On the discretization of Feynman-Kac semi-groups.

Application to rare events sampling and Diffusion Monte Carlo.

PhD Seminar

Grégoire Ferré – Gabriel Stoltz

CERMICS - ENPC

Wednesday, May 10th, 2017

Outline

1. Introduction: Two apparently unrelated problems

Schrödinger ground state

Schrödinger operator, describes the energy of a sytem:

$$H = -\Delta + V,$$

where V is a potential. Typical example:

$$V(x_1,\dots,x_N) = \sum_{i=1}^{N} V_1(x_i) + \sum_{i\neq j} V_2(x_i - x_j).$$

Schrödinger ground state

Schrödinger operator, describes the energy of a sytem:

$$H = -\Delta + V,$$

where V is a potential. Typical example:

$$V(x_1,\ldots,x_N) = \sum_{i=1}^{N} V_1(x_i) + \sum_{i\neq j} V_2(x_i - x_j).$$

Goal (for electronic structure calculation, ground state energy, properties of materials, etc), compute the **ground state energy** E_0, lowest eigenvalue

$$H\psi = E_0\psi.$$

Averages and fluctuations

Example. Consider the dynamics

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dB_t.$$

Goal: estimate a long time average

$$\varphi_t = \frac{1}{t}\int_0^t \varphi(X_s)ds \underset{t\to+\infty}{\longrightarrow} \int_{\mathcal{D}} \varphi \, d\mu, \quad \mu(dx) = Z^{-1}\mathrm{e}^{-\beta V}dx.$$

Averages and fluctuations

Example. Consider the dynamics

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dB_t.$$

Goal: estimate a long time average

$$\varphi_t = \frac{1}{t}\int_0^t \varphi(X_s)ds \underset{t\to+\infty}{\longrightarrow} \int_{\mathcal{D}} \varphi\, d\mu, \quad \mu(dx) = Z^{-1}\mathrm{e}^{-\beta V}dx.$$

Problem: for a finite time t,

$$\varphi_t \neq \mu(\varphi)$$

New goal: estimate probabilities of fluctuations around the mean.

Large deviations in time

Idea of large deviations:

$$\mathbb{P}\left[\frac{1}{t}\int_0^t \varphi(X_s)ds = a\right] \asymp \mathrm{e}^{-tI(a)},$$

where I is the **rate function**. This suggests the formula

$$I(a) = -\lim_{t\to\infty}\frac{1}{t}\log P_{\varphi_t}(a).$$

Large deviations in time

Idea of large deviations:

$$\mathbb{P}\left[\frac{1}{t}\int_0^t \varphi(X_s)ds = a\right] \asymp \mathrm{e}^{-tI(a)},$$

where I is the **rate function**. This suggests the formula

$$I(a) = -\lim_{t\to\infty} \frac{1}{t}\log P_{\varphi_t}(a).$$

Fact: $I(\mu(\varphi)) = 0$.

Large deviations in time

Idea of large deviations:

$$\mathbb{P}\left[\frac{1}{t}\int_0^t \varphi(X_s)ds = a\right] \asymp \mathrm{e}^{-tI(a)},$$

where I is the **rate function**. This suggests the formula

$$I(a) = -\lim_{t\to\infty} \frac{1}{t} \log P_{\varphi_t}(a).$$

Fact: $I(\mu(\varphi)) = 0$.

Donsker-Varadhan [1975]: if one sets

$$\lambda(k) := \sup_{a\in\mathbb{R}} \{ka - I(a)\},$$

Then $\lambda(k)$ is the largest eigenvalue of $\mathcal{L} + k\varphi$ where $\mathcal{L}$ is the generator of (X_t). We are back to a problem of **ground state** estimation.

