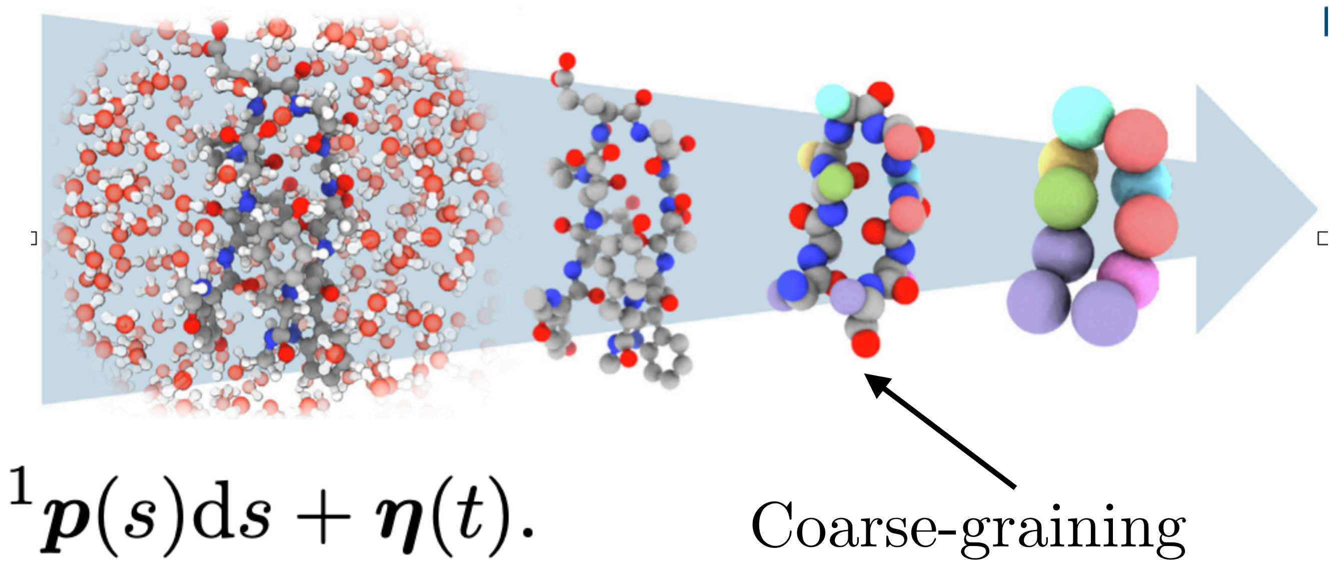
Dynamics with Memory (e.g., Generalized Langevin Equation)



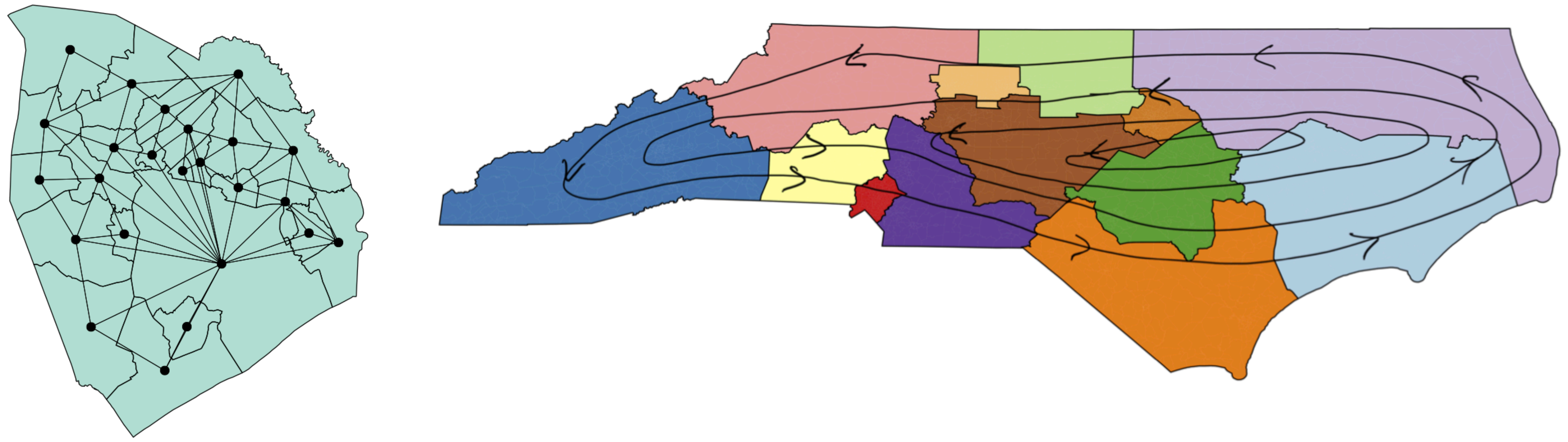
$$\dot{\boldsymbol{q}} = \boldsymbol{M}^{-1}\boldsymbol{p},$$

$$\dot{\boldsymbol{p}} = -\nabla_{\boldsymbol{q}} U(\boldsymbol{q}) - \int_0^t \boldsymbol{K}(t-s)\boldsymbol{M}^{-1}\boldsymbol{p}(s)\mathrm{d}s + \boldsymbol{\eta}(t).$$

Memory-kernel

Time-correlated noise

Sampling Graph Partitions and Redistricting maps to quantify Gerrymandering



Adaptive Langevin Dynamics

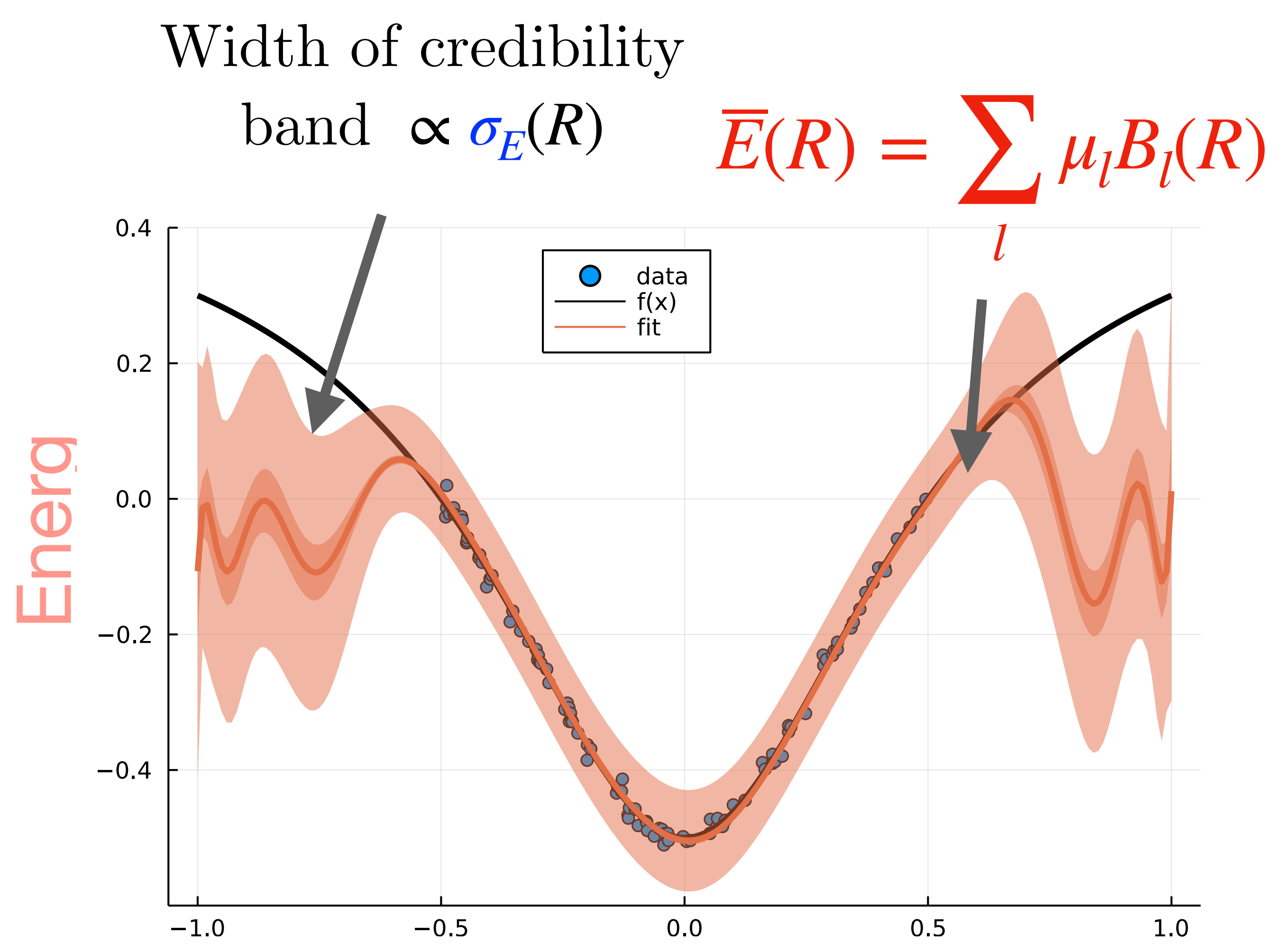
$$\begin{aligned} \mathrm{d}\boldsymbol{\theta} &= \boldsymbol{\theta}\,\mathrm{d}t, \\ \mathrm{d}\boldsymbol{v} &= \nabla \log \pi(\boldsymbol{\theta})\,\mathrm{d}t + \sqrt{h}\,\sigma\,\mathrm{d}W - \xi\,\boldsymbol{v}\,\mathrm{d}t, \\ \mathrm{d}\xi &= \frac{1}{\nu}\left(|\boldsymbol{v}|^2 - p\right)\mathrm{d}t. \end{aligned}$$

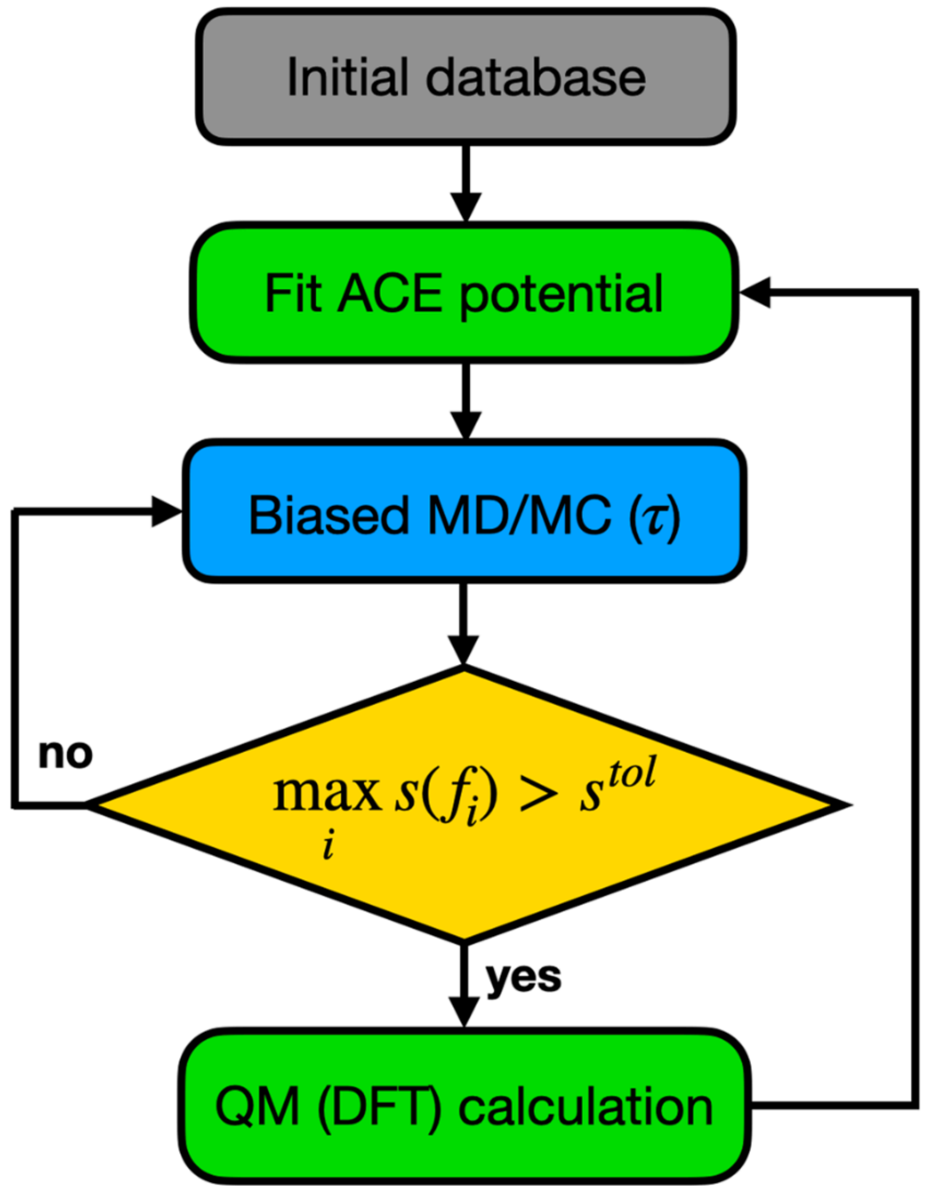
Ergodic with invariant measure:

$$\widetilde{\pi}(\theta, \boldsymbol{v}, \xi) \propto \pi(\theta) e^{-\frac{1}{2}|\boldsymbol{v}|^2} e^{-\frac{\nu}{2}(\xi - h\sigma^2)^2}$$

