Adaptive Langevin Algorithms for Canonical Sampling with Noisy Forces in Scale-bridging Molecular Dynamics

Ben Leimkuhler
University of Edinburgh

Problem: use stochastic dynamics to accurately sample a distribution with given positive smooth density

$$\rho \propto \exp(-U)$$

in case the force $-\nabla U$ can only be computed approximately

Examples:

Multiscale models
several flavors of hybrid **ab initio MD Methods**
QM/MM methods

…Many applications in **Bayesian Inference & Big Data Analytics**

What to do about the force error?

Methods for Gibbs sampling with a noisy gradient

- Ignore the perturbation
- Estimate the perturbation/correct for it
- SGLD (Langevin with a diminishing stepsize sequence)
- Adaptive thermostats Ad-L, Ad-NH,…

With a clean gradient: $F(x) = -\nabla U(x)$

Brownian Dynamics

$$\mathrm{d}x = F(x)\mathrm{dt} + \sqrt{2}\mathrm{dW}$$

- SDEs which can be solved to generate a path $x(t)$
- Under typical conditions, for almost all paths,

$$\lim_{\tau\to\infty} \tau^{-1} \int_0^\tau \varphi(x(t))\mathrm{d}t = \int_\Omega \varphi(x)\rho(x)\mathrm{d}x$$

How to discretize?

Euler-Maruyama?   Stochastic Heun?

Euler-Maruyama Method

$$x_0 \to x_1 \to \ldots \to x_N, \qquad Nh = \tau$$

discrete Brownian path

$$x_{n+1} = x_n + \cancel{hF(x_n)} + \sqrt{2h}R_n$$

$$R_n \sim \mathcal{N}(0,1)$$

Leimkuhler-Matthews Method

$$x_{n+1} = x_n + hF(x_n) + \sqrt{h/2}(R_n + R_{n-1})$$

[L. & Matthews, AMRX, 2013]

[L., Matthews & Tretyakov, Proc Roy Soc A, 2014]

Theorem *(BL-CM-MT Proc Roy Soc A 2014)*
For the **L-M method**, under suitable conditions,

$$\mathbf{E}\varphi(\mathbf{X}_x(\tau)) - \mathbf{E}\varphi(\mathbf{X}_N) = C_0(\tau, x)h + C(\tau, x)h^2$$

$$|C_0(\tau, x)| \leq K_0(1 + |x|^\eta)e^{-\lambda_0 \tau}$$

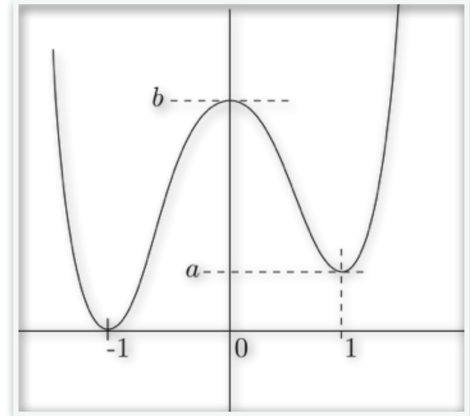
$$\lambda_0, \lambda > 0$$

$$|C(\tau, x)| \leq K(1 + |x|^\eta e^{-\lambda \tau})$$

Euler-Maruyama: Weak 1st order, also as $\tau \to \infty$

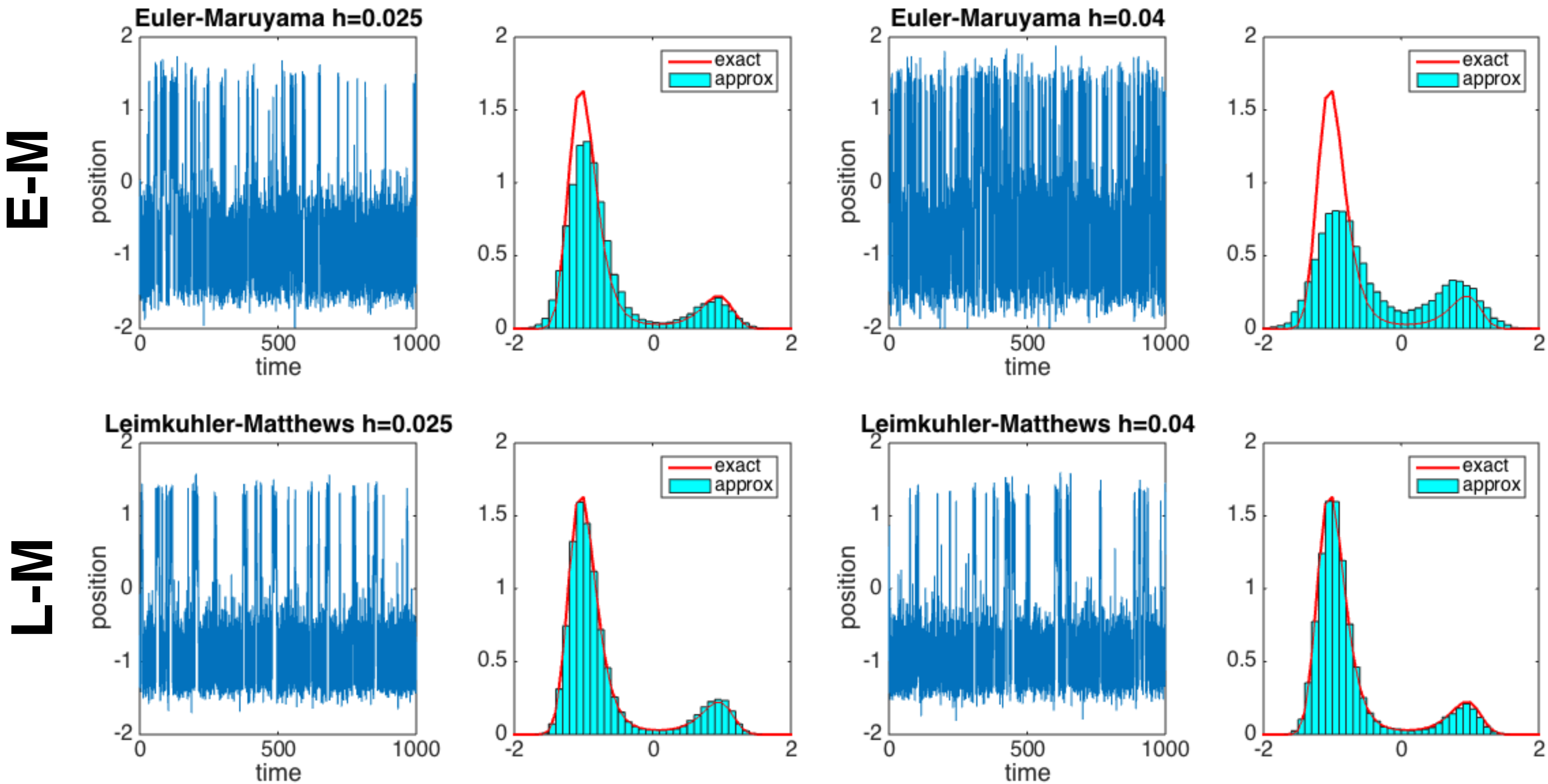
L-M: **Weak first order** -> **weak asymptotic second order**