Probabilistic representation of ground state energy

Consider the generator $\mathcal{L}$ of a process (X_t) and (λ, h_W) the principal eigenvalue and eigenfunction of $\mathcal{L} + W$. **Feynman-Kac** formula gives

$$\mathbb{E}\left[\mathrm{e}^{\int_0^t W(X_s)ds}\right] \sim \mathrm{e}^{\lambda t},$$

so that λ has the representation

$$\lambda = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\left[\mathrm{e}^{\int_0^t W(X_s)ds}\right].$$

Probabilistic representation of ground state energy

Consider the generator $\mathcal{L}$ of a process (X_t) and (λ, h_W) the principal eigenvalue and eigenfunction of $\mathcal{L} + W$. **Feynman-Kac** formula gives

$$\mathbb{E}\left[\mathrm{e}^{\int_0^t W(X_s)ds}\right] \sim \mathrm{e}^{\lambda t},$$

so that λ has the representation

$$\lambda = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\left[\mathrm{e}^{\int_0^t W(X_s)ds}\right].$$

We then turn to the following more general quantity, for an observable φ,

$$\Phi_t(\mu)(\varphi) = \frac{\mathbb{E}_\mu\left[\varphi(X_t)\mathrm{e}^{\int_0^t W(X_s)ds}\right]}{\mathbb{E}_\mu\left[\mathrm{e}^{\int_0^t W(X_s)ds}\right]} \xrightarrow[t\to\infty]{} \int_{\mathcal{D}} \varphi h_W d\nu,$$

where $(\mathcal{L} + W)h_W = \lambda h_W$.

2. Our problem: Time step bias

Statistical approximation

Path average over a set of replicas $(X_t^m)_{m=1}^M$ with initial distribution,

$$\Phi_t(\mu)(\varphi) \approx \frac{\dfrac{1}{M}\displaystyle\sum_{m=1}^{M} \varphi(X_t^m)\mathrm{e}^{\int_0^t W(X_s^m)\,ds}}{\dfrac{1}{M}\displaystyle\sum_{m=1}^{M} \mathrm{e}^{\int_0^t W(X_s^m)\,ds}}$$

Statistical approximation

Path average over a set of replicas $(X_t^m)_{m=1}^M$ with initial distribution,

$$\Phi_t(\mu)(\varphi) \approx \frac{\displaystyle\frac{1}{M}\sum_{m=1}^{M} \varphi(X_t^m)\mathrm{e}^{\int_0^t W(X_s^m)\,ds}}{\displaystyle\frac{1}{M}\sum_{m=1}^{M} \mathrm{e}^{\int_0^t W(X_s^m)\,ds}}$$

Various sampling techniques: interacting particle systems, populations dynamics...

Problem: huge variance of the exponential weights.

Discretization error

Idea

- **Fact**: it is impossible to run a continuous simulation on a computer,
- **Solution**: discretize with a time step Δt,
- **Problem**: this induces a bias in the result.

Discretization error

Idea

- **Fact**: it is impossible to run a continuous simulation on a computer,
- **Solution**: discretize with a time step Δt,
- **Problem**: this induces a bias in the result.

We discretize the process (X_t) into a Markov chain (x_n) with evolution operator $Q_{\Delta t}$ and we study quantities of the form:

$$\Phi_{n,\Delta t}(\mu)(\varphi) = \frac{\mathbb{E}_\mu\left[\varphi(x_n)\mathrm{e}^{\Delta t \sum_{i=0}^{n-1} W(x_i)}\right]}{\mathbb{E}_\mu\left[\mathrm{e}^{\Delta t \sum_{i=0}^{n-1} W(x_i)}\right]} \xrightarrow[n\to\infty]{} \int_{\mathcal{D}} \varphi \, d\nu_{W,\Delta t}.$$

Discretization error

Idea

- **Fact**: it is impossible to run a continuous simulation on a computer,
- **Solution**: discretize with a time step Δt,
- **Problem**: this induces a bias in the result.

We discretize the process (X_t) into a Markov chain (x_n) with evolution operator $Q_{\Delta t}$ and we study quantities of the form:

$$\Phi_{n,\Delta t}(\mu)(\varphi) = \frac{\mathbb{E}_\mu\left[\varphi(x_n)\mathrm{e}^{\Delta t \sum_{i=0}^{n-1} W(x_i)}\right]}{\mathbb{E}_\mu\left[\mathrm{e}^{\Delta t \sum_{i=0}^{n-1} W(x_i)}\right]} \xrightarrow[n\to\infty]{} \int_{\mathcal{D}} \varphi \, d\nu_{W,\Delta t}.$$

Natural question: is $\nu_{W,\Delta t}$ close to $\nu_W := h_W \nu$.

