Stochastic Gradient Riemannian Langevin dynamics on the Probability Simplex

Sam Patterson (Gatsby, UCL)
Yee Whye Teh (Oxford) (presenter)

UCL

UNIVERSITY OF OXFORD

Poster: Fri32

Big Data and Bayesian Learning

- Large scale datasets are fast becoming the norm.
- Most current successes in scalable learning are optimization-based and non-Bayesian.

- Stochastic gradient Langevin dynamics:

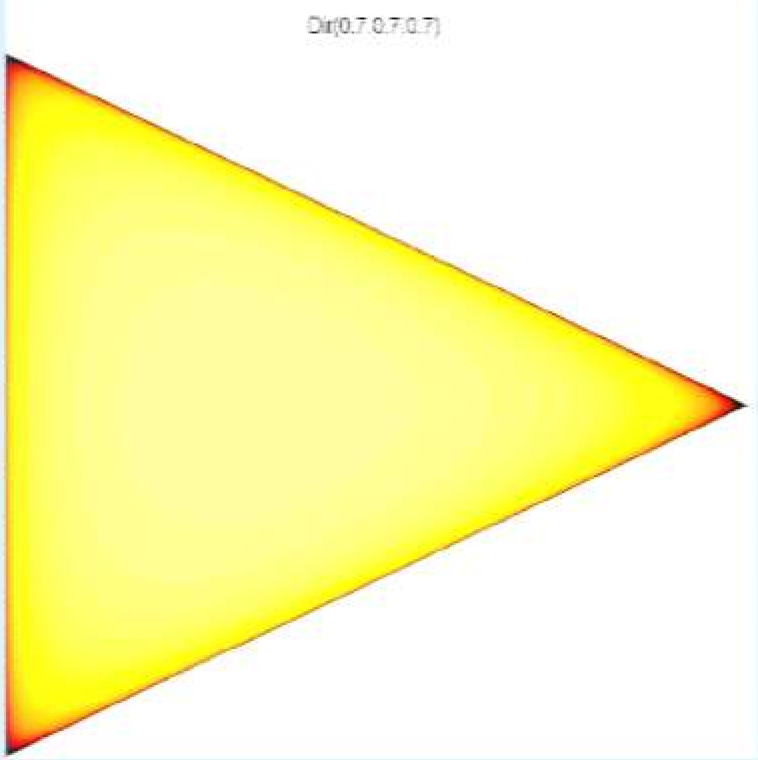
$$x_t = x_{t-1} + \frac{\epsilon_t}{2}\left(\nabla_X \log P(x_{t-1}) + \frac{N}{n}\sum_{i=1}^{n}\nabla_X \log P(y_{S_i}|x_{t-1})\right) + \mathcal{N}(0,\epsilon_t)$$

  - Step-sizes $\varepsilon_t \rightarrow 0$ slowly enough.

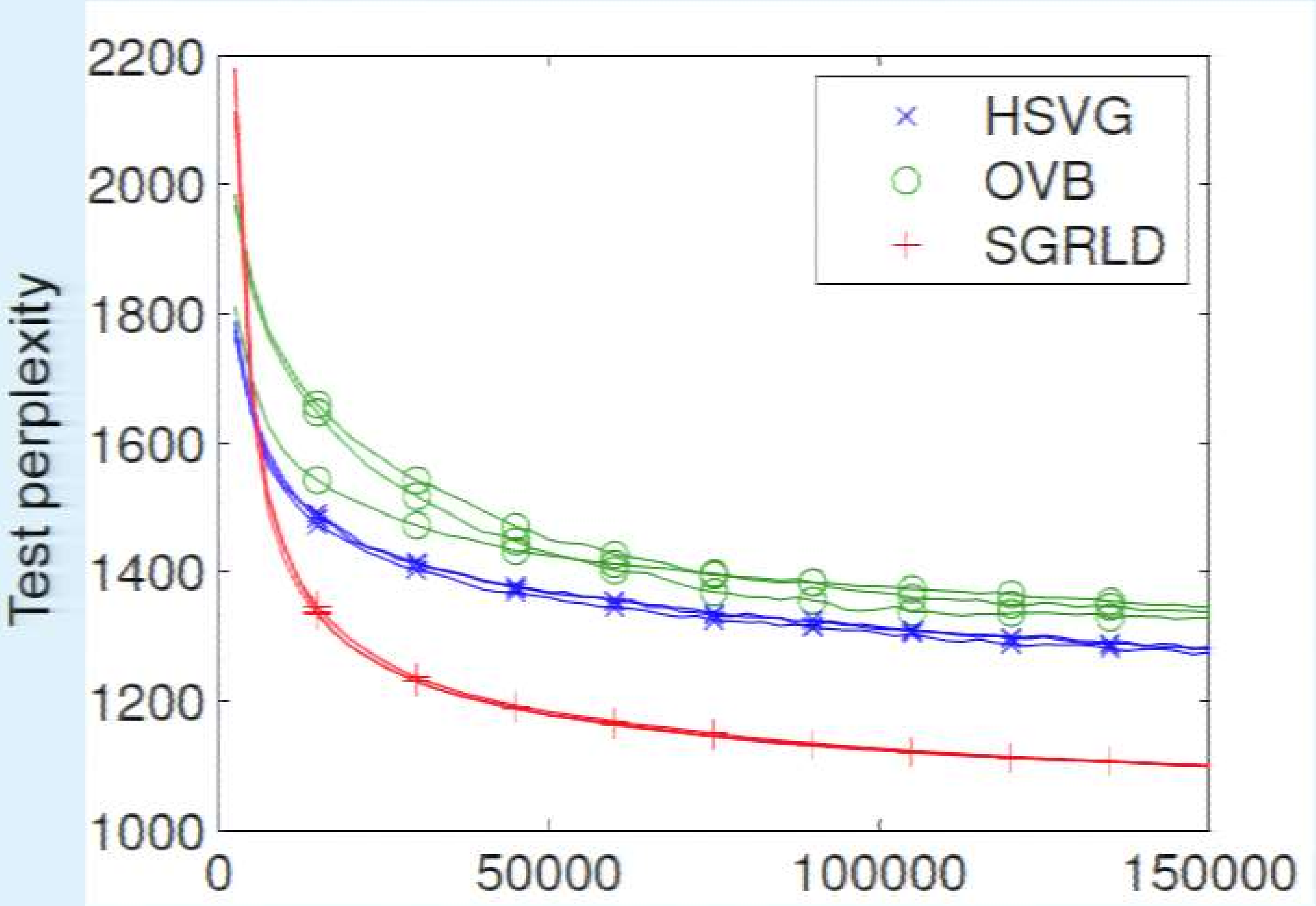
- [Welling & Teh 2011, Boyles et al 2012, Ahn et al 2012].

SGLD on Probability Simplices

- Many models defined over probability simplices, e.g. latent Dirichlet allocation.
- Complications:
  - High dimensional
  - Boundaries and constraints
  - Probability mass located close to boundaries.
  - Riemannian metric structure.
- Contribution:
  - Choice of parameterization
  - Choice of Riemannian metric



Learning LDA from Wikipedia



articles processed

OVB, HSVG
[Hoffman et al,
Mimno et al]

SGRLD