Uneven Double Well



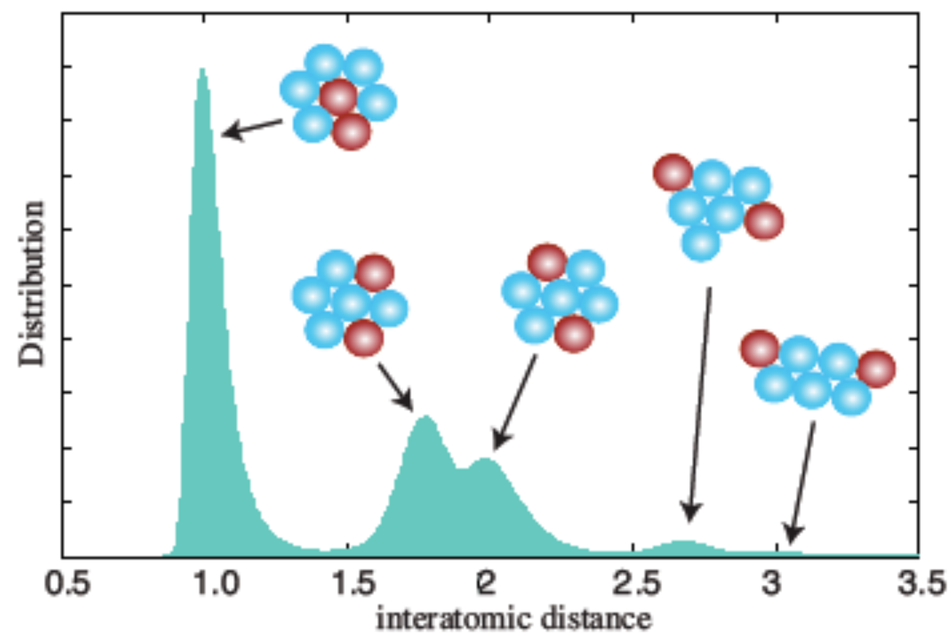
small stepsize

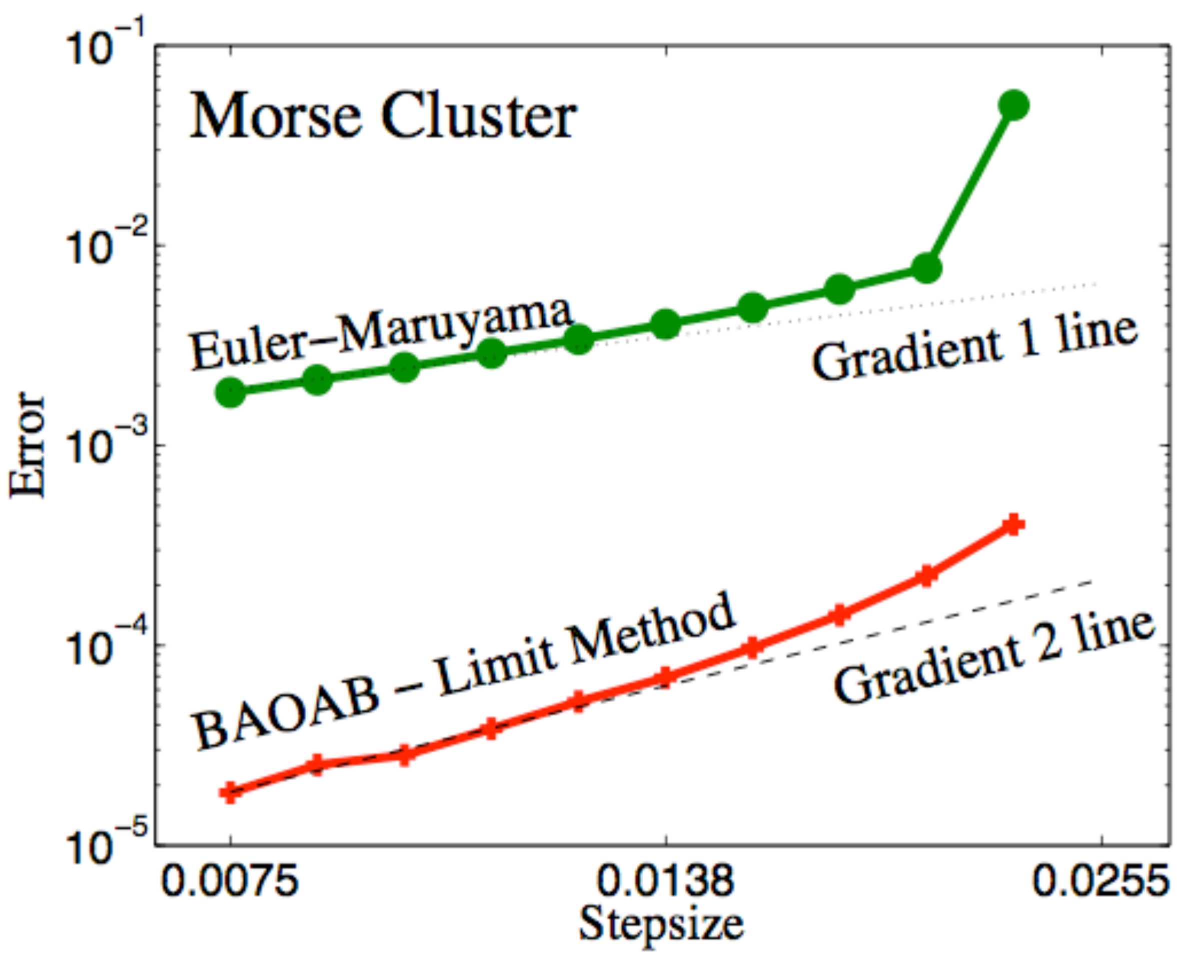
large stepsize

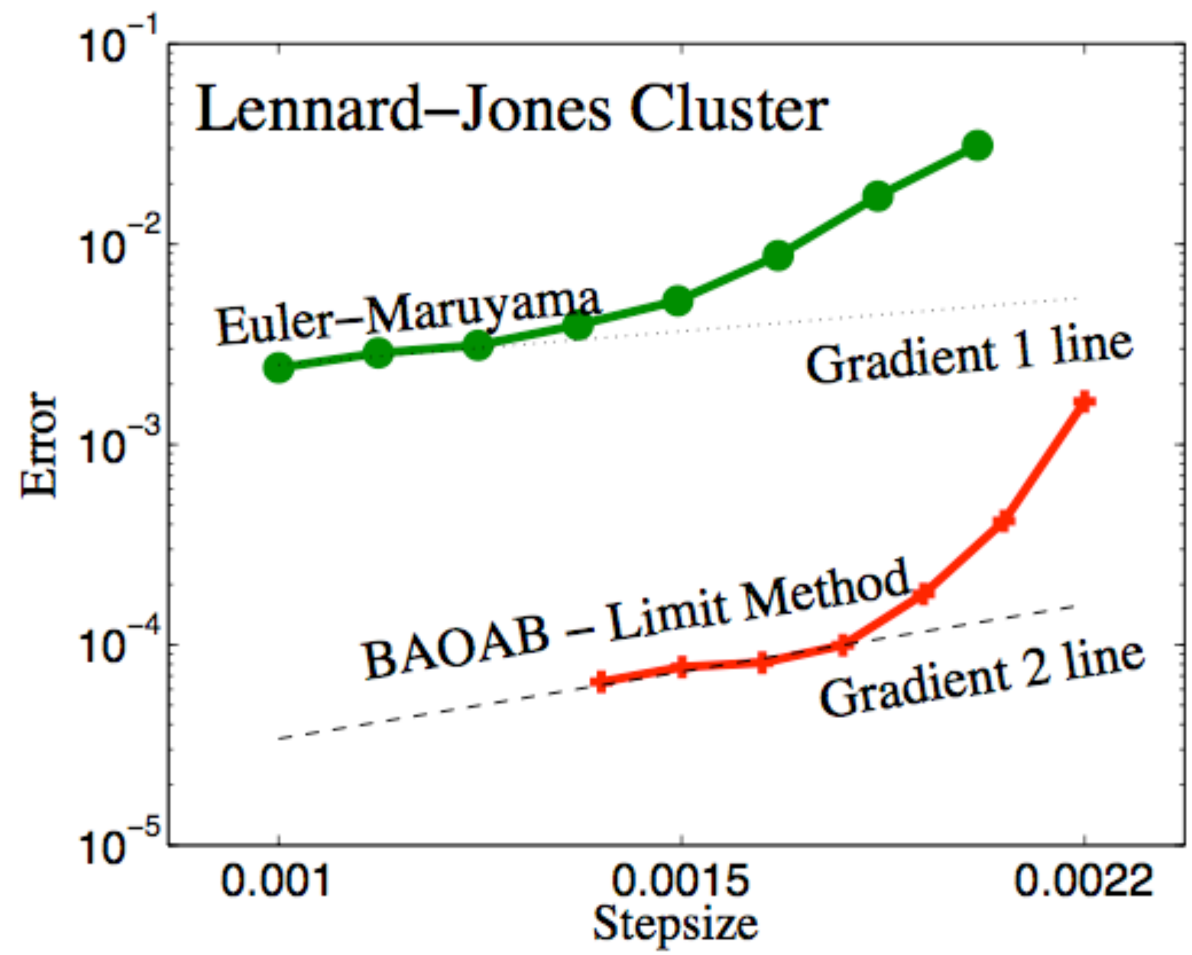


Morse and Lennard Jones Clusters