3. Error on the invariant measure

Discretization of a Markov process

Let's go back to the *linear* case:

$$\mathbb{E}_x[\varphi(X_t)] \underset{t\to\infty}{\longrightarrow} \int_{\mathcal{D}} \varphi \, d\nu.$$

The Markov process (X_t) is discretized into a Markov chain (x_n) with evolution operator

$$Q_{\Delta t}(\varphi)(x) = \mathbb{E}\left[\varphi(x_{n+1}) | x_n = x\right]$$

Discretization of a Markov process

Let's go back to the *linear* case:

$$\mathbb{E}_x[\varphi(X_t)] \xrightarrow[t\to\infty]{} \int_{\mathcal{D}} \varphi \, d\nu.$$

The Markov process (X_t) is discretized into a Markov chain (x_n) with evolution operator

$$Q_{\Delta t}(\varphi)(x) = \mathbb{E}\left[\varphi(x_{n+1}) | x_n = x\right]$$

We then have for a long time average of the Markov chain

$$\mathbb{E}_x[\varphi(x_n)] \xrightarrow[t\to\infty]{} \int_{\mathcal{D}} \varphi \, d\nu_{\Delta t}.$$

Discretization of a Markov process

Let's go back to the *linear* case:

$$\mathbb{E}_x[\varphi(X_t)] \xrightarrow[t\to\infty]{} \int_{\mathcal{D}} \varphi \, d\nu.$$

The Markov process (X_t) is discretized into a Markov chain (x_n) with evolution operator

$$Q_{\Delta t}(\varphi)(x) = \mathbb{E}\left[\varphi(x_{n+1})|x_n = x\right]$$

We then have for a long time average of the Markov chain

$$\mathbb{E}_x[\varphi(x_n)] \xrightarrow[t\to\infty]{} \int_{\mathcal{D}} \varphi \, d\nu_{\Delta t}.$$

Question: how can we relate ν and $\nu_{\Delta t}$?

Error on the invariant measure

Idea: $Q_{\Delta t}$ approximates the exact flow $e^{t\mathcal{L}}$ over a time step Δt

Expansion of the operator

We assume that:

$$Q_{\Delta t}\varphi = \varphi + \Delta t \mathcal{A}_1\varphi + \Delta t^2 \mathcal{A}_2\varphi + \ldots + \Delta t^p \mathcal{A}_p\varphi + \Delta t^{p+1}\mathcal{A}_{p+1}\varphi + O(\Delta t^{p+2})$$

Theorem

Under «mild» assumptions, if for $k = 1, \ldots, p$ we have $\forall\, \varphi \in \mathcal{C}^\infty, \quad \int_{\mathcal{D}} \mathcal{A}_k\varphi \, d\nu = 0$, then

$$\underbrace{\int_{\mathcal{D}} \varphi \, d\nu_{\Delta t}}_{\text{approximate average}} = \underbrace{\int_{\mathcal{D}} \varphi \, d\nu}_{\text{correct average}} + \underbrace{\Delta t^p}_{\text{order}} \underbrace{\int_{\mathcal{D}} \varphi f \, d\nu}_{\text{correction term}} + O(\Delta t^{p+1})$$

Error on the invariant measure

Idea: $Q_{\Delta t}$ approximates the exact flow $e^{t\mathcal{L}}$ over a time step Δt

Expansion of the operator

We assume that:

$$Q_{\Delta t}\varphi = \varphi + \Delta t \mathcal{A}_1\varphi + \Delta t^2 \mathcal{A}_2\varphi + \ldots + \Delta t^p \mathcal{A}_p\varphi + \Delta t^{p+1}\mathcal{A}_{p+1}\varphi + O(\Delta t^{p+2})$$

