Motivation & Contribution
1) Improving stationary mixing efficiency in SGMCMC by leveraging a generalized (potentially heavy-tailed) kinetics.
2) Alleviating numerical issue and satisfying conditions for stationarity by leveraging smooth version of generalized kinetics.
3) Ameliorating convergence issue by introducing additional first order dynamics and stochastic resampling.

**Theorem**

For univariate target distribution, the one time lag autocorrelation $p(x_t, x_{t+1})$ of the analytic MG-SS parameterized by $a$ asymptotically approaches zero when $a \to \infty$, under regularity condition of $U(x)$ and stationary assumption.

In addition to above, the MG-HMC with large $a$ is particularly advantageous for sampling multimodal distributions. Such a performance gain does not come in free.

SGMCMC
Sampling from $f(\theta) \propto \exp(-U(\theta))$ using minibatch data.

SGHMC (stochastic gradient Hamiltonian Monte Carlo)

- $\text{d} p = \partial p K(\theta)p d\theta$
- $\text{d} p = -\partial p U(\theta)d\theta - B(\theta) \partial p K(\theta)p d\theta + \mathcal{N}(0, 2B(\theta)d\theta)$.
- SGNHT (stochastic gradient Nosé-Hoover thermostat)

- $\text{d} p = -\partial p U(\theta)d\theta - \xi \partial p K(\theta)p d\theta + \mathcal{N}(0, 2B(\theta)d\theta)$
- $d\xi = (\dot{p} - 1)dt$.

Conclusion
- Scalable MCMC inference with generalized HMC variants.
- Future direction:
  - Adaptive selection of monomial parameters
  - Connection to optimization methods.