binned radial density







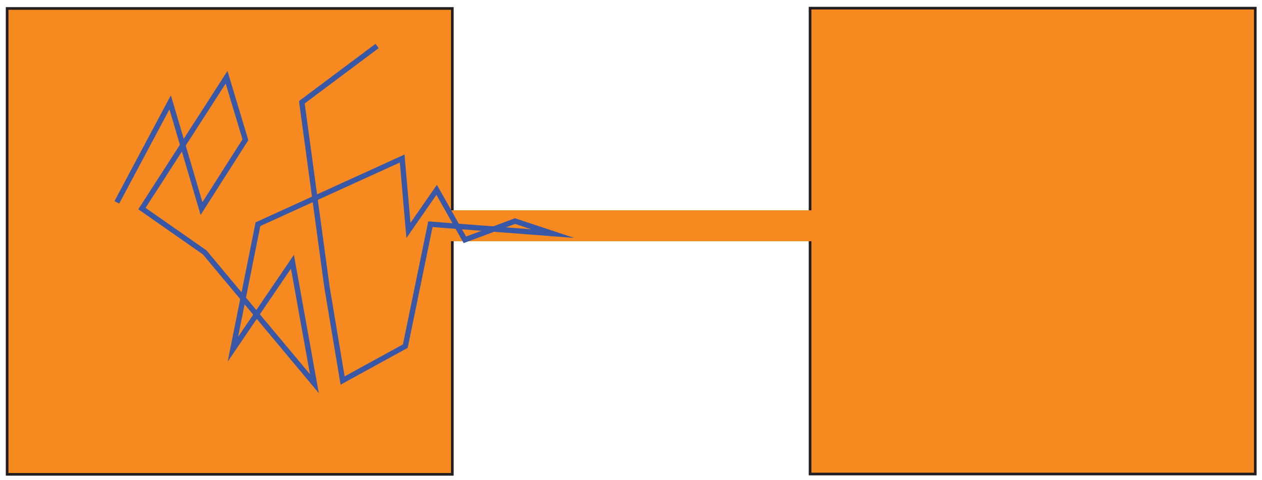
Accuracy ≠ Sampling Efficiency

Most sampling calculations are performed in the **pre-converged regime** (not at infinite time).