Theorem

Under «mild» assumptions, if for $k = 1, \ldots, p$ we have $\forall\, \varphi \in \mathcal{C}^\infty, \quad \int_{\mathcal{D}} \mathcal{A}_k \varphi \, d\nu = 0$, then

$$\underbrace{\int_{\mathcal{D}} \varphi \, d\nu_{\Delta t}}_{\text{approximate average}} = \underbrace{\int_{\mathcal{D}} \varphi \, d\nu}_{\text{correct average}} + \underbrace{\Delta t^p}_{\text{order}} \underbrace{\int_{\mathcal{D}} \varphi f \, d\nu}_{\text{correction term}} + O(\Delta t^{p+1})$$

Error on the invariant measure

Idea: $Q_{\Delta t}$ approximates the exact flow $\mathrm{e}^{t\mathcal{L}}$ over a time step Δt

Expansion of the operator

We assume that:

$$Q_{\Delta t}\varphi = \varphi + \Delta t \mathcal{A}_1\varphi + \Delta t^2 \mathcal{A}_2\varphi + \ldots + \Delta t^p \mathcal{A}_p\varphi + \Delta t^{p+1}\mathcal{A}_{p+1}\varphi + O(\Delta t^{p+2})$$

Theorem

Under «mild» assumptions, if for $k = 1, \ldots, p$ we have $\forall\, \varphi \in \mathcal{C}^\infty, \quad \int_{\mathcal{D}} \mathcal{A}_k \varphi \, d\nu = 0$, then

$$\underbrace{\int_{\mathcal{D}} \varphi \, d\nu_{\Delta t}}_{\text{approximate average}} = \underbrace{\int_{\mathcal{D}} \varphi \, d\nu}_{\text{correct average}} + \underbrace{\Delta t^p}_{\text{order}} \underbrace{\int_{\mathcal{D}} \varphi f \, d\nu}_{\text{correction term}} + O(\Delta t^{p+1})$$

Error on the invariant measure

Idea: $Q_{\Delta t}$ approximates the exact flow $\mathrm{e}^{t\mathcal{L}}$ over a time step Δt

Expansion of the operator

We assume that:

$$Q_{\Delta t}\varphi = \varphi + \Delta t \mathcal{A}_1\varphi + \Delta t^2 \mathcal{A}_2\varphi + \ldots + \Delta t^p \mathcal{A}_p\varphi + \Delta t^{p+1}\mathcal{A}_{p+1}\varphi + O(\Delta t^{p+2})$$

Theorem

Under «mild» assumptions, if for $k = 1, \ldots, p$ we have
$\forall\, \varphi \in \mathcal{C}^\infty, \quad \int_{\mathcal{D}} \mathcal{A}_k \varphi \, d\nu = 0$, then

$$\underbrace{\int_{\mathcal{D}} \varphi \, d\nu_{\Delta t}}_{\text{approximate average}} = \underbrace{\int_{\mathcal{D}} \varphi \, d\nu}_{\text{correct average}} + \underbrace{\Delta t^p}_{\text{order}} \underbrace{\int_{\mathcal{D}} \varphi f \, d\nu}_{\text{correction term}} + O(\Delta t^{p+1})$$

Question: can we extend this strategy to our case ?

4. Discretizing Feynman-Kac semi-groups

A double discretization

Idea: we need to discretize the flow of

$$\mathbb{E}\left[\varphi(X_t)\mathrm{e}^{\int_0^t W(X_s)ds}\right] = \mathrm{e}^{t(\mathcal{L}+W)}\varphi$$

over a time step Δt.

A double discretization

Idea: we need to discretize the flow of

$$\mathbb{E}\left[\varphi(X_t)\mathrm{e}^{\int_0^t W(X_s)ds}\right] = \mathrm{e}^{t(\mathcal{L}+W)}\varphi$$

over a time step Δt.

