MCMC with Multiple Proposals

Andrew Holbrook

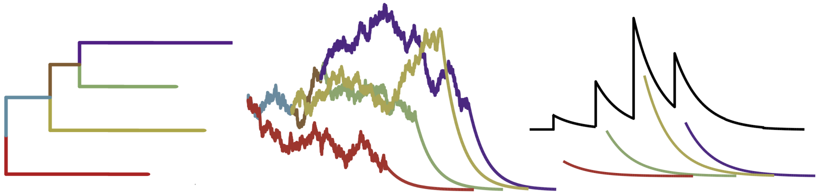
UCLA Biostatistics

May 18, 2023

At the Intersection of Big Data and Big Model

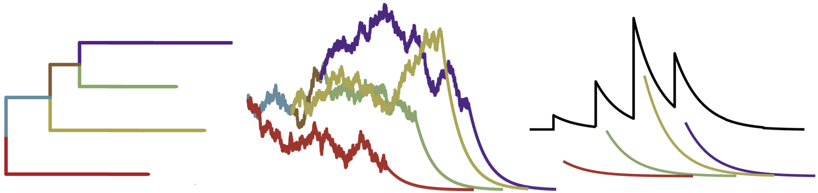
A Unified Model for Viral Spread

- Holbrook, Ji and Suchard (2022). *From viral evolution to spatial contagion: a biologically modulated Hawkes model*, Bioinformatics.



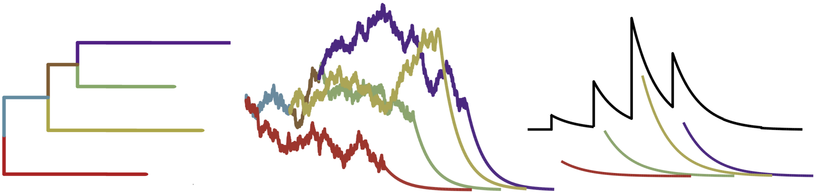
A Unified Model for Viral Spread

- Holbrook, Ji and Suchard (2022). *From viral evolution to spatial contagion: a biologically modulated Hawkes model*, Bioinformatics.
- Virus-specific latent variables connect a spatiotemporal Hawkes process model with a phylogenetic diffusion prior.



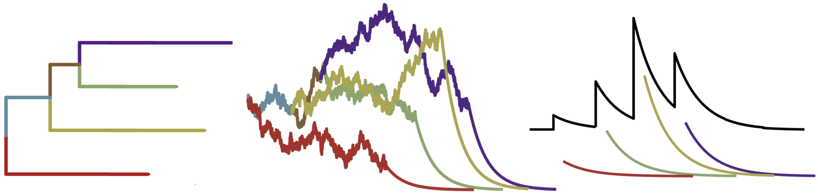
A Unified Model for Viral Spread

- Holbrook, Ji and Suchard (2022). *From viral evolution to spatial contagion: a biologically modulated Hawkes model*, Bioinformatics.
- Virus-specific latent variables connect a spatiotemporal Hawkes process model with a phylogenetic diffusion prior.
- The number of latent variables is $\mathcal{O}(N)$, for N the number of observed viruses.



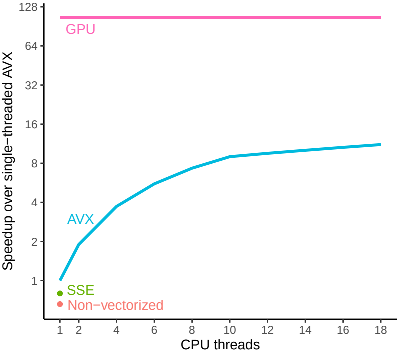
A Unified Model for Viral Spread

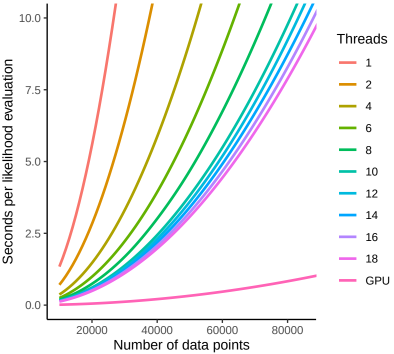
- Holbrook, Ji and Suchard (2022). *From viral evolution to spatial contagion: a biologically modulated Hawkes model*, Bioinformatics.
- Virus-specific latent variables connect a spatiotemporal Hawkes process model with a phylogenetic diffusion prior.
- The number of latent variables is $\mathcal{O}(N)$, for N the number of observed viruses.
- Hawkes likelihood computations require $\mathcal{O}(N^2)$ floating-point operations.