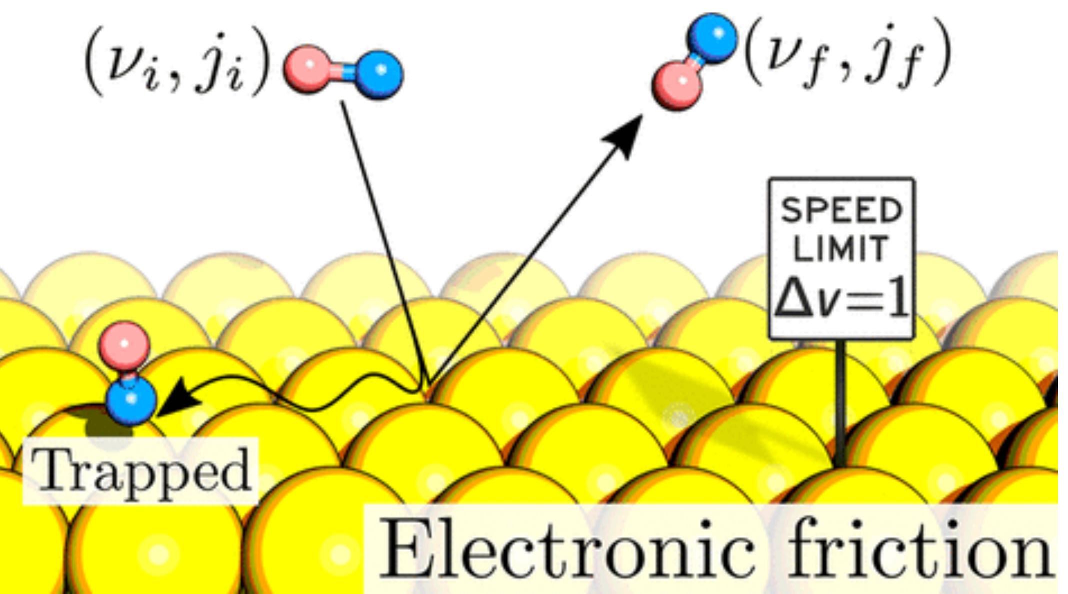
Convergence rates: $\lambda_{\nu,\gamma} \geq \overline{\lambda} \min\left(\gamma\nu, \frac{1}{\gamma}, \frac{\nu}{\gamma}, \frac{\gamma}{\nu}\right), \quad \gamma = h\sigma^2/2$

Hyperactive Learning for Machine-Learned Interatomic Potentials





Machine Learning for Quantum Molecular Dynamics (Electronic Friction)



$$\dot{\boldsymbol{r}}_i = m_i^{-1}\boldsymbol{p}_i,$$

$$\dot{\boldsymbol{p}}_i = -\nabla_{\boldsymbol{r}_i}\mathcal{E}((\boldsymbol{r}_\iota, z_\iota)_\iota) - \sum_{j=1}^{N} \boldsymbol{\Gamma}_{ij}((\boldsymbol{r}_\iota, z_\iota)_\iota)\boldsymbol{p}_j/m_j + \sqrt{2\beta^{-1}}\boldsymbol{\Sigma}_i((\boldsymbol{r}_\iota, z_\iota)_\iota)\dot{\boldsymbol{W}}$$

Posterior Computation with the Gibbs Zig-Zag Sampler

Joint work with Deborshee Sen, Jianfeng Lu, David Dunson

Matthias Sachs

University of Birmingham

Algorithms and Computationally Intensive Inference Seminar, University of Warwick

June 21, 2024

Outline

1. Background
   - Continuous-time Monte Carlo
   - The Zig-zag sampler (ZZ)
   - Bayesian hierarchical models
2. The Gibbs Zig-zag sampler (GZZ)
   - Construction
   - Theoretical properties
3. Application to posterior sampling problems
   - Random effect model
   - Logistic regression with Spike-and-Slab Prior

Outline

1. Background
   - Continuous-time Monte Carlo
   - The Zig-zag sampler (ZZ)
   - Bayesian hierarchical models
2. The Gibbs Zig-zag sampler (GZZ)
   - Construction
   - Theoretical properties
3. Application to posterior sampling problems
   - Random effect model
   - Logistic regression with Spike-and-Slab Prior

Computation of expectations

$$\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)] = \int_{\mathbb{R}^d} \varphi(\zeta)\pi(\mathrm{d}\zeta)$$

with

- probability measure π known up to a normalization constant.
- φ some π-integrable real valued function (aka “observable”).
- number of dimensions, d, of integration domain “large”

Monte Carlo approximations

- **Markov chain Monte Carlo:**

$$\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)] \approx \frac{1}{N}\sum_{k=0}^{N-1}\varphi(\zeta_k)$$

  with $(\zeta_k)_{k\in\mathbb{N}} \subset \mathbb{R}^d$ ergodic Markov chain with unique invariant measure π.

- **Continuous time Monte Carlo:**

$$\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)] \approx \frac{1}{T}\int_0^T \varphi(\zeta(t))\mathrm{d}t$$

  with $(\zeta(t))_{t\geq 0} \subset \mathbb{R}^d$ ergodic stochastic (Markov-)process with unique invariant measure π.

Computation of expectations

$$\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)] = \int_{\mathbb{R}^d} \varphi(\zeta)\pi(\mathrm{d}\zeta)$$

with

- probability measure π known up to a normalization constant.
- φ some π-integrable real valued function (aka "observable").
- number of dimensions, d, of integration domain "large"

Monte Carlo approximations

- **Markov chain Monte Carlo:**

$$\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)] \approx \frac{1}{N}\sum_{k=0}^{N-1}\varphi(\boldsymbol{\zeta}_k)$$

with $(\boldsymbol{\zeta}_k)_{k\in\mathbb{N}} \subset \mathbb{R}^d$ ergodic Markov chain with unique invariant measure π.

- **Continuous time Monte Carlo:**

$$\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)] \approx \frac{1}{T}\int_0^T \varphi(\boldsymbol{\zeta}(t))\mathrm{d}t$$

with $(\boldsymbol{\zeta}(t))_{t\geq 0} \subset \mathbb{R}^d$ ergodic stochastic (Markov-)process with unique invariant measure π.

Computation of expectations

$$\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)] = \int_{\mathbb{R}^d} \varphi(\zeta)\pi(\mathrm{d}\zeta)$$

with

- probability measure π known up to a normalization constant.
- φ some π-integrable real valued function (aka "observable").
- number of dimensions, d, of integration domain "large"

Monte Carlo approximations

- **Markov chain Monte Carlo:**

$$\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)] \approx \frac{1}{N}\sum_{k=0}^{N-1} \varphi(\boldsymbol{\zeta}_k)$$

  with $(\boldsymbol{\zeta}_k)_{k\in\mathbb{N}} \subset \mathbb{R}^d$ ergodic Markov chain with unique invariant measure π.

- **Continuous time Monte Carlo:**

$$\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)] \approx \frac{1}{T}\int_0^T \varphi(\boldsymbol{\zeta}(t))\mathrm{d}t$$

  with $(\boldsymbol{\zeta}(t))_{t\geq 0} \subset \mathbb{R}^d$ ergodic stochastic (Markov-)process with unique invariant measure π.

Computation of expectations

$$\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)] = \int_{\mathbb{R}^d} \varphi(\zeta)\pi(\mathrm{d}\zeta)$$

with

- probability measure π known up to a normalization constant.
- φ some π-integrable real valued function (aka "observable").
- number of dimensions, d, of integration domain "large"

Monte Carlo approximations

- **Markov chain Monte Carlo:**

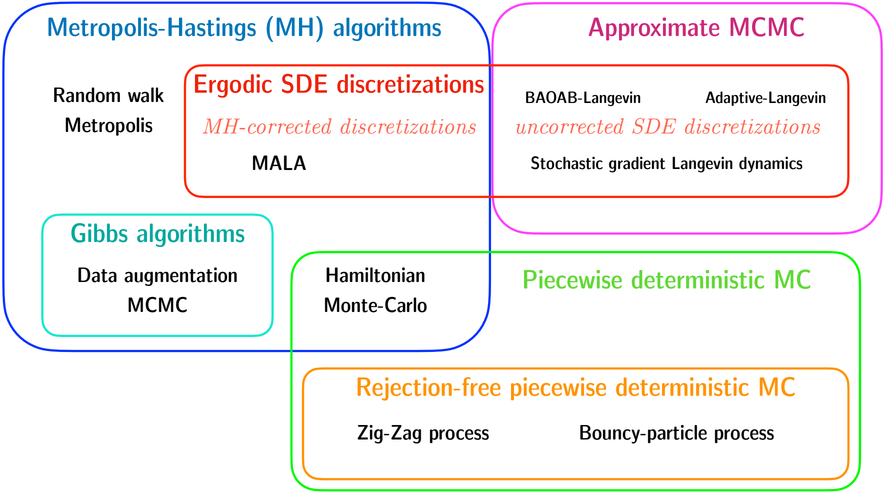
$$\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)] \approx \frac{1}{N}\sum_{k=0}^{N-1} \varphi(\boldsymbol{\zeta}_k)$$

  with $(\boldsymbol{\zeta}_k)_{k\in\mathbb{N}} \subset \mathbb{R}^d$ ergodic Markov chain with unique invariant measure π.
- **Continuous time Monte Carlo:**

$$\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)] \approx \frac{1}{T}\int_0^T \varphi(\boldsymbol{\zeta}(t))\mathrm{d}t$$

  with $(\boldsymbol{\zeta}(t))_{t\geq 0} \subset \mathbb{R}^d$ ergodic stochastic (Markov-)process with unique invariant measure π.

An (incomplete) map of the Monte Carlo world

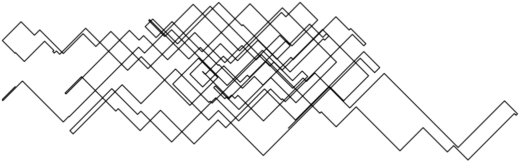


Zig-zag process: construction

- the process is defined on the **augmented space** $\mathbb{R}^d \times \{-1, 1\}^d$, and is **continuous in time**, i.e.,

$$(\boldsymbol{\zeta}(t), \boldsymbol{\theta}(t))_{t \geq 0} \subset \mathbb{R}^d \times \{-1, 1\}^d, \tag{1}$$

  we refer to
  - $\boldsymbol{\zeta}(t)$ as the **position vector** of the process
  - $\boldsymbol{\theta}(t)$ as the **velocity vector** of the process
- signs of components of the velocity vector are flipped at random event times sampled from a non-homogenous Poisson process
- the process evolves linearly as $\dot{\boldsymbol{\zeta}} = \boldsymbol{\theta}$ between event times.



