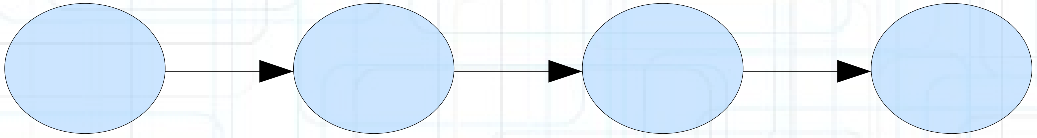
Markov Chain Monte Carlo

Overview

- Definitions
  - Markov Chain
  - Monte Carlo
- Kernels and Sampling
  - Gibbs / Metropolis-Hastings
- What I Did
  - Some example images

Markov Chain

- "a mathematical model for stochastic systems whose states, … are governed by a transition probability. The current state in a Markov chain only depends on the most recent previous states." [1]



Monte Carlo

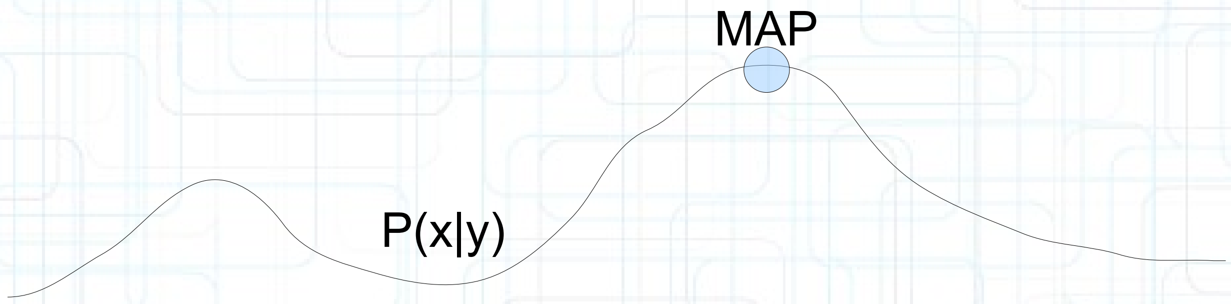
- The “random” part of the simulation:
  - Drawing from a given distribution with replacement.

MCMC

- A general technique for:
  - sampling from a posterior distribution,
  - integration in high dimensional space,
  - simulated annealing, and
  - learning. [1]

Differences

- Classical regularization gives a single answer
- MCMC explores a *distribution*
  - Both employ the same “regularization” (Tikhonov with Discrete Smoothing Norm, etc)



MCMC Kernel

- Metropolis Hastings
  - Propose a new distribution $x' = q(x', x^t)$
  - Accept dist. with probability $a = \frac{P(x')}{P(x^t)}$
  - Repeat $x^{t+1} = x'$ OR $x^{t+1} = x^t$
- Gibbs Sampling (a special case of ▲)
  - $x_1 \sim P(x_1 | x_2)$
  - $x_2 \sim P(x_2 | x_1)$
  - Repeat
    - NOTE: always accepted = no wasted work!

Burnin

- Samplers will converge to the target distribution
  - NEAT!

... but how long?

- Some number of iterations are required to achieve convergence before the samples will accurately represent the distribution

(unless the first guess is good)

Fancier Kernels

- Some redundant work is done by the samplers since they are performing a “random walk” and may be sampling the same region repeatedly.
  - by adding concept of “momentum”
- No free lunch...
  - There's no one recipe for all problems

What Did I Do?

- CT Back Projection
- Gibbs Sampler
  - Modelled randomized normally distributed errors
    - Additive measurement noise
    - Additive prior (x0 ~ 0)
    - Modelling error (projection angle)

      = multiplicative modelling noise
    - Hyper-parameter (Rayleigh dist.)

Tikhonov Regularization

$$min\{\ \|Ax-b\|_2^2\ +\ \lambda^2\|L(x-x_0)\|_2^2\ \}$$

- Solved via Least Squares

$$x=\begin{bmatrix}Ax\\ \lambda L\end{bmatrix}\backslash\begin{bmatrix}b\\ \lambda L x_0\end{bmatrix}$$

Gibbs

- Calculate and Repeat for K samples:

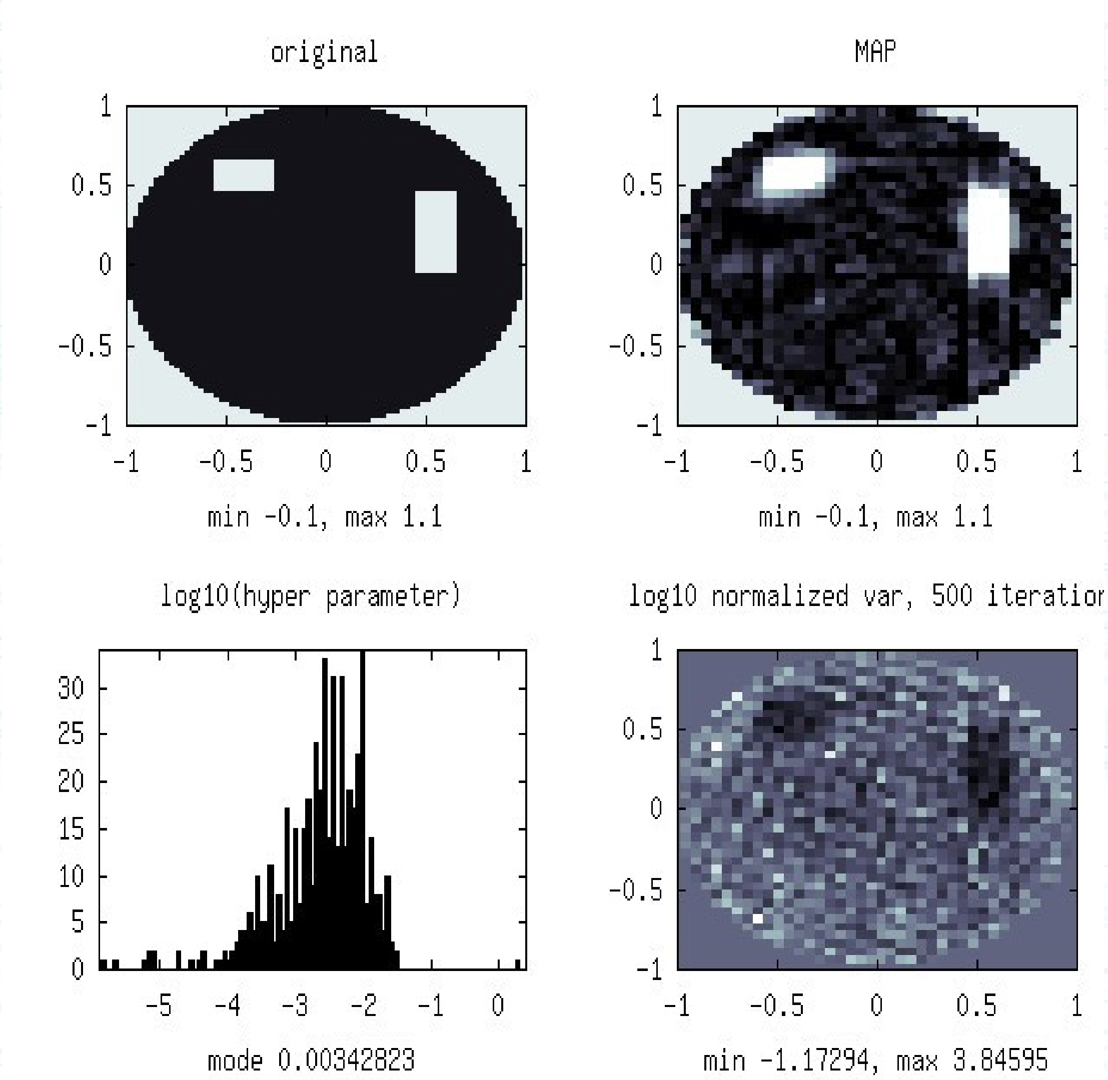
$$\lambda \sim P(\lambda|x) = \lambda^{(n+2)} \exp\left(\frac{-1}{2}\frac{\lambda^2}{\sigma_p^2}\|L(x-x_0)\|^2 - \frac{1}{2}\left(\frac{\lambda^2}{\lambda_0^2}\right)\right)$$

$$A = proj(angles + noise)$$

$$x = \begin{bmatrix} \sigma_b^{-1} A x \\ \lambda \sigma_p^{-1} L \end{bmatrix} \backslash \begin{bmatrix} \sigma_b^{-1} b + \eta \\ \lambda \sigma_p^{-1} L x_0 + \zeta \end{bmatrix}$$