Fact: we need to discretize both the process *and* the integral. We denote $Q_{\Delta t}^W$ the approximated flow, for example

$$(Q_{\Delta t}^W\varphi)(x) = \mathrm{e}^{\Delta t W(x)}(Q_{\Delta t}\varphi)(x), \quad (Q_{\Delta t}^W\varphi)(x) = \mathrm{e}^{\frac{\Delta t}{2}W(x)}(Q_{\Delta t}\varphi\mathrm{e}^{\frac{\Delta t}{2}W})(x).$$

A double discretization

Idea: we need to discretize the flow of

$$\mathbb{E}\left[\varphi(X_t)\mathrm{e}^{\int_0^t W(X_s)\,ds}\right] = \mathrm{e}^{t(\mathcal{L}+W)}\varphi$$

over a time step Δt.

Fact: we need to discretize both the process *and* the integral. We denote $Q_{\Delta t}^W$ the approximated flow, for example

$$(Q_{\Delta t}^W\varphi)(x) = \mathrm{e}^{\Delta t W(x)}(Q_{\Delta t}\varphi)(x), \quad (Q_{\Delta t}^W\varphi)(x) = \mathrm{e}^{\frac{\Delta t}{2}W(x)}(Q_{\Delta t}\varphi \mathrm{e}^{\frac{\Delta t}{2}W})(x).$$

Same strategy for the long time error, show that

$$Q_{\Delta t}^W \approx \mathrm{e}^{\Delta t(\mathcal{L}+W)}$$

with an expansion in Δt.

Main result

Expansion of the operator

We assume that:

$$Q_{\Delta t}^{W}\varphi = \varphi + \Delta t \widetilde{\mathcal{A}}_1\varphi + \Delta t^2 \widetilde{\mathcal{A}}_2\varphi + \ldots + \Delta t^p \widetilde{\mathcal{A}}_p\varphi + \Delta t^{p+1}\widetilde{\mathcal{A}}_{p+1}\varphi + O(\Delta t^{p+2})$$

Theorem: error on the invariant measure

Under «mild» assumptions, if for $k = 1, \ldots, p$ there exists $a_k \in \mathbb{R}$ s.t.

$$\forall\, \varphi \in \mathcal{C}^\infty, \quad \int_{\mathcal{D}} \widetilde{\mathcal{A}}_k \varphi \, d\nu_W = a_k \int_{\mathcal{D}} \varphi \, d\nu_W,$$

then

$$\underbrace{\int_{\mathcal{D}} \varphi \, d\nu_{W,\Delta t}}_{\text{approximate average}} = \underbrace{\int_{\mathcal{D}} \varphi \, d\nu_W}_{\text{correct average}} + \underbrace{\Delta t^p}_{\text{order}} \underbrace{\int_{\mathcal{D}} \varphi f \, d\nu_W}_{\text{correction term}} + O(\Delta t^{p+1})$$

Main result

Expansion of the operator

We assume that:

$$Q_{\Delta t}^{W}\varphi = \varphi+\Delta t\widetilde{\mathcal{A}}_1\varphi+\Delta t^2\widetilde{\mathcal{A}}_2\varphi+\ldots+\Delta t^p\widetilde{\mathcal{A}}_p\varphi+\Delta t^{p+1}\widetilde{\mathcal{A}}_{p+1}\varphi+O(\Delta t^{p+2})$$

Theorem: error on the invariant measure

Under «mild» assumptions, if for $k=1,\ldots,p$ there exists $a_k\in\mathbb{R}$ s.t.

$$\forall\,\varphi\in\mathcal{C}^\infty,\quad \int_{\mathcal{D}}\widetilde{\mathcal{A}}_k\varphi\,d\nu_W = a_k\int_{\mathcal{D}}\varphi\,d\nu_W,$$

then

$$\underbrace{\int_{\mathcal{D}}\varphi\,d\nu_{W,\Delta t}}_{\text{approximate average}} = \underbrace{\int_{\mathcal{D}}\varphi\,d\nu_W}_{\text{correct average}} + \underbrace{\Delta t^p}_{\text{order}}\underbrace{\int_{\mathcal{D}}\varphi f\,d\nu_W}_{\text{correction term}} + O(\Delta t^{p+1})$$

