself. It is an iterative process where algorithm starts with a trial ground state wavefunction ϕ_0 which is used in constructing potential followed by solving Schrödinger equation with this potential to obtain a new ground state ϕ_1 . It is then used to build a new potential and so on.

Hartree-Fock equation is solved by introducing a set of **known spatial basis functions** to reformulate it as an algebraic equation. Suppose $\{\varphi_\mu\}$ a set of basis function (may or may not be complete),

$$\phi(r) = \sum_{\mu=1}^K C_\mu \varphi_\mu$$

$$\left[-\frac{1}{2m_e} \nabla_1^2 - \frac{2e^2}{r_1} + \sum_{\lambda, \rho} C_\lambda C_\rho \int d^3 r_2 \frac{\varphi_\lambda(r_2) \varphi_\rho(r_2)}{|r_1 - r_2|} \right] \sum_\nu C_\nu \varphi_\nu(r_1) = E' \sum_\nu C_\nu \varphi_\nu(r_1)$$

Multiplying φ_μ^* from left and integrating it, it turns into a computer friendly matrix equation,

$$[F_{\mu\nu} + C_\lambda C_\rho Q_{\lambda\rho\mu\nu}] C_\nu = E' S_{\mu\nu} C_\nu$$

where, $F_{\mu\nu} = \left\langle \varphi_\mu \left| -\frac{1}{2m_e} \nabla_1^2 - \frac{2e^2}{r_1} \right| \varphi_\nu \right\rangle$

$$S_{\mu\nu} = \langle \varphi_\mu | \varphi_\nu \rangle \quad \text{etc}$$

$$\mathcal{F}_{\mu\nu} \equiv F_{\mu\nu} + C_\lambda C_\rho Q_{\lambda\rho\mu\nu}$$

This is not a **generalized eigenvalue problem** because of presence of C_λ, C_ρ inside square bracket. One way is to keep them fixed to a guess value and solve for C_ν and use to replace them.

The non-orthogonal basis function $\{\varphi_\mu\}$ can be transformed to an orthogonal set by a transformation matrix X

$$\varphi'_\mu = \sum_\nu X_{\mu\nu} \varphi_\nu \quad \text{where} \quad \langle \varphi'_\mu | \varphi'_\nu \rangle = \delta_{\mu\nu} \Rightarrow X^\dagger S X = \mathbb{1}$$

Transformation of two-electrons integral is time consuming. Self consistent Hartree-Fock uses a new coefficient matrix C' .

$$\begin{aligned} C' &= X^\dagger C \Rightarrow \mathcal{F} X C' = E' S X C' \Rightarrow (X^\dagger \mathcal{F} X) C' = E' (X^\dagger S X) C' \\ \Rightarrow \mathcal{F}' C' &= E' C' \end{aligned}$$

Thus, the steps to solve Hartree-Fock equation self-consistently are

1. Choose trial Gaussian wavefunctions $\varphi_\mu = \exp(-\alpha_\mu r^2)$ and orthogonalize them. Making appropriate choices of α_μ is extremely non-trivial, generally relies on previously performed calculations.
2. Guess initial values for C_μ , they can all be equal or whatever. Normalize them to reduce overlap matrix to unity $C^\dagger S C = \mathbb{1}$
3. Determine F and Q and solve eigenvalue problem $\mathcal{F}' C' = E' C'$ followed by E .
4. Solution C' is then used to build new \mathcal{F} and keep iterating the steps till convergence.