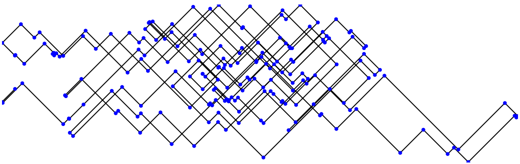
Trace of $\boldsymbol{\zeta}(t) = (\boldsymbol{\zeta}_1(t), \boldsymbol{\zeta}_2(t))$, $t \geq 0$.

Zig-zag process: construction

- the process is defined on the **augmented space** $\mathbb{R}^d \times \{-1, 1\}^d$, and is **continuous in time**, i.e.,

$$(\boldsymbol{\zeta}(t), \boldsymbol{\theta}(t))_{t \geq 0} \subset \mathbb{R}^d \times \{-1, 1\}^d, \quad (1)$$

  we refer to
  - $\boldsymbol{\zeta}(t)$ as the **position vector** of the process
  - $\boldsymbol{\theta}(t)$ as the **velocity vector** of the process
- signs of components of the velocity vector are flipped at random event times sampled from a non-homogenous Poisson process
- the process evolves linearly as $\dot{\boldsymbol{\zeta}} = \boldsymbol{\theta}$ between event times.



Trace of $\boldsymbol{\zeta}(t) = (\boldsymbol{\zeta}_1(t), \boldsymbol{\zeta}_2(t))$, $t \geq 0$.

Zig-zag process: algorithm

Algorithm

Input $T^{(0)}, \boldsymbol{\zeta}^{(0)}, \boldsymbol{\theta}^{(0)}$.

For $k = 1, 2, 3, \ldots$

1. Compute bouncing time:
   - Draw $\tau_1, \ldots, \tau_d$ such that

     $$\mathbb{P}(\tau_i \geq t) = \exp\left\{ -\int_0^t m_i(s)\,\mathrm{d}s \right\}.$$

   - $i_0 = \operatorname{argmin}_i \{\tau_i\}$.
2. Evolve position:
   $$(T^{k+1}, \boldsymbol{\zeta}^{k+1}) \leftarrow (T^k + \tau_{i_0}, \boldsymbol{\zeta}^k + \boldsymbol{\theta}^k \tau_{i_0}),$$
   $$\boldsymbol{\theta}^{k+1} \leftarrow \boldsymbol{\theta}^k.$$
3. Bounce: $\boldsymbol{\theta}_{i_0}{}^{k+1} \leftarrow -\boldsymbol{\theta}_{i_0}{}^{k}$.

Output $\left(T^k, \boldsymbol{\zeta}^k, \boldsymbol{\theta}^k\right)_{k=0,1,2,\ldots}$.

Example trajectory

Zig-zag process: algorithm

Algorithm

Input $T^{(0)}, \boldsymbol{\zeta}^{(0)}, \boldsymbol{\theta}^{(0)}$.

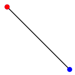
For $k = 1, 2, 3, \ldots$

1. Compute bouncing time:
   - Draw $\tau_1, \ldots, \tau_d$ such that

     $$\mathbb{P}(\tau_i \geq t) = \exp\left\{ -\int_0^t m_i(s)\,\mathrm{d}s \right\}.$$

   - $i_0 = \operatorname{argmin}_i \{\tau_i\}$.
2. Evolve position:
   $(T^{k+1}, \boldsymbol{\zeta}^{k+1}) \leftarrow (T^k + \tau_{i_0}, \boldsymbol{\zeta}^k + \boldsymbol{\theta}^k \tau_{i_0})$,
   $\boldsymbol{\theta}^{k+1} \leftarrow \boldsymbol{\theta}^k$.
3. Bounce: ${\boldsymbol{\theta}_{i_0}}^{k+1} \leftarrow -{\boldsymbol{\theta}_{i_0}}^{k}$.

Output $\left(T^k, \boldsymbol{\zeta}^k, \boldsymbol{\theta}^k\right)_{k=0,1,2,\ldots}$.



Example trajectory

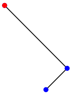
Zig-zag process: algorithm

Algorithm

Input $T^{(0)}, \boldsymbol{\zeta}^{(0)}, \boldsymbol{\theta}^{(0)}$.
For $k = 1, 2, 3, \dots$

1. Compute bouncing time:
   - Draw $\tau_1, \dots, \tau_d$ such that
   $$\mathbb{P}(\tau_i \geq t) = \exp\left\{-\int_0^t m_i(s)\,\mathrm{d}s\right\}.$$
   - $i_0 = \mathrm{argmin}_i\{\tau_i\}$.
2. Evolve position:
   $$(T^{k+1}, \boldsymbol{\zeta}^{k+1}) \leftarrow (T^k + \tau_{i_0}, \boldsymbol{\zeta}^k + \boldsymbol{\theta}^k \tau_{i_0}),$$
   $$\boldsymbol{\theta}^{k+1} \leftarrow \boldsymbol{\theta}^k.$$
3. Bounce: ${\boldsymbol{\theta}_{i_0}}^{k+1} \leftarrow -{\boldsymbol{\theta}_{i_0}}^{k}$.

Output $(T^k, \boldsymbol{\zeta}^k, \boldsymbol{\theta}^k)_{k=0,1,2,\dots}$.



Example trajectory

Zig-zag process: algorithm

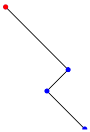
Algorithm

Input $T^{(0)}, \boldsymbol{\zeta}^{(0)}, \boldsymbol{\theta}^{(0)}$.

For $k = 1, 2, 3, \ldots$

1. Compute bouncing time:
   - Draw $\tau_1, \ldots, \tau_d$ such that
   $$\mathbb{P}(\tau_i \geq t) = \exp\left\{-\int_0^t m_i(s)\,\mathrm{d}s\right\}.$$
   - $i_0 = \operatorname{argmin}_i\{\tau_i\}$.
2. Evolve position:
   $$(T^{k+1}, \boldsymbol{\zeta}^{k+1}) \leftarrow (T^k + \tau_{i_0}, \boldsymbol{\zeta}^k + \boldsymbol{\theta}^k \tau_{i_0}),$$
   $$\boldsymbol{\theta}^{k+1} \leftarrow \boldsymbol{\theta}^k.$$
3. Bounce: ${\boldsymbol{\theta}_{i_0}}^{k+1} \leftarrow -{\boldsymbol{\theta}_{i_0}}^k$.

Output $\left(T^k, \boldsymbol{\zeta}^k, \boldsymbol{\theta}^k\right)_{k=0,1,2,\ldots}$.



Example trajectory

Zig-zag process: algorithm

Algorithm

Input $T^{(0)}, \boldsymbol{\zeta}^{(0)}, \boldsymbol{\theta}^{(0)}$.

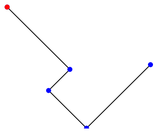
For $k = 1, 2, 3, \ldots$

1. Compute bouncing time:
   - Draw $\tau_1, \ldots, \tau_d$ such that

     $$\mathbb{P}(\tau_i \geq t) = \exp\left\{-\int_0^t m_i(s)\,\mathrm{d}s\right\}.$$

   - $i_0 = \mathrm{argmin}_i\{\tau_i\}$.
2. Evolve position:
   $(T^{k+1}, \boldsymbol{\zeta}^{k+1}) \leftarrow (T^k + \tau_{i_0}, \boldsymbol{\zeta}^k + \boldsymbol{\theta}^k \tau_{i_0})$,
   $\boldsymbol{\theta}^{k+1} \leftarrow \boldsymbol{\theta}^k$.
3. Bounce: ${\boldsymbol{\theta}_{i_0}}^{k+1} \leftarrow -{\boldsymbol{\theta}_{i_0}}^k$.

Output $\left(T^k, \boldsymbol{\zeta}^k, \boldsymbol{\theta}^k\right)_{k=0,1,2,\ldots}$.



Example trajectory

Zig-zag process: algorithm

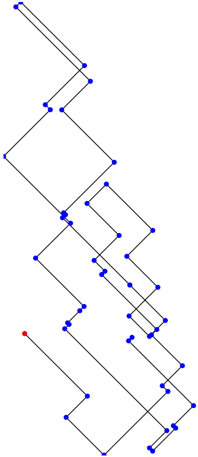
Algorithm

Input $T^{(0)}, \boldsymbol{\zeta}^{(0)}, \boldsymbol{\theta}^{(0)}$.

For $k = 1, 2, 3, \dots$

1. Compute bouncing time:
   - Draw $\tau_1, \dots, \tau_d$ such that
   $$\mathbb{P}(\tau_i \geq t) = \exp\left\{ -\int_0^t m_i(s)\, \mathrm{d}s \right\}.$$
   - $i_0 = \operatorname{argmin}_i \{\tau_i\}$.
2. Evolve position:
   $(T^{k+1}, \boldsymbol{\zeta}^{k+1}) \leftarrow (T^k + \tau_{i_0}, \boldsymbol{\zeta}^k + \boldsymbol{\theta}^k \tau_{i_0})$,
   $\boldsymbol{\theta}^{k+1} \leftarrow \boldsymbol{\theta}^k$.
3. Bounce: ${\boldsymbol{\theta}_{i_0}}^{k+1} \leftarrow -{\boldsymbol{\theta}_{i_0}}^{k}$.

Output $\left(T^k, \boldsymbol{\zeta}^k, \boldsymbol{\theta}^k\right)_{k=0,1,2,\dots}$.



Example trajectory

Zig-zag process: ergodic properties

- Probability measure: $\pi(\mathrm{d}\zeta) \propto e^{-U(\zeta)}\mathrm{d}\zeta$, $\quad U$ "Potential function"

Theorem [Bierkens et al., 2016]

If $m_i(s) = \lambda_i\left(\boldsymbol{\zeta}(t^k+s), \boldsymbol{\theta}(t^k+s)\right)$ with

$$\lambda_i(\zeta,\theta) = \left\{\theta_i \partial_{\zeta_i} U(\zeta)\right\}^+ + \underbrace{\gamma_i(\zeta)}_{\geq 0 \text{ refreshment rate}},$$

then

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \varphi(\boldsymbol{\zeta}(t))\mathrm{d}t = \mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)], \text{ almost surely} \tag{2}$$

for all $\varphi \in L^2(\pi)$.

$\Rightarrow$ For finite $T > 0$, the trajectory average $\frac{1}{T}\int_0^T \varphi(\boldsymbol{\zeta}(t))\mathrm{d}t$ may be used as a Monte-Carlo estimate of $\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)]$.

Zig-zag process: ergodic properties

- Probability measure: $\pi(\mathrm{d}\zeta) \propto e^{-U(\zeta)}\mathrm{d}\zeta$, $\quad U$ "Potential function"

Theorem [Bierkens et al., 2016]

If $m_i(s) = \lambda_i\left(\boldsymbol{\zeta}(t^k + s), \boldsymbol{\theta}(t^k + s)\right)$ with

$$\lambda_i(\zeta, \theta) = \left\{\theta_i \partial_{\zeta_i} U(\zeta)\right\}^+ + \underbrace{\gamma_i(\zeta)}_{\geq 0 \text{ refreshment rate}},$$

then

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \varphi(\boldsymbol{\zeta}(t))\mathrm{d}t = \mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)], \text{ almost surely} \tag{2}$$

for all $\varphi \in L^2(\pi)$.

