Parallel Chromatic MCMC with Spatial Partitioning

Jun Song, David Moore

UC Berkeley

2/5/2017

Introduction

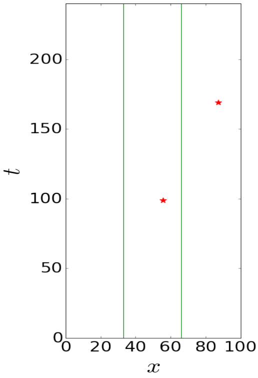
- New approach for parallel MCMC inference exploiting spatially induced conditional independence.
- Motivated by scaling "Global seismic monitoring as probabilistic inference" (Arora et al. 2010)
- Complex model, big data, open-universe: no fixed set of random variables
- This talk: we'll introduce a toy spatial model to motivate and demonstrate our proposed approach.

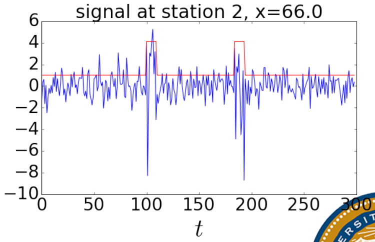
Related work

- Fugue (Kumar et al. 2014)
- On model parallelization and scheduling strategies for distributed machine learning (Lee et al. 2014)
- Distributed algorithms for topic models (Newman et al. 2009)
- An architecture for parallel topic models (Smola and Narayanamurthy 2010)
- Parallel Markov chain Monte Carlo for nonparametric mixture models (Williamson, Dubey,and Xing 2013)
- Asymptotically exact, embarrassingly parallel mcmc (Neiswanger, Wang, and Xing 2013)

Our Model

- Simplified seismic event detection model (Arora et al. 2010)
  - Prior on unknown number of events, location/times.
  - Likelihood: given events, models arrival time and signal generation at stations.



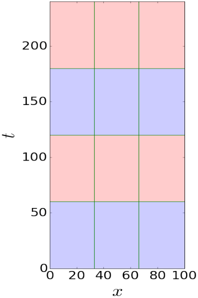


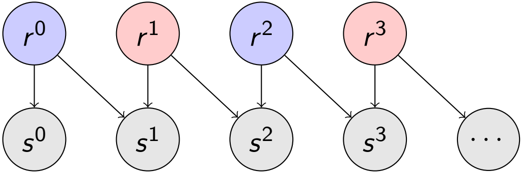
MCMC Inference

- Serial Metropolis-Hastings with multiple proposal types
    - Birth and Death Moves
    - Event Location Move
    - Arrival Time Move
- Birth/death moves change model structure by adding/removing random variables.

Partitioning the Space

- Travel times are bounded
- Induces conditional independence between distant events





Parallel MCMC algorithms

Baseline: Naive Parallel Metropolis Hastings:

- Runs independently inference on each region
- Incorrect, does not converge to the correct stationary distribution

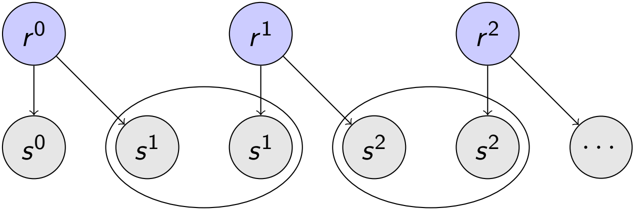
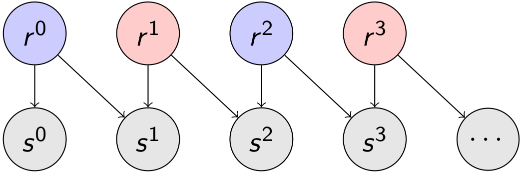


Figure 2: factor graph of naive m-h

Chromatic Metropolis Hastings

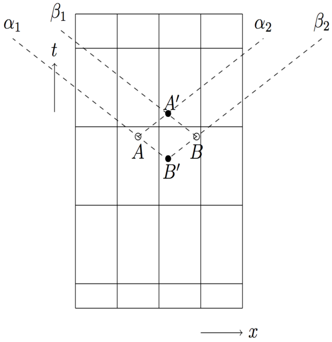
- Inspired by Chromatic Gibbs Sampling (Gonzalez et al., 2011)
- We color spatial regions instead of fixed random variables.
- Can run MH moves in parallel on all regions of each color.



- The joint chain is ergodic and has the correct stationary distribution

Dynamic Coloring

- Static region boundaries can slow mixing
- A simple trick for fast mixing: dynamic coloring
- Offset the region boundary after a round of inference



Experimental Results

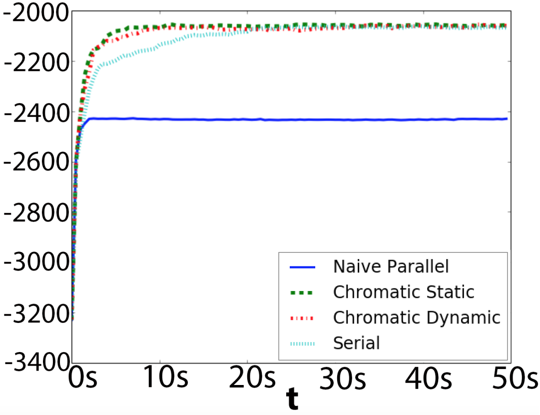


Figure 3: Log probability, higher is better

Experimental Results

- Recall: fraction of true events recovered by inference.
- Dynamic coloring yields faster mixing.

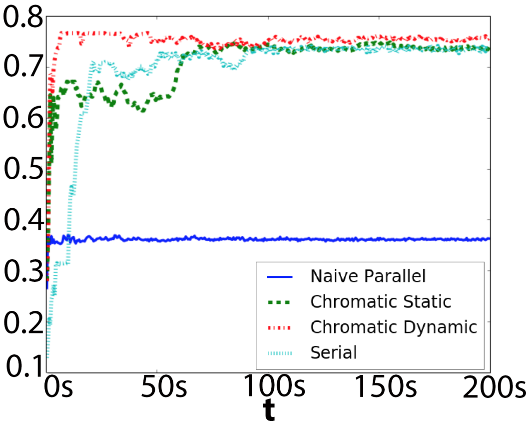


Figure 4: Recall vs Time, higher is better

Experimental Results

- Recall over 5 inference runs on 50 sampled possible worlds.
- Even though static regions are correct, faster mixing from dynamic partitions yields better results.

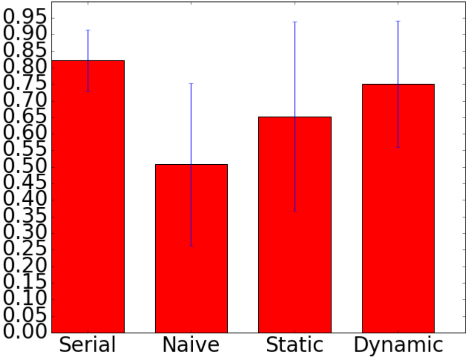


Figure 5: Recall

Conclusion and Future Work

- Conclusion
  - Introduced a novel approach for parallel MCMC inference exploiting value-dependent conditional independence induced by spatial structure rather than a fixed graphical model.
  - Achieved speedup while still maintaining a similar level of error, precision and recall as Serial Metropolis Hastings.
- Future Work
  - Extend this approach to production-scale models such as "Global seismic monitoring as probabilistic inference" (Arora et al. 2010).
  - Extend to other applications involving spatial object detection and/or localization.