The challenge is often **effective search in a high dimensional space riddled with entropic barriers**

Brownian (first order) dynamics is *"non-inertial"*

Langevin (inertial) stochastic dynamics, at low friction, **can enhance diffusion** through entropic barriers



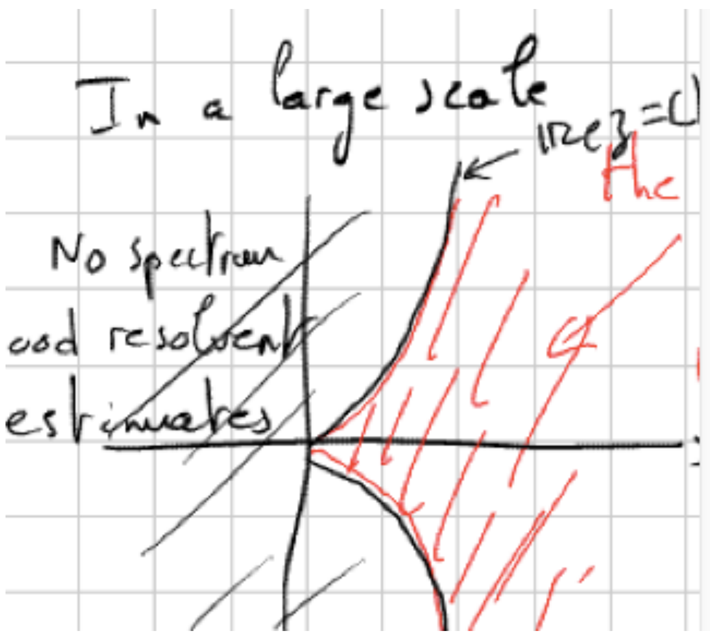
Langevin Dynamics

$$\mathrm{d}x = M^{-1}p\mathrm{d}t$$

$$\mathrm{d}p = -\nabla U \mathrm{d}t - \gamma M^{-1}p\mathrm{d}t + \sqrt{2\beta^{-1}\gamma}\mathrm{dW}$$

With Periodic Boundary Conditions
and smooth potential, ergodic sampling
of the canonical distribution with density

$$\rho_{\mathrm{can}} \propto e^{-\beta(p^T M^{-1} p/2 + U(q))}$$



courtesy F.Nier

Splitting Methods for Langevin Dynamics

$$\mathscr{L} = \mathscr{L}_{\mathrm{A}} + \mathscr{L}_{\mathrm{B}} + \mathscr{L}_{\mathrm{O}}$$

$$\mathscr{L}_{\mathrm{A}} = (M^{-1}p) \cdot \nabla_x$$

$$\mathscr{L}_{\mathrm{B}} = -\nabla U(x) \cdot \nabla_p$$

$$\mathscr{L}_{\mathrm{O}} = -\gamma(M^{-1}p) \cdot \nabla_p + \gamma\beta^{-1}\Delta_p$$

$$e^{h\hat{\mathscr{L}}_{\mathrm{BAOAB}}} = e^{\frac{h}{2}\mathscr{L}_{\mathrm{B}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{A}}} e^{h\mathscr{L}_{\mathrm{O}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{A}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{B}}}$$

Expansion of the invariant distribution

$$[\mathscr{L}^\dagger + h^2 \mathscr{L}_2^\dagger + \ldots] e^{-\beta(H + h^2 f_2 + \ldots)} = 0$$

Leading order:

$$\mathscr{L}^\dagger(\rho_{\rm can} f_2) = \beta^{-1} \mathscr{L}_2^\dagger \rho_{\rm can}$$

L. & Matthews, AMRX, 2013

L., Matthews, & Stoltz, IMA J. Num. Anal. 2015

- detailed treatment of all 1st and 2nd order splittings
- estimates for the operator inverse and justification of the expansion
- treatment of nonequilibrium (e.g. transport coefficients)

Configurational Sampling

The Magic Cancellation: (BL&CM 2013)
The marginal (configurational) distribution of the **BAOAB** method has an expansion of the form

$$\bar{\rho}_h = e^{-\beta U}[1 + O(h^2/\gamma^2) + O(h^4)]$$

In the high friction limit: 4th order, and with just one force evaluation per timestep.

Weak accuracy order = 2 but for high friction,
4th order in the invariant measure.

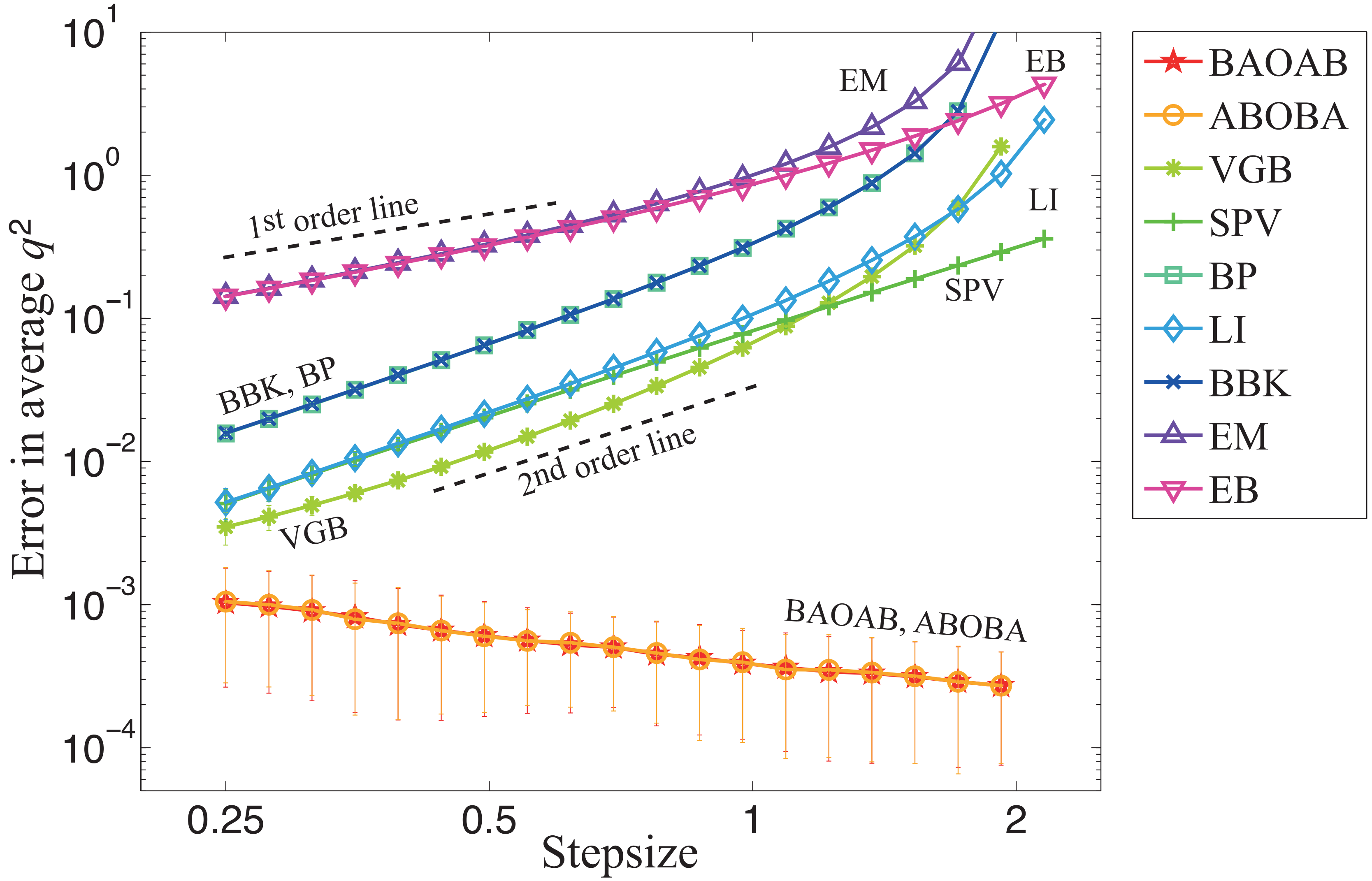
Harmonic Oscillator Configurational Sampling