$\Rightarrow$ For finite $T > 0$, the trajectory average $\frac{1}{T}\int_0^T \varphi(\boldsymbol{\zeta}(t))\mathrm{d}t$ may be used as a Monte-Carlo estimate of $\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)]$.

Zig-zag process: ergodic properties

- Probability measure: $\pi(\mathrm{d}\zeta) \propto e^{-U(\zeta)}\mathrm{d}\zeta$, $\quad U$ "Potential function"

Theorem [Bierkens et al., 2016]

If $m_i(s) = \lambda_i\left(\boldsymbol{\zeta}(t^k+s), \boldsymbol{\theta}(t^k+s)\right)$ with

$$\lambda_i(\zeta,\theta) = \left\{\theta_i \partial_{\zeta_i} U(\zeta)\right\}^+ + \underbrace{\gamma_i(\zeta)}_{\geq 0 \text{ refreshment rate}},$$

then

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \varphi(\boldsymbol{\zeta}(t))\mathrm{d}t = \mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)], \text{ almost surely} \tag{2}$$

for all $\varphi \in L^2(\pi)$.

$\Rightarrow$ For finite $T > 0$, the trajectory average $\frac{1}{T}\int_0^T \varphi(\boldsymbol{\zeta}(t))\mathrm{d}t$ may be used as a Monte-Carlo estimate of $\mathbb{E}_{\zeta\sim\pi}[\varphi(\zeta)]$.

Zig-zag process: nice properties ☺

- the Zig-Zag process is a non-reversible stochastic process
  - ⇒ non-diffusive (kinetic-like) dynamics
  - ⇒ better mixing
- can be modified so as to allow (data) sub-sampling without the introduction of any systematic bias:
  - Potential function: $U(\zeta) = \frac{1}{n}\sum_{j=1}^{n} U^j(\zeta)$,
  - Unbiased Estimator: $U^J(\zeta), J \sim \text{Uniform}(\{1,\ldots,n\})$.
  - Example:

Bayesian Posterior with i.i.d observations

$$U^j(\zeta) = -\log \underbrace{p_0(\zeta)}_{\text{Prior density}} - n \log \underbrace{f(X_j \mid \zeta)}_{\text{Likelihood of } j\text{-th observation}}, \quad j = 1, \ldots, n.$$

Zig-zag process: nice properties ☺

- the Zig-Zag process is a non-reversible stochastic process
  - $\Rightarrow$ non-diffusive (kinetic-like) dynamics
  - $\Rightarrow$ better mixing
- can be modified so as to allow (data) sub-sampling without the introduction of any systematic bias:
  - Potential function: $U(\zeta) = \frac{1}{n}\sum_{j=1}^{n} U^j(\zeta)$,
  - Unbiased Estimator: $U^J(\zeta), J \sim \text{Uniform}(\{1, \ldots, n\})$.
  - Example:

Bayesian Posterior with i.i.d observations

$$U^j(\zeta) = -\log \underbrace{p_0(\zeta)}_{\text{Prior density}} \quad -n \log \underbrace{f(X_j \mid \zeta)}_{\text{Likelihood of } j\text{-th observation}}, \quad j = 1, \ldots, n.$$

Zig-zag process: not so nice properties... ☹

1. Standard implementation via Poisson-thinning requires upper bounds $M_i(t), i = 1, \ldots, d$ satisfying

$$\left\{\theta_i \partial_{\zeta_i} U(\zeta + \theta t)\right\}^+ \leq M_i(t), \quad \forall\, t \geq 0$$

   and all $\zeta, \theta \in \mathbb{R}^d \times \{-1, 1\}^d$.

2. Standard implementation employing sub-sampling requires upper bounds $M_i(t), i = 1, \ldots, d$ satisfying

$$\max_{j \in \{1, \ldots, n\}} \left\{\theta_i \partial_{\zeta_i} U^j(\zeta + \theta t)\right\}^+ \leq M_i(t), \quad \forall\, t \geq 0$$

   and all $\zeta, \theta \in \mathbb{R}^d \times \{-1, 1\}^d$.

Zig-zag process: not so nice properties... ☹

1. Standard implementation via Poisson-thinning requires upper bounds $M_i(t), i = 1, \ldots, d$ satisfying

$$\left\{\theta_i \partial_{\zeta_i} U(\zeta + \theta t)\right\}^+ \leq M_i(t), \quad \forall\, t \geq 0$$

and all $\zeta, \theta \in \mathbb{R}^d \times \{-1, 1\}^d$.

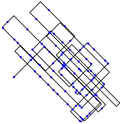
2. Standard implementation employing sub-sampling requires upper bounds $M_i(t), i = 1, \ldots, d$ satisfying

$$\max_{j \in \{1, \ldots, n\}} \left\{\theta_i \partial_{\zeta_i} U^j(\zeta + \theta t)\right\}^+ \leq M_i(t), \quad \forall\, t \geq 0$$

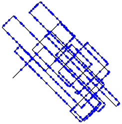
and all $\zeta, \theta \in \mathbb{R}^d \times \{-1, 1\}^d$.

Zig-zag process: not so nice porperties... ☹

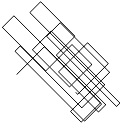
1. Bounds are problem-specific and often difficult to construct.
2. If bounds are not tight, computational efficiency decreases dramatically:



v.s.



3. Sub-sampling may result in an increased refreshment rate.
   ⇒ diffusive/quasi-reversible sampling dynamics:



v.s.

We address

- points 2 and 3 in the specific context of sub-sampling with **sparse data** in
  - [1] *Efficient posterior sampling for high-dimensional imbalanced logistic regression*, Biometrika, 2020.
- points 1 and 2 in the specific context of **Bayesian hierarchical models** in
  - [2] *Posterior computation with the Gibbs zig-zag sampler*, Bayesian Analysis, 2022.

Bayesian hierarchical models

Bayesian posterior with hierarchical prior

$$\underbrace{X_1, \ldots, X_n}_{\text{Observations}} \overset{\text{iid}}{\sim} \underbrace{f(x \mid \xi)}_{\text{Likelihood}}, \quad \xi \mid \alpha \sim \underbrace{p_0(\xi \mid \alpha)}_{\text{Prior on of } \xi \text{ given } \alpha}, \quad \alpha \sim \underbrace{p_{\mathrm{h}}(\alpha)}_{\text{hyper-prior}} \;,$$

- $\xi \in \mathbb{R}^p$ model parameters
- $\alpha \in \mathbb{R}^r$ hyper parameters
- Examples: Horseshoe prior, Spike-and-slab prior

Inference requires sampling of the joint posterior distribution:

$$\pi(\mathrm{d}\xi \, \mathrm{d}\alpha) \propto \exp\Big\{ -U^0(\xi, \alpha) - \sum_{j=1}^{n} U^j(\xi) \Big\} \mathrm{d}\xi \mathrm{d}\alpha.$$

where $U^0(\xi, \alpha) = -\log p_0(\xi \mid \alpha) - \log p_{\mathrm{h}}(\alpha)$ and
$U^j(\xi) = -\log f(X_j \mid \xi)$.

Outline

1. Background
   - Continuous-time Monte Carlo
   - The Zig-zag sampler (ZZ)
   - Bayesian hierarchical models
2. The Gibbs Zig-zag sampler (GZZ)
   - Construction
   - Theoretical properties
3. Application to posterior sampling problems
   - Random effect model
   - Logistic regression with Spike-and-Slab Prior

The Gibbs Zig-Zag sampler: idea

Potential function: $U(\xi, \alpha) = U^0(\xi, \alpha) + \sum_{j=1}^{n} U^j(\xi)$.

Combine

- updates of the component α via a Markov kernel $\mathcal{Q}\{(\alpha, \xi), \mathrm{d}\alpha'\}$ which preserves

$$\pi(\mathrm{d}\alpha \mid \xi) \propto \exp\{-U^0(\xi, \alpha)\}\, \mathrm{d}\alpha.$$

Updates don't depend on likelihood/data [Cheap]

with

- updates of the component ξ via a ZZ process which preserves

$$\pi(\mathrm{d}\xi \mid \alpha) \propto \exp\{-U(\xi, \alpha)\}\, \mathrm{d}\xi,$$

Requires bounds for $\left\{\theta_i \partial_{\xi_i} U(\xi + \theta t, \alpha)\right\}^+$ [Easier]

so that the resulting process is a PDMP preserving $\pi(\mathrm{d}\xi\, \mathrm{d}\alpha)$

The Gibbs Zig-Zag sampler: idea

Potential function: $U(\xi, \alpha) = U^0(\xi, \alpha) + \sum_{j=1}^n U^j(\xi)$.

Combine

- updates of the component α via a Markov kernel $\mathcal{Q}\{(\alpha, \xi), \mathrm{d}\alpha'\}$ which preserves

$$\pi(\mathrm{d}\alpha \mid \xi) \propto \exp\{-U^0(\xi, \alpha)\}\, \mathrm{d}\alpha.$$

Updates don't depend on likelihood/data [Cheap]

with

- updates of the component ξ via a ZZ process which preserves

$$\pi(\mathrm{d}\xi \mid \alpha) \propto \exp\{-U(\xi, \alpha)\}\, \mathrm{d}\xi,$$

Requires bounds for $\{\theta_i \partial_{\xi_i} U(\xi + \theta t, \alpha)\}^+$ [Easier]

so that the resulting process is a PDMP preserving $\pi(\mathrm{d}\xi\, \mathrm{d}\alpha)$

The Gibbs Zig-Zag sampler: idea

Potential function: $U(\xi, \alpha) = U^0(\xi, \alpha) + \sum_{j=1}^n U^j(\xi)$.

Combine

- updates of the component α via a Markov kernel $\mathcal{Q}\{(\alpha, \xi), \mathrm{d}\alpha'\}$ which preserves

$$\pi(\mathrm{d}\alpha \mid \xi) \propto \exp\{-U^0(\xi, \alpha)\} \, \mathrm{d}\alpha.$$

Updates don't depend on likelihood/data [Cheap]

with

- updates of the component ξ via a ZZ process which preserves

$$\pi(\mathrm{d}\xi \mid \alpha) \propto \exp\{-U(\xi, \alpha)\} \, \mathrm{d}\xi,$$

Requires bounds for $\left\{\theta_i \partial_{\xi_i} U(\xi + \theta t, \alpha)\right\}^+$ [Easier]

so that the resulting process is a PDMP preserving $\pi(\mathrm{d}\xi \, \mathrm{d}\alpha)$

The Gibbs Zig-Zag sampler: idea

Potential function: $U(\xi, \alpha) = U^0(\xi, \alpha) + \sum_{j=1}^n U^j(\xi)$.