- Converges towards expected hyper-parameter vs. L-curve

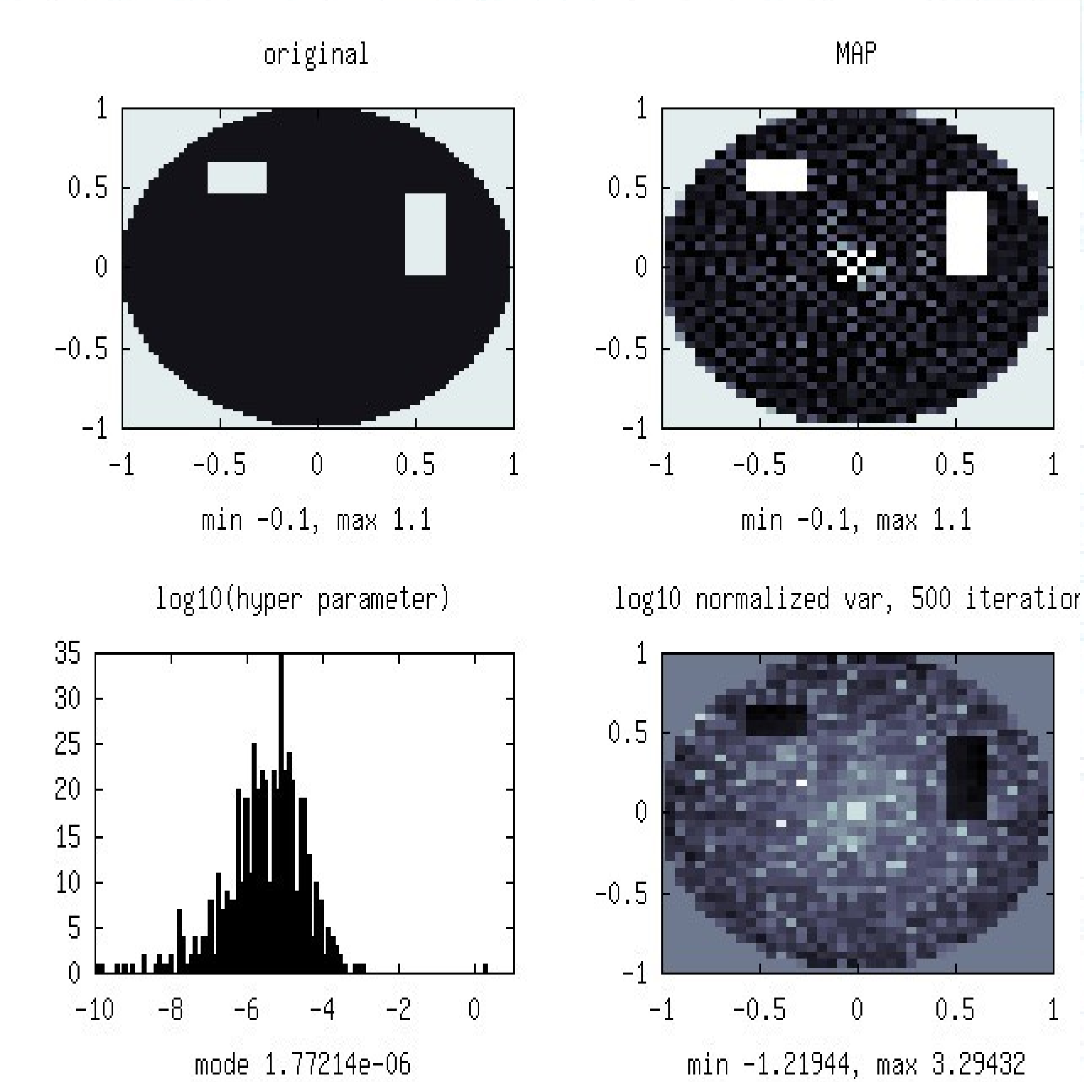
500 samples, 10 angles



1h

$\lambda_0 = 10$

500 samples, 60 angles



1.5h

$\lambda_0 = 10$

Discussion

[http://www.clipartguide.com/_pages/0808-0712-3117-5830.html]

References

- MCMC Tutorial, http://www.civs.ucla.edu/MCMC/MCMC_tutorial.htm, visited Apr 5, 2009
- Wikipedia: Markov Chain Monte Carlo,http://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo, visited Apr 5, 2009
- J Kaipio, E Somersalo, Statistical and Computational Inverse Problems, Springer, 2005