L. & Matthews, *J. Chem. Phys.*, 2014

ABOBA and **BAOAB** are ***exact*** for configurations

Scheme	$\langle q^2 \rangle$	Scheme	$\langle q^2 \rangle$
Exact	$K^{-1}\beta^{-1}$	SPV	$K^{-1}\beta^{-1}\left(\gamma\,\delta t\frac{1-e^{-2\gamma\delta t}}{2(1-e^{-\gamma\delta t})^2}\right)$
BAOAB	$K^{-1}\beta^{-1}$	LI	$K^{-1}\beta^{-1} - \frac{\delta t^2}{12M\beta} + O\left(\delta t^4\right)$
ABOBA	$K^{-1}\beta^{-1}$	VGB	$K^{-1}\beta^{-1} + \frac{\gamma^2 M - 2K}{24M\beta K}\delta t^2 + O\left(\delta t^4\right)$
BBK	$K^{-1}\beta^{-1}\left(1 - \frac{\delta t^2 K}{4M}\right)^{-1}$	EM	$K^{-1}\beta^{-1}\left(1 - \frac{\delta t\, K}{2\gamma M}\right)^{-1}$
BP	$K^{-1}\beta^{-1}\left(1 - \frac{\delta t^2 K}{4M}\right)^{-1}$	EB	$K^{-1}\beta^{-1} + \frac{\delta t}{2\gamma M\beta} + O\left(\delta t^2\right)$

Harmonic Oscillator Configurational Sampling



BAOAB and **ABOBA** <u>both</u> are exact for PE

*But...this is only part of the story since **BAOAB** is much better than **ABOBA** for real molecules*

Anharmonic model problem

$$U(q) = \frac{q^2}{2} + \eta q^4$$

Asymptotic analysis of sampling error of $\langle q^2 \rangle$

typical, e.g. **OBABO**:	$O(\delta t^2)$
ABOBA:	$O(\eta \delta t^2)$
BAOAB:	$\dfrac{9\eta^2 \delta t^2}{6\gamma^2 + 8} + O(\eta^2 \delta t^4)$

but….

What to do about the force error?

$$\tilde{F}(x) = -\nabla U(x) + \eta(x)$$

a sampling error… it seems natural to take

$$\eta(x) \sim \mathcal{N}(0, \sigma^2(x))$$

and also, at least in the first stage, to assume $\sigma(x) \approx \sigma$

$$\begin{aligned} x_{n+1} &= x_n + hF(x_n) + h\sigma\tilde{R}_n + \sqrt{2h}R_n \\ &= x_n + hF(x_n) + \sqrt{h}\sqrt{h\sigma^2 + 2}\hat{R}_n \end{aligned}$$

Like Euler-Maruyama discretization of

$$\mathrm{d}x = F(x)\mathrm{d}t + \sqrt{2 + \sigma^2 h}\mathrm{dW}$$

$$\mathrm{d}x = F(x)\mathrm{d}t + \sqrt{2 + \sigma^2 h}\mathrm{dW}$$

1. Stepsize-dependent dynamics (like in B.E.A.)
2. Distorts temperature
3. Easy to correct - if we know σ
4. Computing/estimating σ can be difficult in practice

Options:

Monte-Carlo based approach [Ceperley et al, ‘Quantum Monte Carlo’ 1999]
Stochastic Gradient Langevin Dynamics [Welling, Teh, 2011]
Adaptive Thermostat [Jones and L., 2011]

The Adaptive Property

Jones & L. 2011

Applying Nosé-Hoover Dynamics to a system which is driven by white noise restores the canonical distribution.

<u>Adaptive (Automatic) Langevin</u>

$$\mathrm{d}x = M^{-1}p\mathrm{d}t$$

$$\mathrm{d}p = -\nabla U \mathrm{d}t - \sqrt{h}\sigma \mathrm{d}W - \xi p \mathrm{d}t + \sigma_A \mathrm{d}W_A$$

$$\mathrm{d}\xi = \mu^{-1}[p^T M^{-1} p - n\beta^{-1}]\mathrm{d}t$$

$$\tilde{\rho} = e^{-\beta[p^T M^{-1} p/2 + U(x)]} \times e^{-\beta\mu(\xi-\gamma)^2/2}$$ ergodic!

Shift in auxiliary variable by $\gamma = \dfrac{\beta(h\sigma^2 + \sigma_A^2)}{2\mathrm{Tr}(M)}$

Discretization

[With X. Shang, 2015]

generator: $\mathscr{L} = \mathscr{L}_{\mathrm{A}} + \mathscr{L}_{\mathrm{B}} + \mathscr{L}_{\mathrm{O}} + \mathscr{L}_{\mathrm{D}}$

$$\mathscr{L}_{\mathrm{A}} = (M^{-1}p) \cdot \nabla_x$$

$$\mathscr{L}_{\mathrm{B}} = -\nabla U(x) \cdot \nabla_p + \frac{h\sigma^2}{2}\Delta_p$$

$$\mathscr{L}_{\mathrm{O}} = -\xi p \cdot \nabla_p + \frac{\sigma_{\mathrm{A}}^2}{2}\Delta_p$$

$$\mathscr{L}_{\mathrm{D}} = G(p)\frac{\partial}{\partial \xi}$$

define related operator by composition, e.g. **BADODAB**

$$e^{h\hat{\mathscr{L}}} = e^{\frac{h}{2}\mathscr{L}_{\mathrm{B}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{A}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{D}}} e^{h\mathscr{L}_{\mathrm{O}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{D}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{A}}} e^{\frac{h}{2}\mathscr{L}_{\mathrm{B}}}$$

typically anticipate 2nd order (IM)