Combine

- updates of the component α via a Markov kernel $\mathcal{Q}\{(\alpha, \xi), \mathrm{d}\alpha'\}$ which preserves

$$\pi(\mathrm{d}\alpha \mid \xi) \propto \exp\{-U^0(\xi, \alpha)\} \, \mathrm{d}\alpha.$$

Updates don't depend on likelihood/data [Cheap]

with

- updates of the component ξ via a ZZ process which preserves

$$\pi(\mathrm{d}\xi \mid \alpha) \propto \exp\{-U(\xi, \alpha)\} \, \mathrm{d}\xi,$$

Requires bounds for $\left\{\theta_i \partial_{\xi_i} U(\xi + \theta t, \alpha)\right\}^+$ [Easier]

so that the resulting process is a PDMP preserving $\pi(\mathrm{d}\xi \, \mathrm{d}\alpha)$

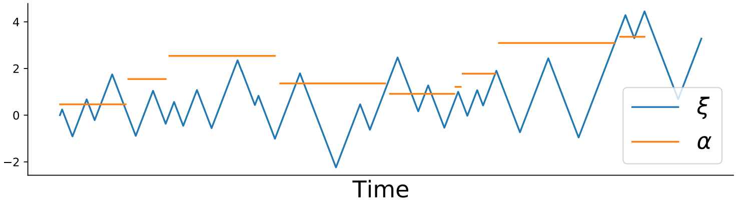
GZZ: construction

GZZ process:

$$\Big(\underbrace{\boldsymbol{\alpha}(t)}_{\text{MC-part}}, \underbrace{\boldsymbol{\xi}(t), \boldsymbol{\theta}(t)}_{\text{ZZ-part}} \Big)_{t\geq 0} \subset \mathbb{R}^r \times \mathbb{R}^p \times \{-1, 1\}^p.$$

- $(\boldsymbol{\alpha}(t))_{t\geq 0}$ is resampled according to $\mathcal{Q}\{(\boldsymbol{\alpha}(t_k), \boldsymbol{\xi}(t_k), \ \cdot \ ), \ k \in \mathbb{N}$ at random times $(t_k)_{k\in\mathbb{N}}$ given by a Poisson arrival process with constant rate η. $(\boldsymbol{\alpha}(t))_{t\geq 0}$ is constant between these arrival times.
- $(\boldsymbol{\xi}(t), \boldsymbol{\theta}(t))_{t\geq 0}$ is evolved as a ZZ-process with rate functions

$$m_i(t) = [\theta_i \partial_{\xi_i} U\{\boldsymbol{\xi}(t), \boldsymbol{\alpha}(t)\}]^+ + \underbrace{\gamma_i\{\boldsymbol{\xi}(t), \boldsymbol{\alpha}(t)\}}_{\geq 0} \quad (i = 1, \ldots, p; \ t \geq 0);$$



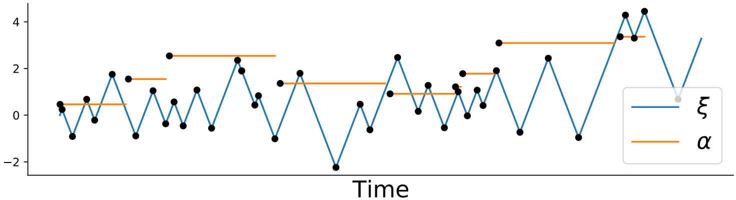
GZZ: construction

GZZ process:

$$\Big(\underbrace{\boldsymbol{\alpha}(t)}_{\text{MC-part}},\underbrace{\boldsymbol{\xi}(t),\boldsymbol{\theta}(t)}_{\text{ZZ-part}}\Big)_{t\geq 0}\subset\mathbb{R}^r\times\mathbb{R}^p\times\{-1,1\}^p.$$

- $(\boldsymbol{\alpha}(t))_{t\geq 0}$ is resampled according to $\mathcal{Q}\{(\boldsymbol{\alpha}(t_k),\boldsymbol{\xi}(t_k),\ \cdot\ ),\ k\in\mathbb{N}$ at random times $(t_k)_{k\in\mathbb{N}}$ given by a Poisson arrival process with constant rate η. $(\boldsymbol{\alpha}(t))_{t\geq 0}$ is constant between these arrival times.
- $(\boldsymbol{\xi}(t),\boldsymbol{\theta}(t))_{t\geq 0}$ is evolved as a ZZ-process with rate functions

$$m_i(t)=[\theta_i\partial_{\xi_i}U\{\boldsymbol{\xi}(t),\boldsymbol{\alpha}(t)\}]^{+}+\underbrace{\gamma_i\{\boldsymbol{\xi}(t),\boldsymbol{\alpha}(t)\}}_{\geq 0}\quad(i=1,\ldots,p;\ t\geq 0);$$



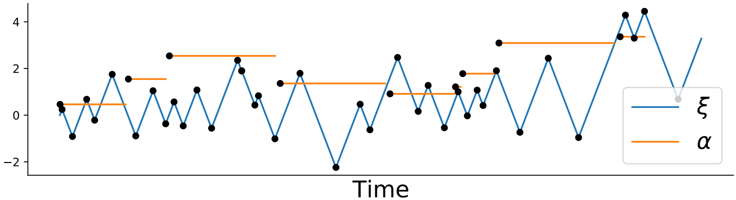
GZZ: construction

GZZ process:

$$\Big(\underbrace{\boldsymbol{\alpha}(t)}_{\text{MC-part}},\underbrace{\boldsymbol{\xi}(t),\boldsymbol{\theta}(t)}_{\text{ZZ-part}}\Big)_{t\geq 0}\subset\mathbb{R}^r\times\mathbb{R}^p\times\{-1,1\}^p.$$

- $(\boldsymbol{\alpha}(t))_{t\geq 0}$ is resampled according to $\mathcal{Q}\{(\boldsymbol{\alpha}(t_k),\boldsymbol{\xi}(t_k),\ \cdot\ ),\ k\in\mathbb{N}$ at random times $(t_k)_{k\in\mathbb{N}}$ given by a Poisson arrival process with constant rate η. $(\boldsymbol{\alpha}(t))_{t\geq 0}$ is constant between these arrival times.
- $(\boldsymbol{\xi}(t),\boldsymbol{\theta}(t))_{t\geq 0}$ is evolved as a ZZ-process with rate functions

$$m_i(t)=[\theta_i\partial_{\xi_i}U\{\boldsymbol{\xi}(t),\boldsymbol{\alpha}(t)\}]^+ + \underbrace{\gamma_i\{\boldsymbol{\xi}(t),\boldsymbol{\alpha}(t)\}}_{\geq 0}\quad (i=1,\ldots,p;\ t\geq 0);$$



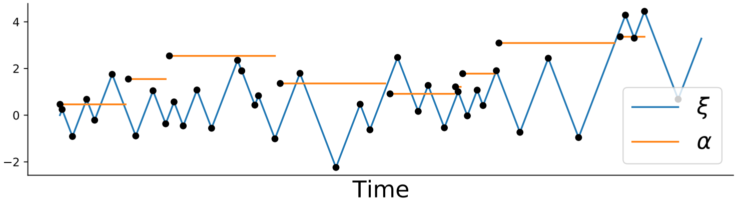
GZZ: construction

GZZ process:

$$\big(\underbrace{\boldsymbol{\alpha}(t)}_{\text{MC-part}},\underbrace{\boldsymbol{\xi}(t),\boldsymbol{\theta}(t)}_{\text{ZZ-part}}\big)_{t\geq 0}\subset\mathbb{R}^r\times\mathbb{R}^p\times\{-1,1\}^p.$$

- $(\boldsymbol{\alpha}(t))_{t\geq 0}$ is resampled according to $\mathcal{Q}\{(\boldsymbol{\alpha}(t_k),\boldsymbol{\xi}(t_k),\ \cdot\ ),\ k\in\mathbb{N}$ at random times $(t_k)_{k\in\mathbb{N}}$ given by a Poisson arrival process with constant rate η. $(\boldsymbol{\alpha}(t))_{t\geq 0}$ is constant between these arrival times.
- $(\boldsymbol{\xi}(t),\boldsymbol{\theta}(t))_{t\geq 0}$ is evolved as a ZZ-process with rate functions

$$m_i(t)=[\theta_i\partial_{\xi_i}U\{\boldsymbol{\xi}(t),\boldsymbol{\alpha}(t)\}]^+ + \underbrace{\gamma_i\{\boldsymbol{\xi}(t),\boldsymbol{\alpha}(t)\}}_{\geq 0}\quad (i=1,\ldots,p;\ t\geq 0);$$



Algorithm

Input $T^{(0)}, \boldsymbol{\xi}^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\theta}^{(0)}$.

For $k = 1, 2, 3, \ldots$

1. Compute event time:
   - Draw (independently)
     - $\tau_0 \sim \text{Exponential}(\eta)$,
     - $\tau_1, \ldots, \tau_d$ such that

       $$\mathbb{P}(\tau_i \geq t) = \exp\left\{-\int_{T^k}^{T^k+t} m_i(s)\,\mathrm{d}s\right\}.$$

   - $i_0 = \text{argmin}_i\{\tau_i\}$.
2. Evolve Zig-Zag: $\boldsymbol{\xi}^{k+1} \leftarrow \boldsymbol{\xi}^k + \tau_{i_0}\,\boldsymbol{\theta}^k$,
   $\boldsymbol{\theta}^{k+1} \leftarrow \boldsymbol{\theta}^k$, $T^{k+1} \leftarrow T^k + \tau_{i_0}$.
3. **If** $i_0 = 0$ **then:**
   $\boldsymbol{\alpha}^{k+1} \sim \mathcal{Q}\{(\boldsymbol{\xi}^{k+1}, \boldsymbol{\alpha}^k), \ \cdot\ \}$.

   Else:
   $\boldsymbol{\theta}^{k+1}_{i_0} \leftarrow -\boldsymbol{\theta}^k_{i_0}$.
   $\boldsymbol{\alpha}^{k+1} \leftarrow \boldsymbol{\alpha}^k$.

Output $\left(T^k, \boldsymbol{\xi}^k, \boldsymbol{\alpha}^k, \boldsymbol{\theta}^k\right)_{k=0,1,2,\ldots}$.

A GZZ process is obtained from the skeleton points $\{(\boldsymbol{\xi}^k, \boldsymbol{\theta}^k, \boldsymbol{\alpha}^k, T^k)\}_{k\in\mathbb{N}}$ as

$$\boldsymbol{\xi}(t) = \boldsymbol{\xi}^k + \boldsymbol{\theta}^k(t - T^k), \quad \boldsymbol{\alpha}(t) = \boldsymbol{\alpha}^k, \quad \boldsymbol{\theta}(t) = \boldsymbol{\theta}^k, \quad \text{for } T^k \leq t < T^{k+1}.$$

GZZ: Ergodic properties

Proposition

The GZZ process has

$$\widetilde{\pi}(\mathrm{d}\xi\,\mathrm{d}\alpha,\theta) = 2^{-p}\pi(\mathrm{d}\xi\,\mathrm{d}\alpha),$$

as an invariant measure.

Easy to show because

$$\underbrace{\mathcal{L}_{\mathrm{GZZ}}}_{\text{Generator of GZZ}} = \underbrace{\mathcal{L}_{\mathrm{ZZ}}}_{\text{Generator of ZZ-part}} + \eta\,\underbrace{\mathcal{L}_{\mathrm{Gibbs}}}_{\text{Generator of MC-part}}.$$

And thus

$$\int \mathcal{L}_{\mathrm{GZZ}}\varphi\,\mathrm{d}\widetilde{\pi} = \int \mathcal{L}_{\mathrm{ZZ}}\varphi\,\mathrm{d}\widetilde{\pi} + \eta\int \mathcal{L}_{\mathrm{Gibbs}}\varphi\,\mathrm{d}\widetilde{\pi} = 0 + 0.$$

for any test function φ.

Assumption 1: (on $\mathcal{Q}$ and γ_i $(i = 1, \ldots, p)$)

(A) The Markov transition kernel $\mathcal{Q}$ possesses a smooth density, and for any $(\xi, \alpha) \in \Omega_\xi \times \Omega_\alpha$, its associated probability measure has full support on Ω_α, i.e.,

$$\mathcal{Q}\{(\xi, \alpha), A\} = \int_A q\{(\xi, \alpha), \alpha'\} \, \mathrm{d}\alpha',$$

with $q \in \mathcal{C}^\infty[(\Omega_\xi \times \Omega_\alpha) \times \Omega_\alpha, (0, \infty)]$ and $q\{(\xi, \alpha), \cdot\} > 0$ for all $(\xi, \alpha) \in \Omega_\xi \times \Omega_\alpha$ and all measurable sets $A \subset \Omega_\alpha$.

(B) The refreshment rates are bounded away from zero, i.e., there exists $\underline{\gamma} > 0$ such that

$$\gamma_i(\xi, \alpha) \geq \underline{\gamma} \quad \forall\, i = 1, \ldots, p, \ \forall\, (\xi, \alpha) \in \mathbb{R}^p \times \mathbb{R}^r.$$

Theorem (Uniqueness of invariant measure)

If Assumption 1 is satisfied, then the GZZ process is ergodic with unique invariant measure $\widetilde{\pi}$. In particular, the process is path-wise ergodic in the sense that

$$\lim_{t \to \infty} \widehat{\varphi}_t = \mathbb{E}_{(\xi, \alpha, \theta) \sim \widetilde{\pi}}\{\varphi(\xi, \alpha, \theta)\} \quad \text{almost surely}$$

for any real-valued $\widetilde{\pi}$-integrable test function φ.

