SGLD

Motivation

MCMC需要computate over the whole dataset

Large dataset会慢，

用Stochastic的话没有weight uncertainty并且会over fitting

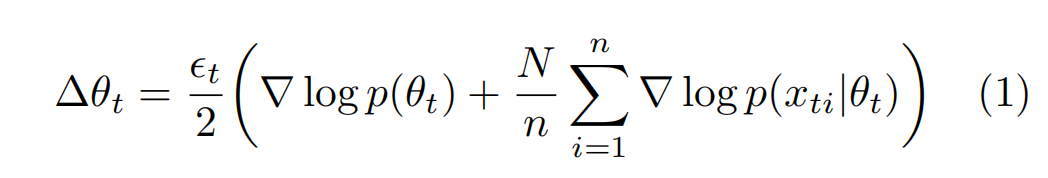
所以要先用Stochastic在一定的时候改用mcmc

Stochastic的parameter updating 公式和langevin equation相似所以采用LMC

Smooth

Stochastic

一次只用dataset中的一部分



x : subset

N : number of items in dataset

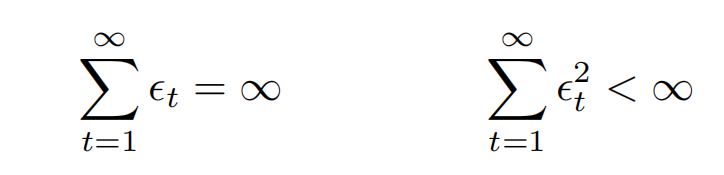
n : number of items subset of dataset

: sequence of step sizes.

每次iteration用不同的subset

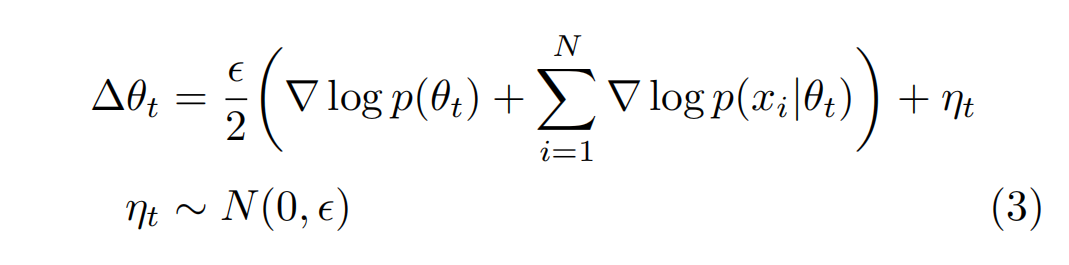
从Stochastic所产生的noise会被average out

Step size should satisfy



,

Langevin



Langevin dynamics is motivated and originally derived as a discretization of a stochastic differential equation whose equilibrium distribution is the posterior distribution.