Useful corollary

We are mainly interested in the eigenvalue λ. We have:

$$\frac{1}{\Delta t}\log\left[\int_{\mathcal{D}} \mathrm{e}^{\Delta t W}\, d\nu_{W,\Delta t}\right] \approx \lambda.$$

Useful corollary

We are mainly interested in the eigenvalue λ. We have:

$$\frac{1}{\Delta t}\log\left[\int_{\mathcal{D}} \mathrm{e}^{\Delta t W} d\nu_{W,\Delta t}\right] \approx \lambda.$$

More precisely

Theorem: Eigenvalue as a partition function

If $Q_{\Delta t}^{W}$ verifies the hypotheses of last theorem, then

$$\lambda_{\Delta t} := \frac{1}{\Delta t}\log\left[\int_{\mathcal{D}} Q_{\Delta t}^{W}(\mathbb{1}) d\nu_{W,\Delta t}\right] = \lambda + \Delta t^{p} C + O(\Delta t^{p+1}),$$

where $C \in \mathbb{R}$ is a constant depending on f.

Important consequences

Statistical physics

If the dynamics is discretized with a second order scheme $Q_{\Delta t}$, then the splitting

$$Q_{\Delta t}^{W} = \mathrm{e}^{\frac{\Delta t}{2}W}\Big(Q_{\Delta t}(\mathrm{e}^{\frac{\Delta t}{2}W}\cdot)\Big)$$

provides a second-order discretization of the Feynman-Kac semi-group.

Diffusion Monte Carlo

In this case, the dynamics (X_t) is a brownian motion, so the flow $Q_{\Delta t}$ is always *exact*. The order of the scheme is then the order of discretization of the integral.

Important consequences

Statistical physics

If the dynamics is discretized with a second order scheme $Q_{\Delta t}$, then the splitting

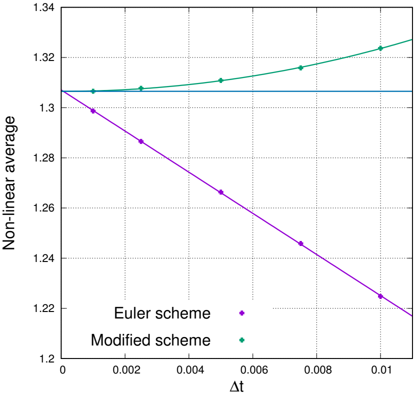
$$Q_{\Delta t}^{W} = \mathrm{e}^{\frac{\Delta t}{2}W}\left(Q_{\Delta t}(\mathrm{e}^{\frac{\Delta t}{2}W}\cdot)\right)$$

provides a second-order discretization of the Feynman-Kac semi-group.

Diffusion Monte Carlo

In this case, the dynamics (X_t) is a brownian motion, so the flow $Q_{\Delta t}$ is always *exact*. The order of the scheme is then the order of discretization of the integral.

Overdamped Langevin dynamics on a one dimensional torus.



Estimation of $\lambda_{\Delta t}$ for:

- $dX_t = -V'(X_t)dt + dB_t$,
- $V(x) = cos(2\pi x)$,
- $W = |V|^2$,
- Euler-Maruyama scheme and 2nd order modified scheme.

We indeed observe first and second order convergence.

Conclusion & Tracks

Conlusion

Results:

- error estimates on the invariant measure of Feynman-Kac semi-groups,
- alternative representation of the principal eigenvalue of a Schrödinger operator,
- immediate applications to rare events sampling and Diffusion Monte Carlo.

Future works

- unbounded state-space, degenerate dynamics,
- singular potentials,
- adaptative scheme.

Conclusion & Tracks

Conlusion

Results:

- error estimates on the invariant measure of Feynman-Kac semi-groups,
- alternative representation of the principal eigenvalue of a Schrödinger operator,
- immediate applications to rare events sampling and Diffusion Monte Carlo.

Future works

- unbounded state-space, degenerate dynamics,
- singular potentials,
- adaptative scheme.