Proof follows in large parts: Bierkens, Roberts, Zitt, (2019).

Assumption 1: (on $\mathcal{Q}$ and γ_i $(i = 1, \ldots, p)$)

(A) The Markov transition kernel $\mathcal{Q}$ possesses a smooth density, and for any $(\xi, \alpha) \in \Omega_\xi \times \Omega_\alpha$, its associated probability measure has full support on Ω_α, i.e.,

$$\mathcal{Q}\{(\xi, \alpha), A\} = \int_A q\{(\xi, \alpha), \alpha'\} \, \mathrm{d}\alpha',$$

with $q \in \mathcal{C}^\infty[(\Omega_\xi \times \Omega_\alpha) \times \Omega_\alpha, (0, \infty)]$ and $q\{(\xi, \alpha), \cdot\} > 0$ for all $(\xi, \alpha) \in \Omega_\xi \times \Omega_\alpha$ and all measurable sets $A \subset \Omega_\alpha$.

(B) The refreshment rates are bounded away from zero, i.e., there exists $\underline{\gamma} > 0$ such that

$$\gamma_i(\xi, \alpha) \geq \underline{\gamma} \quad \forall\, i = 1, \ldots, p, \ \forall\, (\xi, \alpha) \in \mathbb{R}^p \times \mathbb{R}^r.$$

Theorem (Uniqueness of invariant measure)

If Assumption 1 is satisfied, then the GZZ process is ergodic with unique invariant measure $\widetilde{\pi}$. In particular, the process is path-wise ergodic in the sense that

$$\lim_{t \to \infty} \widehat{\varphi}_t = \mathbb{E}_{(\xi, \alpha, \theta) \sim \widetilde{\pi}}\{\varphi(\xi, \alpha, \theta)\} \ \text{ almost surely}$$

for any real-valued $\widetilde{\pi}$-integrable test function φ.

Proof follows in large parts: Bierkens, Roberts, Zitt, (2019).

Assumption 1: (on $\mathcal{Q}$ and γ_i $(i = 1, \dots, p)$)

(A) The Markov transition kernel $\mathcal{Q}$ possesses a smooth density, and for any $(\xi, \alpha) \in \Omega_\xi \times \Omega_\alpha$, its associated probability measure has full support on Ω_α, i.e.,

$$\mathcal{Q}\{(\xi,\alpha), A\} = \int_A q\{(\xi,\alpha),\alpha'\}\,\mathrm{d}\alpha',$$

with $q \in \mathcal{C}^\infty[(\Omega_\xi \times \Omega_\alpha) \times \Omega_\alpha, (0,\infty)]$ and $q\{(\xi,\alpha),\cdot\} > 0$ for all $(\xi,\alpha) \in \Omega_\xi \times \Omega_\alpha$ and all measurable sets $A \subset \Omega_\alpha$.

(B) The refreshment rates are bounded away from zero, i.e., there exists $\underline{\gamma} > 0$ such that

$$\gamma_i(\xi,\alpha) \geq \underline{\gamma} \quad \forall\, i = 1,\dots,p,\ \forall\,(\xi,\alpha) \in \mathbb{R}^p \times \mathbb{R}^r.$$

Theorem (Uniqueness of invariant measure)

If Assumption 1 is satisfied, then the GZZ process is ergodic with unique invariant measure $\widetilde{\pi}$. In particular, the process is path-wise ergodic in the sense that

$$\lim_{t\to\infty} \widehat{\varphi}_t = \mathbb{E}_{(\xi,\alpha,\theta)\sim\widetilde{\pi}}\{\varphi(\xi,\alpha,\theta)\} \ \text{ almost surely}$$

for any real-valued $\widetilde{\pi}$-integrable test function φ.

Proof follows in large parts: Bierkens, Roberts, Zitt, (2019).

Central limit theorem

Let

(i) $\mathcal{Q}\{(\xi,\alpha),\cdot\} = \pi(\cdot\,|\,\xi,\alpha)$,

(ii) $0 < \inf_{\xi\in\mathbb{R}^p,\alpha\in\mathbb{R}^r} \gamma_i(\xi,\alpha) \leq \sup_{\xi\in\mathbb{R}^p,\alpha\in\mathbb{R}^r} \gamma_i(\xi,\alpha) < \infty$,

(iii) certain growth conditions on U (see Assumption 2 in [Sachs et al., 2022]),

hold.

Then, there is $\sigma_\varphi^2 > 0$ so that

$$\sqrt{t}\left[\frac{1}{t}\int_0^t \varphi(\boldsymbol{\xi}(s),\boldsymbol{\alpha}(s))\mathrm{d}s - \mathbb{E}_{(\xi,\alpha)\sim\pi}\{\varphi(\xi,\alpha)\}\right] \xrightarrow[t\to\infty]{\text{law}} \mathcal{N}(0,\sigma_\varphi^2).$$

Outline

1. Background
   - Continuous-time Monte Carlo
   - The Zig-zag sampler (ZZ)
   - Bayesian hierarchical models
2. The Gibbs Zig-zag sampler (GZZ)
   - Construction
   - Theoretical properties
3. Application to posterior sampling problems
   - Random effect model
   - Logistic regression with Spike-and-Slab Prior

Random effect model

For $i = 1, \ldots, \underbrace{n}_{\text{Subject index}}$, and $j = 1, \ldots, \underbrace{K}_{\text{group index}}$ let

$$\underbrace{Y_{ij}}_{\text{Response}} \sim \text{Bernoulli}\left(\frac{1}{1+e^{-\psi_{ij}}}\right), \quad \psi_{ij} = \underbrace{m + v_j}_{\text{Random effects}} + \underbrace{\overbrace{X_{ij}^T}^{\text{Predictor} \in \mathbb{R}^p} \beta}_{\text{Fixed effect}},$$

with hierarchical prior specified by

$$m \sim \text{Normal}(0, \phi^{-1}), \quad v_j \overset{\text{iid}}{\sim} \text{Normal}(0, \phi^{-1}) \ \ (j = 1, \ldots, K),$$

$$\beta_l \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2) \ \ (l = 1, \ldots, p), \quad \phi \sim \text{Ga}(a_\phi, b_\phi), \quad \sigma^2 \sim \text{IG}(a_\sigma, b_\sigma),$$

Variable decomposition in GZZ: $\xi = (m, v_1 \ldots, v_K, \beta_1, \ldots, \beta_p)$, $\alpha = (\phi, \sigma^2)$.

Results

Random effect model

For $i = 1, \ldots, \underbrace{n}_{\text{Subject index}}$, and $j = 1, \ldots, \underbrace{K}_{\text{group index}}$ let

$$\underbrace{Y_{ij}}_{\text{Response}} \sim \text{Bernoulli}\left(\frac{1}{1+e^{-\psi_{ij}}}\right), \quad \psi_{ij} = \underbrace{m + v_j}_{\text{Random effects}} + \underbrace{\overbrace{X_{ij}^T}^{\text{Predictor} \in \mathbb{R}^p} \beta}_{\text{Fixed effect}},$$

with hierarchical prior specified by

$$m \sim \text{Normal}(0, \phi^{-1}), \quad v_j \overset{\text{iid}}{\sim} \text{Normal}(0, \phi^{-1}) \quad (j = 1, \ldots, K),$$
$$\beta_l \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2) \quad (l = 1, \ldots, p), \quad \phi \sim \text{Ga}(a_\phi, b_\phi), \quad \sigma^2 \sim \text{IG}(a_\sigma, b_\sigma),$$

Variable decomposition in GZZ: $\xi = (m, v_1 \ldots, v_K, \beta_1, \ldots, \beta_p)$, $\alpha = (\phi, \sigma^2)$.

Results

Random effect model

For $i = 1, \ldots, \underbrace{n}_{\text{Subject index}}$, and $j = 1, \ldots, \underbrace{K}_{\text{group index}}$ let

$$\underbrace{Y_{ij}}_{\text{Response}} \sim \text{Bernoulli}\left(\frac{1}{1 + e^{-\psi_{ij}}}\right), \quad \psi_{ij} = \underbrace{m + v_j}_{\text{Random effects}} + \underbrace{\overbrace{X_{ij}^T}^{\text{Predictor} \in \mathbb{R}^p} \beta}_{\text{Fixed effect}},$$

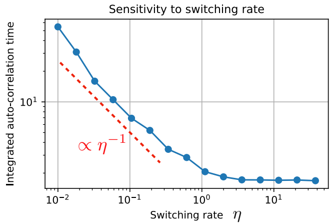
with hierarchical prior specified by

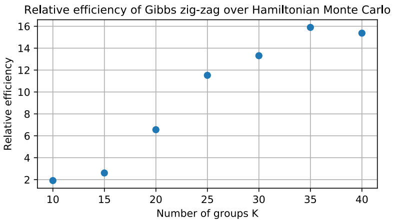
$$m \sim \text{Normal}(0, \phi^{-1}), \quad v_j \overset{\text{iid}}{\sim} \text{Normal}(0, \phi^{-1}) \quad (j = 1, \ldots, K),$$

$$\beta_l \overset{\text{iid}}{\sim} \text{Normal}(0, \sigma^2) \quad (l = 1, \ldots, p), \quad \phi \sim \text{Ga}(a_\phi, b_\phi), \quad \sigma^2 \sim \text{IG}(a_\sigma, b_\sigma),$$

Variable decomposition in GZZ: $\xi = (m, v_1 \ldots, v_K, \beta_1, \ldots, \beta_p)$, $\alpha = (\phi, \sigma^2)$.

Results





Logistic regression with Spike-and-Slab Prior

For $j = 1, \ldots, n$,

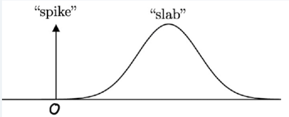
$$\underbrace{Y_j}_{\text{Response}} \sim \text{Bernoulli}\left(\frac{1}{1+e^{-X_j\beta}}\right), \quad \underbrace{X_j}_{\text{Predictors}} \in \mathbb{R}^p.$$

with hierarchical prior specified by

$$\beta_i \overset{\text{ind}}{\sim} \gamma_i\,\delta(\cdot) + (1-\gamma_i)\,\text{Normal}(0, \nu\tau_i^2),$$

$$\gamma_i \overset{\text{iid}}{\sim} \text{Bernoulli}(\pi), \quad \tau_i \sim C^+(0,1) \quad (i = 1, \ldots, p),$$

$$\nu \sim \text{IG}(a_\nu, b_\nu), \quad \pi \sim \text{Beta}(a_\pi, b_\pi).$$



Variable decomposition in GZZ: $\xi = \beta$, $\alpha = (\gamma_1, \ldots, \gamma_p, \tau_1, \ldots, \tau_p, \nu, \pi)$.

Results

Logistic regression with Spike-and-Slab Prior

For $j = 1, \ldots, n$,

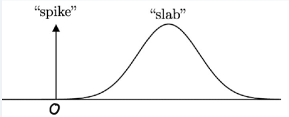
$$\underbrace{Y_j}_{\text{Response}} \sim \text{Bernoulli}\left(\frac{1}{1+e^{-X_j\beta}}\right), \qquad \underbrace{X_j}_{\text{Predictors}} \in \mathbb{R}^p.$$

with hierarchical prior specified by

$$\beta_i \overset{\text{ind}}{\sim} \gamma_i\,\delta(\cdot) + (1-\gamma_i)\,\text{Normal}(0, \nu\tau_i^2),$$

$$\gamma_i \overset{\text{iid}}{\sim} \text{Bernoulli}(\pi), \quad \tau_i \sim C^+(0,1) \quad (i = 1, \ldots, p),$$

$$\nu \sim \text{IG}(a_\nu, b_\nu), \quad \pi \sim \text{Beta}(a_\pi, b_\pi).$$



Variable decomposition in GZZ: $\xi = \beta$, $\alpha = (\gamma_1, \ldots, \gamma_p, \tau_1, \ldots, \tau_p, \nu, \pi)$.