Proposal distribution+MH

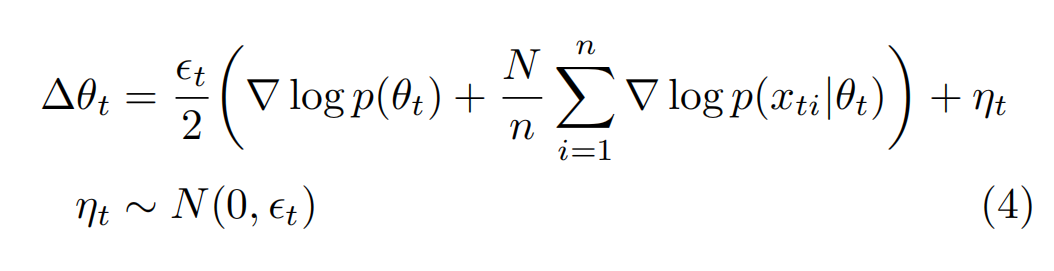
在降低时，error from discretization也会降低

过小的会导致converge变慢

使用HMC

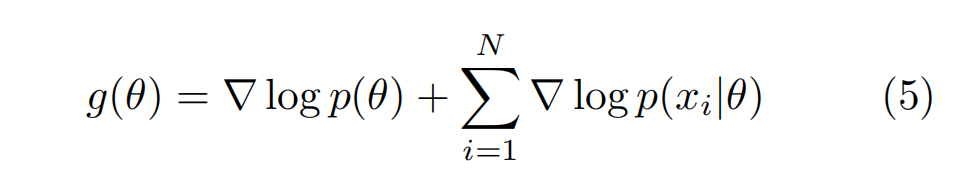
SGLD

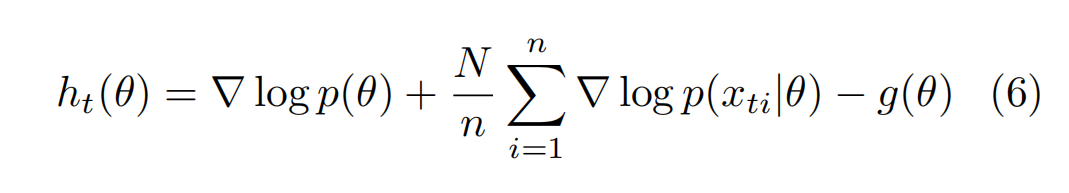
Combine together

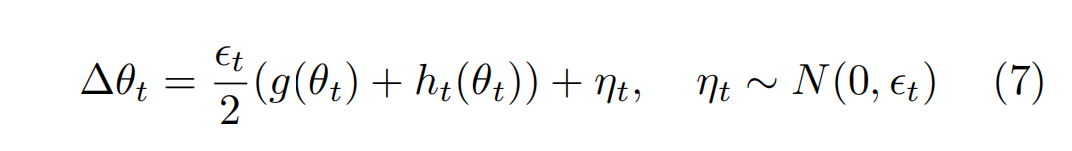


For large t (4) will approach to (3)

证明在 large t时Langevin dynamics会dominate





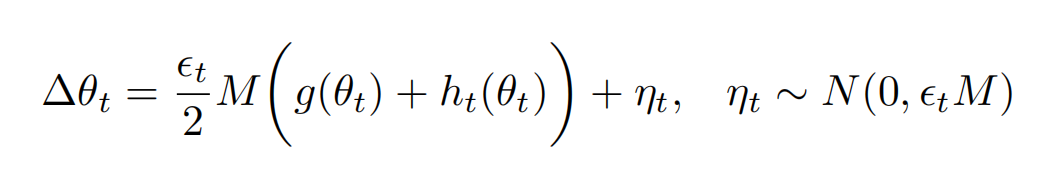


Two parts of

证明

会converge to the posterior

什么时候会变成langevin



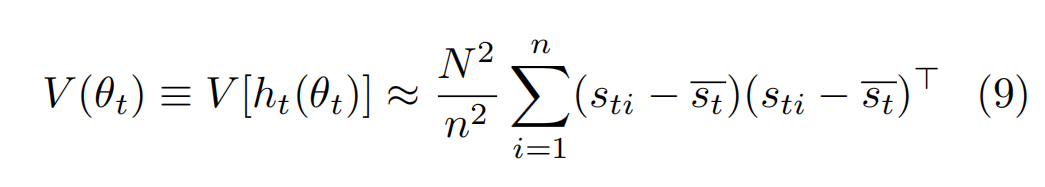
M is for making each dimension in the same scale, symmetric

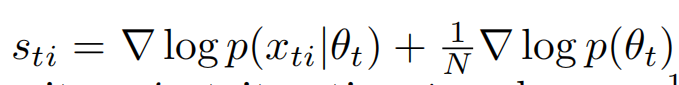
whether the algorithm is in the stochastic optimization phase or Langevin dynamics phase depends on the variance of the injected noise,

比较variance

injected noise的variance is

Stochastic部分：

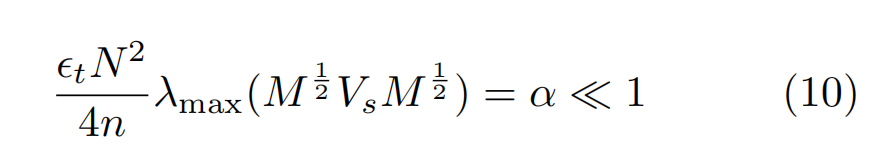




Stochastic 的variance是

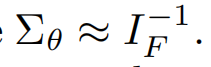
比较

所以为了让injected noise dominate需要



Max eigenvalue



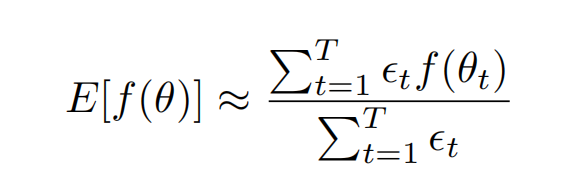


I 是 fisher information

是posterior的variance



Estimate Posterior Expectations



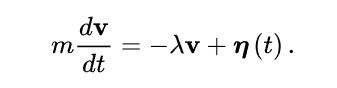
1 step

m=1

Then

Notation

Langevin equation

 ~ gaussian(0,1)