Superconvergence

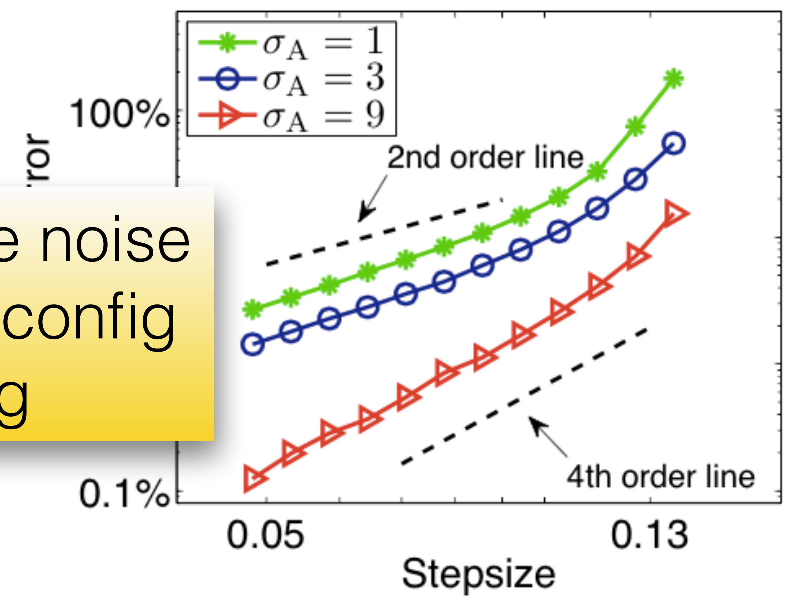
BAOAB, in the high friction limit, gives a superconvergence property for configurational quantities.

By taking large $\gamma \propto \sigma_A^2$ and $\mu \propto \sigma_A^2$ we can make BADODAB behave like BAOAB in the high friction limit after averaging over the auxiliary variable.

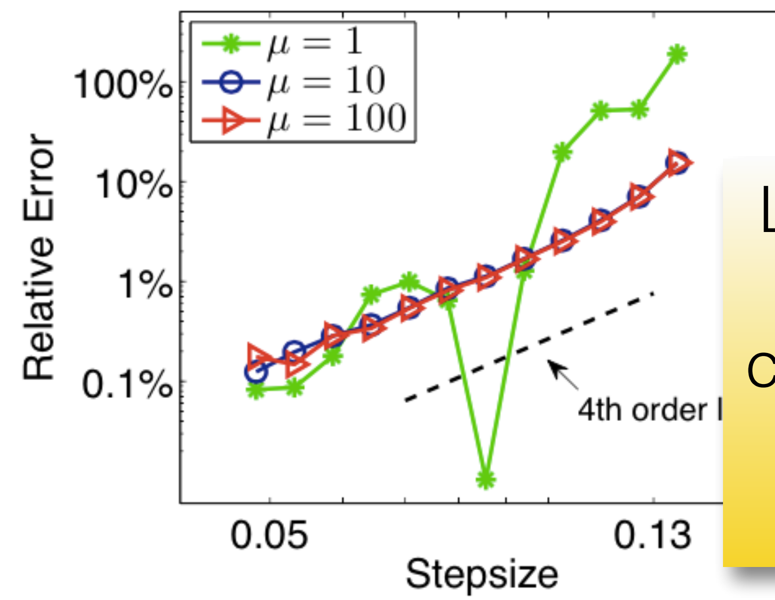
Effectively the extra driving noise implements a projection to the case of Langevin dynamics, **but large driving noise also implies large friction so restricted phase space exploration** (even if better accuracy). So caution is needed…