Results

Logistic regression with Spike-and-Slab Prior

For $j = 1, \ldots, n$,

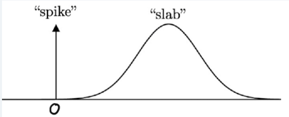
$$\underbrace{Y_j}_{\text{Response}} \sim \text{Bernoulli}\left(\frac{1}{1 + e^{-X_j \beta}}\right), \quad \underbrace{X_j}_{\text{Predictors}} \in \mathbb{R}^p.$$

with hierarchical prior specified by

$$\beta_i \overset{\text{ind}}{\sim} \gamma_i \, \delta(\cdot) + (1 - \gamma_i) \, \text{Normal}(0, \nu \tau_i^2),$$

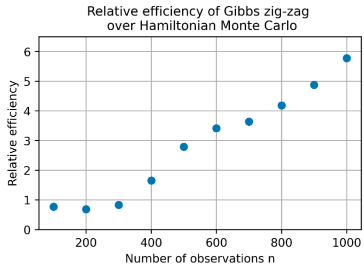
$$\gamma_i \overset{\text{iid}}{\sim} \text{Bernoulli}(\pi), \quad \tau_i \sim C^+(0, 1) \quad (i = 1, \ldots, p),$$

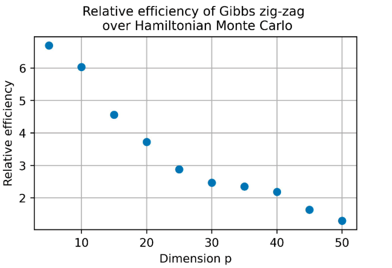
$$\nu \sim \text{IG}(a_\nu, b_\nu), \quad \pi \sim \text{Beta}(a_\pi, b_\pi).$$



Variable decomposition in GZZ: $\xi = \beta$, $\alpha = (\gamma_1, \ldots, \gamma_p, \tau_1, \ldots, \tau_p, \nu, \pi)$.

Results





Summary

1. With the GZZ-sampler we propose a new type of PDMP which
   - combines elements of traditional MCMC with PDMP-sampling in a Gibbs-like construction
   - simplifies construction of (tight) upper bounds: “if you can’t find a bound, just use a MH-update instead”
   - allows to take advantage of both worlds: versatility of MCMC + error-free subsampling with PDMP
2. We show (under rather restrictive conditions) that the GZZ sampler satisfies certain theoretical properties: unique ergodicity + central limit theorem
3. We demonstrate in numerical experiemnts efficiency gains over highly tuned HMC sampling.

Summary

1. With the GZZ-sampler we propose a new type of PDMP which
   - combines elements of traditional MCMC with PDMP-sampling in a Gibbs-like construction
   - simplifies construction of (tight) upper bounds: "if you can't find a bound, just use a MH-update instead"
   - allows to take advantage of both worlds: versatility of MCMC + error-free subsampling with PDMP
2. We show (under rather restrictive conditions) that the GZZ sampler satisfies certain theoretical properties: unique ergodicity + central limit theorem
3. We demonstrate in numerical experiemnts efficiency gains over highly tuned HMC sampling.

Summary

1. With the GZZ-sampler we propose a new type of PDMP which
   - combines elements of traditional MCMC with PDMP-sampling in a Gibbs-like construction
   - simplifies construction of (tight) upper bounds: "if you can't find a bound, just use a MH-update instead"
   - allows to take advantage of both worlds: versatility of MCMC + error-free subsampling with PDMP
2. We show (under rather restrictive conditions) that the GZZ sampler satisfies certain theoretical properties: unique ergodicity + central limit theorem
3. We demonstrate in numerical experiemnts efficiency gains over highly tuned HMC sampling.

References

[1] D. Sen, M. Sachs, J. Lu, and D. B. Dunson, *Efficient posterior sampling for high-dimensional imbalanced logistic regression*, Biometrika, 2020.

[2] J. Bierkens, P. Fearnhead, and G. Roberts, *The zig-zag process and super-efficient sampling for Bayesian analysis of big data 1*, Ann. Stat., 2019.

[3] J. Bierkens, G. Roberts, P. Zitt, *Ergodicity of the zigzag process*, Ann. Appl. Probab., 2019

[4] M. Sachs, D. Sen, J. Lu, and D. B. Dunson, *Posterior computation with the Gibbs zig-zag sampler*, Bayesian Analysis, 2022.

Supplementary slides

Algorithm: zig-zag with Poisson thinning

Input $T^{(0)}, \boldsymbol{\zeta}^{(0)}, \boldsymbol{\theta}^{(0)}$.
For $k = 1, 2, 3, \ldots$

1. Compute event time:
   - Draw $\tau_1, \ldots, \tau_d$ such that
   $$\mathbb{P}(\tau_i \geq t) = \exp\left\{ - \int_0^t M_i(s)\, \mathrm{d}s \right\}.$$
   - $i_0 = \mathrm{argmin}\{\tau_i\,;\, i = 1, \ldots, d\}$.
2. Evolve position:
   $(T^{k+1}, \boldsymbol{\zeta}^{k+1}) = (T^k + \tau_{i_0}, \boldsymbol{\zeta}^k + \boldsymbol{\theta}^k \tau_{i_0})$.
3. Bounce: with probability $p = \frac{m_i(\tau_{i_0})}{M_i(\tau_{i_0})}$:
   $${\boldsymbol{\theta}_{i_0}}^{k+1} \leftarrow -{\boldsymbol{\theta}_{i_0}}^{k},$$
   otherwise:
   $${\boldsymbol{\theta}_{i_0}}^{k+1} \leftarrow {\boldsymbol{\theta}_{i_0}}^{k}$$

Output $\left(T^k, \boldsymbol{\zeta}^k, \boldsymbol{\theta}^k\right)_{k=0,1,2,\ldots}$.

Supplementary slides

Assumption 2 (On potential function U and excess switching rates γ_i).

(A) *There exist continuous functions* $g_i : \Omega_\xi \to [0, \infty)$ $(i = 1, 2)$, *satisfying* $g_i(\xi) \to 0$ *as* $|\xi| \to \infty$ *and a constant* $c > 0$ *so that the inequalities*

$$\frac{\max\{1, \|\mathrm{Hess}_\xi U(\xi, \alpha)\|\}}{|\nabla_\xi U(\xi, \alpha)|} \leq g_1(\xi) \quad \textit{and} \quad \frac{|\nabla_\xi U(\xi, \alpha)|}{U(\xi, \alpha)} \leq g_2(\xi), \tag{1}$$

hold for all $\alpha \in \Omega_\alpha$ *and* $\xi \in \Omega_\xi$ *with* $|\xi| > c$. *Here* $\mathrm{Hess}_\xi U$ *and* $\nabla_\xi U$ *denote the Hessian and gradient of the function* $\xi \mapsto U(\xi, \alpha)$, *respectively, and* $|\cdot|$ *and* $\|\cdot\|$ *denote the Euclidean norm and the Frobenius norm, respectively.*

(B) *The excess switching rates* γ_i $(i = 1, \ldots, p)$ *are bounded from above, that is, there exists* $\overline{\gamma} > 0$ *so that*

$$\sup_{(\xi,\alpha)\in\Omega_\xi\times\Omega_\alpha} \gamma_i(\xi, \alpha) \leq \overline{\gamma}. \tag{2}$$

(C) *Let* $\delta > 0$ *and* $a > 0$ *be such that* $0 \leq \overline{\gamma}\delta < a < 1$ *with* $\overline{\gamma}$ *as specified in Assumption 2. Define the function*

$$V(\xi, \alpha, \theta) = \exp\left[aU(\xi, \alpha) + \sum_{i=1}^{p} \phi\left\{\theta_i \partial_{\xi_i} U(\xi, \alpha)\right\}\right]$$

where $\phi(s) = \mathrm{sign}(s)\log(1 + \delta|s|)/2$. *There exist a choice of* a *and* δ, *and a constants* $r > 0$ *and* $c > 0$ *such that the inequality*

$$\int_{\Omega_\alpha} \frac{V(\xi, \widehat{\alpha}, \theta)}{V(\xi, \alpha, \theta)} \exp\{-U(\xi, \widehat{\alpha})\}\mathrm{d}\widehat{\alpha} + r < \int_{\Omega_\alpha} \exp\{-U(\xi, \widehat{\alpha})\}\mathrm{d}\widehat{\alpha} \tag{3}$$

holds for all $(\xi, \alpha) \in \Omega_\xi \times \Omega_\alpha$ *with* $|(\xi, \alpha)| > c$, *and all* $\theta \in \{-1, 1\}^p$.

Supplementary slides

Theorem 1. *If Assumption 1 is satisfied, then the GZZ process is ergodic with unique invariant measure $\widetilde{\pi}$. In particular, the process is path-wise ergodic in the sense that*

$$\lim_{t\to\infty} \widehat{\varphi}_t = \mathbb{E}_{(\xi,\alpha,\theta)\sim\widetilde{\pi}}\{\varphi(\xi,\alpha,\theta)\} \text{ almost surely}$$

for any real-valued $\widetilde{\pi}$-integrable test function φ.

Theorem 2. *Assumption 2(C) is satisfied if Assumption 1 and Assumption 2(B), hold, and the potential function U can be decomposed as $U(\xi,\alpha) = U_1(\xi) + b(\xi,\alpha) + U_2(\alpha)$, where b is such that the absolute values of b and its derivatives are bounded, that is, there exists $\bar{b} > 0$ such that*

$$|b(\xi,\alpha)| \le \bar{b} \quad \text{and} \quad |\partial_{\xi_i} b(\xi,\alpha)| \le \bar{b}$$

for all $(\xi,\alpha) \in \Omega_\xi \times \Omega_\alpha$, and $i = 1,\dots,p$.

Adaptive Langevin Dynamics

Adaptive Langevin Dynamics

Gradient of log posterior

$$\nabla \log \pi(\theta) = \nabla \log \operatorname{prob}(\theta) + \sum_{j=1}^{N} \nabla \log \operatorname{prob}(\theta \mid x_j, y_j),$$

Adaptive Langevin Dynamics

Gradient of log posterior

Computational cost scales linearly in N

$$\nabla \log \pi(\theta) = \nabla \log \text{prob}(\theta) + \boxed{\sum_{j=1}^{N} \nabla \log \text{prob}(\theta \mid x_j, y_j),}$$

Adaptive Langevin Dynamics

Computational cost scales linearly in N

Gradient of log posterior

$$\nabla \log \pi(\theta) = \nabla \log \text{prob}(\theta) + \boxed{\sum_{j=1}^{N} \nabla \log \text{prob}(\theta \mid x_j, y_j),}$$

Unbiased estimator via sub-sampling:

$$\widehat{\nabla \log}(\theta) = \nabla \log \text{prob}(\theta) + N \nabla \log \text{prob}(\theta \mid x_J, y_J)$$

$$\text{with } J \sim \text{Uniform}(\{1, \ldots, N\}).$$

Adaptive Langevin Dynamics

Computational cost scales linearly in N

Gradient of log posterior

$$\nabla \log \pi(\theta) = \nabla \log \text{prob}(\theta) + \boxed{\sum_{j=1}^{N} \nabla \log \text{prob}(\theta \mid x_j, y_j),}$$

Unbiased estimator via sub-sampling:

For large N

$$\widehat{\nabla \log(\theta)} = \nabla \log \text{prob}(\theta) + N \nabla \log \text{prob}(\theta \mid x_J, y_J) \approx \nabla \log \pi(\theta) + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_p)$$

$$\text{with } J \sim \text{Uniform}(\{1, \ldots, N\}).$$

Adaptive Langevin Dynamics

Gradient of log posterior

Computational cost scales linearly in N

$$\nabla \log \pi(\theta) = \nabla \log \text{prob}(\theta) + \boxed{\sum_{j=1}^{N} \nabla \log \text{prob}(\theta \mid x_j, y_j),}$$

Unbiased estimator via sub-sampling:

For large N

$$\widehat{\nabla \log(\theta)} = \nabla \log \text{prob}(\theta) + N \nabla \log \text{prob}(\theta \mid x_J, y_J) \approx \nabla \log \pi(\theta) + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_p)$$

$$\text{with } J \sim \text{Uniform}(\{1, \ldots, N\}).$$

Adaptive Langevin equation

$$\begin{aligned} \mathrm{d}\boldsymbol{\theta} &= \boldsymbol{\theta}\, \mathrm{d}t, \\ \mathrm{d}\boldsymbol{v} &= \nabla \log \pi(\boldsymbol{\theta})\, \mathrm{d}t + \sqrt{h}\, \sigma\, \mathrm{d}W - \xi\, \boldsymbol{v}\, \mathrm{d}t, \\ \mathrm{d}\xi &= \frac{1}{\nu} \left(|\boldsymbol{v}|^2 - p \right) \mathrm{d}t. \end{aligned}$$

Adaptive Langevin Dynamics

Gradient of log posterior

Computational cost scales linearly in N

$$\nabla \log \pi(\theta) = \nabla \log \operatorname{prob}(\theta) + \boxed{\sum_{j=1}^{N} \nabla \log \operatorname{prob}(\theta \mid x_j, y_j),}$$

Unbiased estimator via sub-sampling:

For large N

$$\widehat{\nabla \log(\theta)} = \nabla \log \operatorname{prob}(\theta) + N \nabla \log \operatorname{prob}(\theta \mid x_J, y_J) \approx \nabla \log \pi(\theta) + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_p)$$

$$\text{with } J \sim \operatorname{Uniform}(\{1, \ldots, N\}).$$

Adaptive Langevin equation

Auxiliary Momentum $\in \mathbb{R}^p$

$$\begin{aligned} \mathrm{d}\boldsymbol{\theta} &= \boldsymbol{\theta} \, \mathrm{d}t, \\ \mathrm{d}\boxed{\boldsymbol{v}} &= \nabla \log \pi(\boldsymbol{\theta}) \, \mathrm{d}t + \sqrt{h} \, \sigma \, \mathrm{d}W - \xi \, \boldsymbol{v} \, \mathrm{d}t, \\ \mathrm{d}\xi &= \frac{1}{\nu} \left(|\boldsymbol{v}|^2 - p \right) \mathrm{d}t. \end{aligned}$$

Adaptive Langevin Dynamics

Gradient of log posterior

Computational cost scales linearly in N

$$\nabla \log \pi(\theta) = \nabla \log \text{prob}(\theta) + \boxed{\sum_{j=1}^{N} \nabla \log \text{prob}(\theta \mid x_j, y_j),}$$

Unbiased estimator via sub-sampling:

For large N

$$\widehat{\nabla \log(\theta)} = \nabla \log \text{prob}(\theta) + N \nabla \log \text{prob}(\theta \mid x_J, y_J) \approx \nabla \log \pi(\theta) + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_p)$$

$$\text{with } J \sim \text{Uniform}(\{1, \ldots, N\}).$$

Adaptive Langevin equation

Adaptive Friction

Auxiliary Momentum $\in \mathbb{R}^p$

$$\mathrm{d}\boldsymbol{\theta} = \boldsymbol{\theta}\, \mathrm{d}t,$$

$$\mathrm{d}\boxed{\boldsymbol{v}} = \nabla \log \pi(\boldsymbol{\theta})\, \mathrm{d}t + \sqrt{h}\, \sigma\, \mathrm{d}W - \boxed{\xi}\, \boldsymbol{v}\, \mathrm{d}t,$$

$$\mathrm{d}\xi = \frac{1}{\nu} \left(|\boldsymbol{v}|^2 - p \right) \mathrm{d}t.$$

Adaptive Langevin Dynamics

Gradient of log posterior

Computational cost scales linearly in N

$$\nabla \log \pi(\theta) = \nabla \log \text{prob}(\theta) + \boxed{\sum_{j=1}^{N} \nabla \log \text{prob}(\theta \mid x_j, y_j),}$$

Unbiased estimator via sub-sampling:

For large N

$$\widehat{\nabla \log(\theta)} = \nabla \log \text{prob}(\theta) + N \nabla \log \text{prob}(\theta \mid x_J, y_J) \approx \nabla \log \pi(\theta) + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_p)$$

with $J \sim \text{Uniform}(\{1, \ldots, N\})$.

Adaptive Langevin equation

Adaptive Friction

Auxiliary Momentum $\in \mathbb{R}^p$

$$\begin{aligned} \mathrm{d}\boldsymbol{\theta} &= \boldsymbol{\theta}\, \mathrm{d}t, \\ \mathrm{d}\boxed{\boldsymbol{v}} &= \nabla \log \pi(\boldsymbol{\theta})\, \mathrm{d}t + \sqrt{h}\, \sigma\, \mathrm{d}W - \boxed{\xi}\, \boldsymbol{v}\, \mathrm{d}t, \\ \mathrm{d}\xi &= \frac{1}{\boxed{\nu}} \left(|\boldsymbol{v}|^2 - p \right) \mathrm{d}t. \end{aligned}$$

Coupling parameter > 0

Adaptive Langevin Dynamics

Gradient of log posterior

Computational cost scales linearly in N

$$\nabla \log \pi(\theta) = \nabla \log \text{prob}(\theta) + \boxed{\sum_{j=1}^{N} \nabla \log \text{prob}(\theta \mid x_j, y_j),}$$

Unbiased estimator via sub-sampling:

For large N

$$\widehat{\nabla \log(\theta)} = \nabla \log \text{prob}(\theta) + N \nabla \log \text{prob}(\theta \mid x_J, y_J) \approx \nabla \log \pi(\theta) + \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_p)$$

$$\text{with } J \sim \text{Uniform}(\{1, \dots, N\}).$$

Adaptive Langevin equation

Adaptive Friction

Auxiliary Momentum $\in \mathbb{R}^p$

$$\left.\begin{aligned} \mathrm{d}\boldsymbol{\theta} &= \boldsymbol{\theta}\, \mathrm{d}t, \\ \mathrm{d}\boxed{\boldsymbol{v}} &= \nabla \log \pi(\boldsymbol{\theta})\, \mathrm{d}t + \sqrt{h}\, \sigma\, \mathrm{d}W - \boxed{\xi}\, \boldsymbol{v}\, \mathrm{d}t, \\ \mathrm{d}\xi &= \frac{1}{\boxed{\nu}} \left(|\boldsymbol{v}|^2 - p \right) \mathrm{d}t. \end{aligned}\right\}$$

Coupling parameter > 0

Ergodic with invariant measure:

$$\widetilde{\pi}(\theta, \boldsymbol{v}, \xi) \propto \pi(\theta) e^{-\frac{1}{2}|\boldsymbol{v}|^2} e^{-\frac{\nu}{2}(\xi - h\sigma^2)^2}$$

Parameter-dependent sampling efficiency

Change of variable: $\gamma := h\sigma^2/2$

Parameter-dependent sampling efficiency

Change of variable: $\gamma := h\sigma^2/2$

Theorem

Let $-\log\pi$ *satisfy a Poincare inequality. There exist* $C, \bar\lambda$ *such that, for any* $\nu, \gamma > 0$, *there is* $\lambda_{\nu,\gamma} > 0$ *for which*

$$\forall t \geq 0, \quad \forall \varphi \in L^2(\pi), \qquad \left\| \mathrm{e}^{t\mathcal{L}_{\mathrm{AdL}}}\varphi - \int \varphi \,\mathrm{d}\pi \right\|_{L^2(\pi)} \leq C\mathrm{e}^{-\lambda_{\nu,\gamma}} \left\| \varphi - \int \varphi \,\mathrm{d}\pi \right\|_{L^2(\pi)},$$

with the lower bound $\lambda_{\nu,\gamma} \geq \bar\lambda \min\left(\gamma\nu, \frac{1}{\gamma}, \frac{\nu}{\gamma}, \frac{\gamma}{\nu}\right)$.

Parameter-dependent sampling efficiency

Change of variable: $\gamma := h\sigma^2/2$

Theorem

Let $-\log \pi$ satisfy a Poincare inequality. There exist $C, \overline{\lambda}$ such that, for any $\nu, \gamma > 0$, there is $\lambda_{\nu,\gamma} > 0$ for which

$$\forall t \geq 0, \quad \forall \varphi \in L^2(\pi), \qquad \left\| \mathrm{e}^{t\mathcal{L}_{\mathrm{AdL}}}\varphi - \int \varphi \, \mathrm{d}\pi \right\|_{L^2(\pi)} \leq C\mathrm{e}^{-\lambda_{\nu,\gamma}} \left\| \varphi - \int \varphi \, \mathrm{d}\pi \right\|_{L^2(\pi)},$$

with the lower bound $\lambda_{\nu,\gamma} \geq \overline{\lambda} \min\left(\gamma\nu, \frac{1}{\gamma}, \frac{\nu}{\gamma}, \frac{\gamma}{\nu}\right)$.

Collorary (Central Limit Theorem for Adaptive Langevin Dynamics)

Consider $\varphi \in L^2(\pi)$. Then

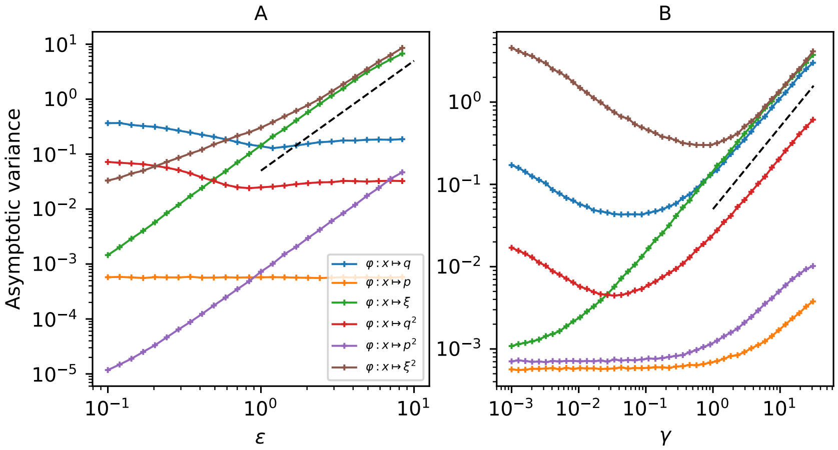
$$\sqrt{t}\left(\frac{1}{t}\int_0^t \varphi(\theta(s), \boldsymbol{v}(s), \xi(s)) \, \mathrm{d}s - \mathbb{E}_\pi \varphi\right) \xrightarrow[t\to+\infty]{\text{law}} \mathcal{N}(0, \sigma^2_{\nu,\gamma}(\varphi)),$$

where the asymptotic variance is bounded as

$$0 \leq \sigma^2_{\nu,\gamma}(\varphi) \leq \frac{2C}{\lambda_{\nu,\gamma}} \|\varphi\|^2_{L^2(\pi)}.$$

Asymptotic variance

$$U(\mathbf{q}) = \frac{b}{a}\left(\mathbf{q}^2 - a\right)^2 + c\,\mathbf{q} \tag{1}$$



Spectral gap

$$\text{Generator: } \mathcal{L}_{\mathrm{AdL}} = \gamma\mathcal{L}_{\mathrm{O}} + \varepsilon\mathcal{L}_{\mathrm{NH}},$$

$$\text{Galerkin subspace: } \psi_{k,l}(\boldsymbol{v},\xi) = h_k(\boldsymbol{v})h_l(\xi), \quad 0 \le l,k \le L-1,$$