500 particles, clean gradient

configurational temperature

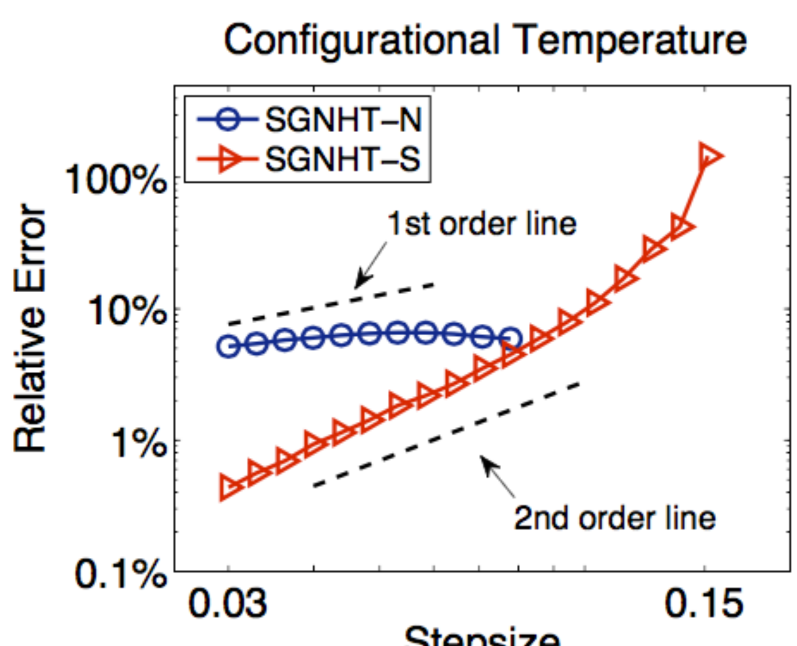


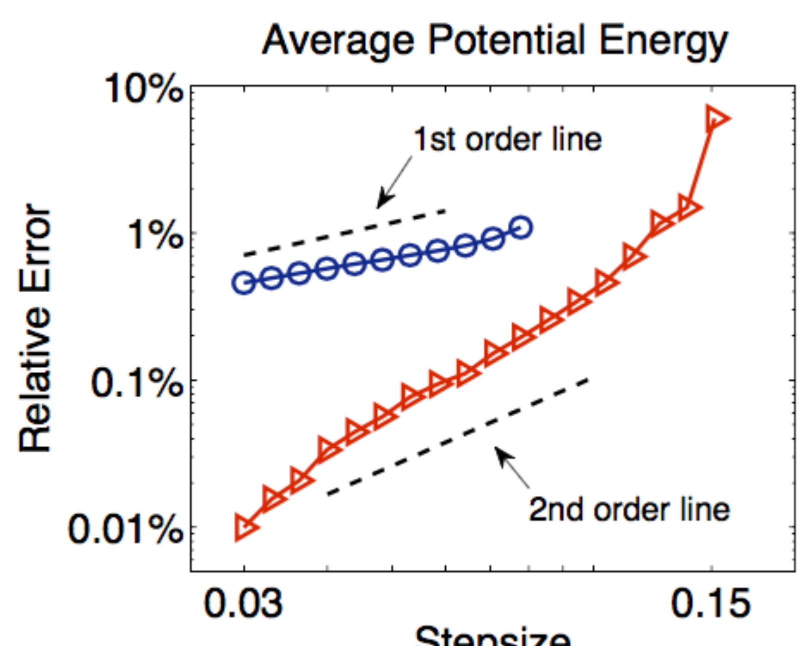
Large additive noise => 4th order config sampling



Large thermal mass = more stable control of distribution (but less responsive)

Blue: N. Ding et al. (Google Inc.) Vs. Red: Our method!





Bayesian Logistic Regression

$$\pi(y_i|\mathbf{x}_i, \boldsymbol{\beta}) = f(y_i \boldsymbol{\beta}^T \mathbf{x}_i)$$

f: logistic function

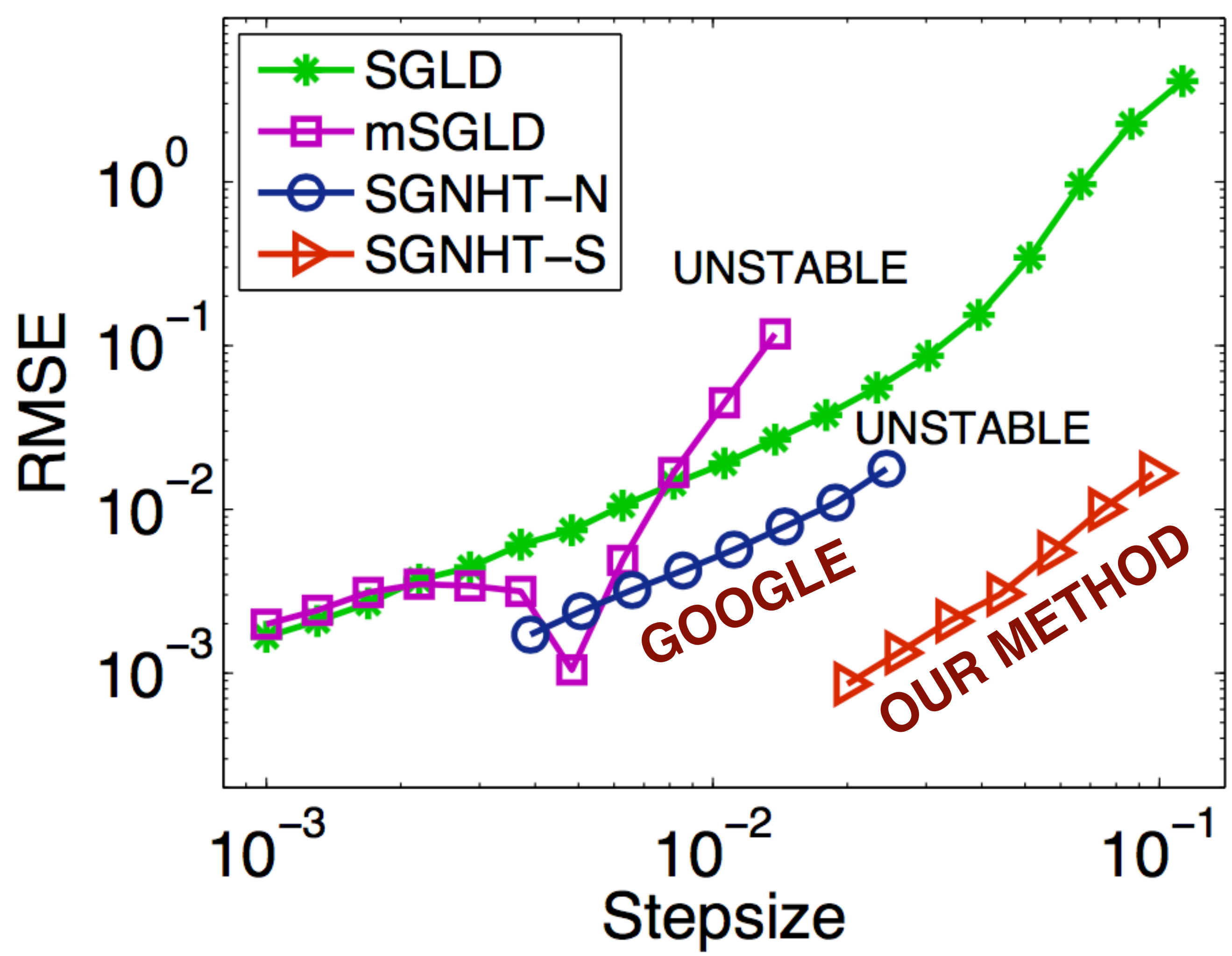
covariates e.g. age, income, …

data e.g. voting intention

posterior parameter distribution

$$\pi(\boldsymbol{\beta}) \propto \exp\left(-\frac{1}{2}\|\boldsymbol{\beta}\|^2\right) \prod_{i=1}^{N} f(y_i \boldsymbol{\beta}^T \mathbf{x}_i)$$

Gaussian prior



Covariance-Controlled Adaptive Langevin Dynamics

In the typical case, the noise may have a multivariate Gaussian distribution but with unknown (and evolving) covariance.

If we assume that we can obtain a covariance estimator then we can use this to enhance the accuracy of the SDEs.

CCAdL=
"Covariance Controlled Adaptive Langevin Dynamics" incorporates such a correction term together with an adaptive Langevin thermostat…

Formulation *[Shang, Zhu, L. & Storkey, NIPS, 2015]*

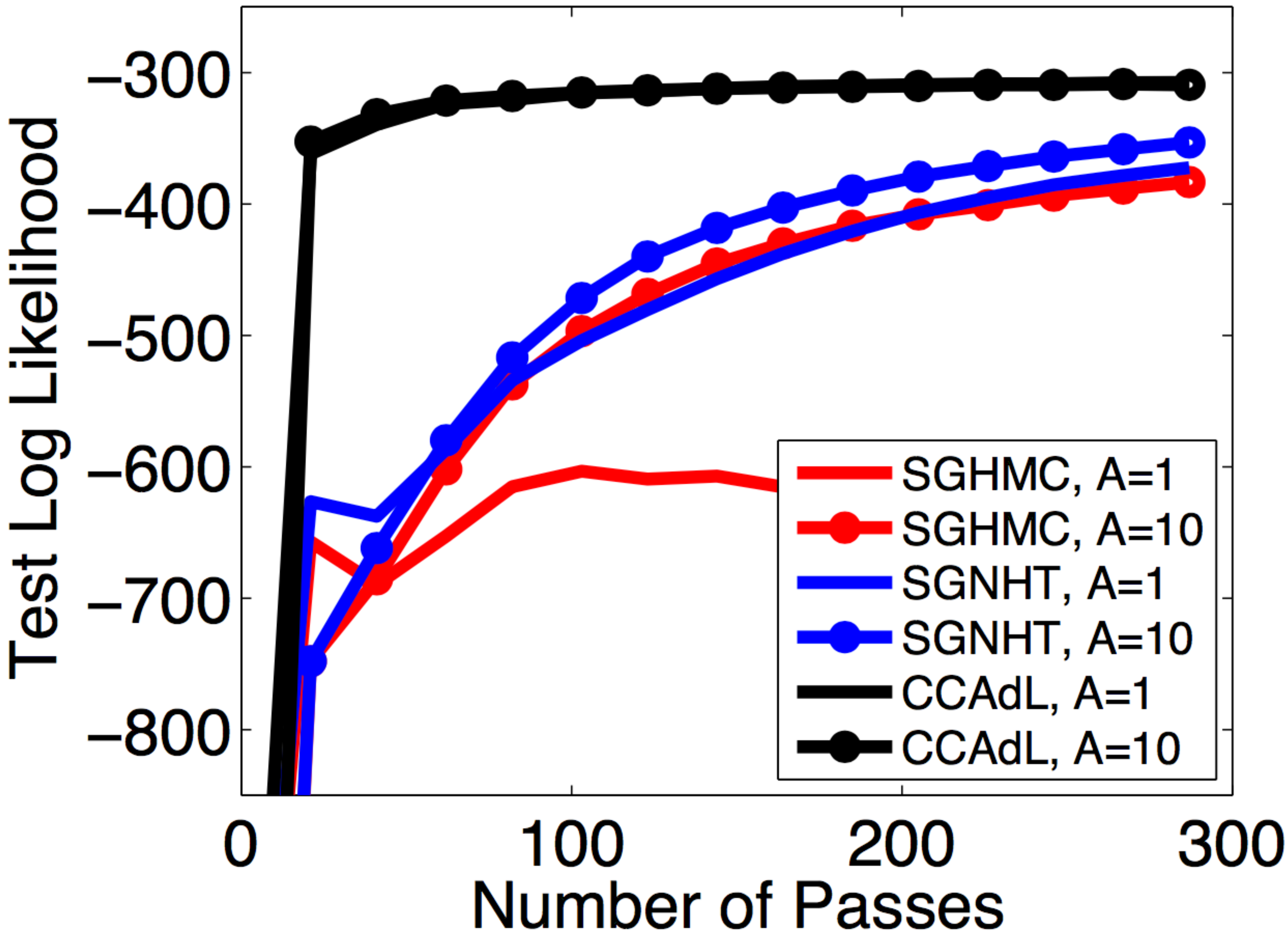
$$
\begin{aligned}
\mathrm{d}q &= M^{-1} p \,\mathrm{d}t \\
\mathrm{d}p &= -\nabla U(q)\mathrm{d}t + \sqrt{h}\Sigma(q)^{1/2} M^{1/2}\mathrm{dW} - \frac{h\beta}{2}\Sigma(q) p \,\mathrm{d}t \\
&\quad - \xi p \,\mathrm{d}t + \sqrt{2\hat{\gamma}\beta^{-1}} M^{1/2}\mathrm{dW}_{\mathrm{A}} \\
\mathrm{d}\xi &= \mu^{-1}\left[p^T M^{-1} p - N\beta^{-1}\right]\mathrm{d}t
\end{aligned}
$$

Invariant Distribution

$$
\tilde{\rho} \propto \exp(-\beta H(q,p)) \times \exp\left(-\frac{\beta\mu}{2}(\xi - \hat{\gamma})^2\right)
$$

Parameter-dependent noise dissipated by the added covariance-dependent term.

Classification Problem



Binary classification of handwritten digits 7 and 9.

Splittings *[Courtesy Xiaocheng Shang]*

$$
\begin{aligned}
\mathrm{d}q &= \underbrace{M^{-1}p\mathrm{d}t}_{\mathbf{A}} \\
\mathrm{d}p &= \underbrace{-\nabla U(q)\mathrm{d}t + \sqrt{h}\Sigma(q)^{1/2}M^{1/2}\mathrm{dW}}_{\mathbf{B}} \underbrace{- \frac{h\beta}{2}\Sigma(q)p\mathrm{d}t}_{\mathbf{C}} \\
&\quad \underbrace{- \xi p\mathrm{d}t + \sqrt{2\hat{\gamma}\beta^{-1}}M^{1/2}\mathrm{dW}_{\mathrm{A}}}_{\mathbf{D}} \\
\mathrm{d}\xi &= \underbrace{\mu^{-1}\left[p^T M^{-1} p - N\beta^{-1}\right]\mathrm{d}t}_{\mathbf{E}}
\end{aligned}
$$

Many palindromic words in this alphabet like
ABCDEDCBA, **BACDEDCAB**, etc.

But most are not 2nd order!

But **BAECDCEAB** is!

Summary

- Theory for invariant distributions for splitting type integrators for Langevin dynamics.

- Adaptive thermostat methods allowing control of error in the invariant distribution under stochastic perturbation of the force field.

- Extensions to exploit estimates of the covariance

- Superconvergence results but also acceleration of data analytics techniques

Ongoing work

- Application of these methods in non-equilibrium modelling

- Integrator order of accuracy (CCAdL)

- Accuracy/exploration balance in rare event modelling?

- Application to ab-initio MD (e.g. BOMD) [w. Jianfeng Lu (Duke) and Matthias Sachs (Edinburgh)

- Treatment of coloured noise.

More Details: ***Xiaocheng's Poster!***