A Unified Model of Viral Spread

Proposal #1: use parallel computing to accelerate MH bottleneck, i.e., likelihood computations.

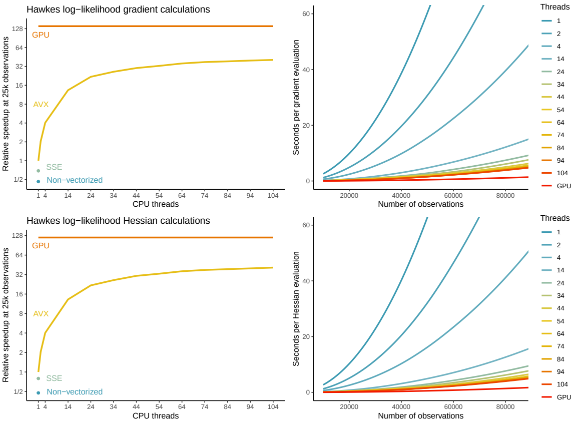




What about high dimensionality?

A Unified Model of Viral Spread

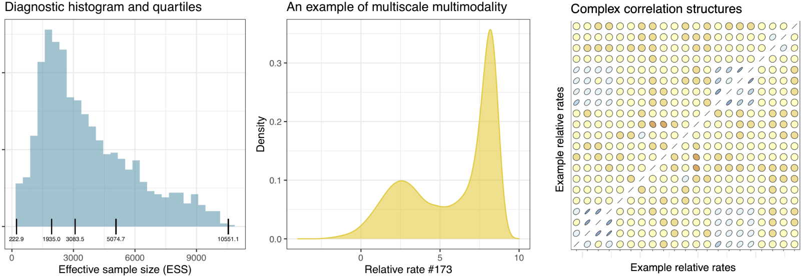
Proposal #2: also use parallel computing to accelerate adaptive HMC bottlenecks, i.e., log-likelihood gradient/Hessian.



What about bad geometry (non-linearity, multimodality)?

A Unified Model of Viral Spread

Proposal #3: to analyze over 23k Ebola cases (2014-2016 West Africa), run the chain for 30 days using Nvidia GV100 GPU.



Inexhaustive Taxonomy of Parallel MCMC

- Between-chain parallelization: multiple independent chains;
- Within-chain parallelization:
  - Model-dependent parallelization: likelihood computations;
  - Model-independent parallelization:
    - Parallel tempering;
    - Multiple-try metropolis;
    - Multiple proposals, single acceptance step.

MCMC with Multiple Proposals

A Complicated Landscape

- Ceperly et al. (1977). *Monte Carlo simulation of a many fermion study*, Phys. Rev. B. (Physics)

A Complicated Landscape

- Ceperly et al. (1977). *Monte Carlo simulation of a many fermion study*, Phys. Rev. B. (Physics)
- Neal (2003). *Markov chain sampling for non-linear state space models using embedded hidden Markov models*, Preprint. (Stats)

A Complicated Landscape

- Ceperly et al. (1977). *Monte Carlo simulation of a many fermion study*, Phys. Rev. B. (Physics)
- Neal (2003). *Markov chain sampling for non-linear state space models using embedded hidden Markov models*, Preprint. (Stats)
- Frenkel (2004). *Speed-up of Monte Carlo simulations by sampling of rejected states*, PNAS. (Physics), ("Waste-Recycling", "Coarse-grained moves", "Macro-steps", "Superdetailed balance")

A Complicated Landscape

- Ceperly et al. (1977). *Monte Carlo simulation of a many fermion study*, Phys. Rev. B. (Physics)
- Neal (2003). *Markov chain sampling for non-linear state space models using embedded hidden Markov models*, Preprint. (Stats)
- Frenkel (2004). *Speed-up of Monte Carlo simulations by sampling of rejected states*, PNAS. (Physics), ("Waste-Recycling", "Coarse-grained moves", "Macro-steps", "Superdetailed balance")
- Tjelmeland (2004). *Using all Metropolis-Hastings proposals to estimate mean values*, Preprint. (Stats), ("Weighted-mean", not "Rao-Blackwellized")

A Complicated Landscape

- Ceperly et al. (1977). *Monte Carlo simulation of a many fermion study*, Phys. Rev. B. (Physics)
- Neal (2003). *Markov chain sampling for non-linear state space models using embedded hidden Markov models*, Preprint. (Stats)
- Frenkel (2004). *Speed-up of Monte Carlo simulations by sampling of rejected states*, PNAS. (Physics), ("Waste-Recycling", "Coarse-grained moves", "Macro-steps", "Superdetailed balance")
- Tjelmeland (2004). *Using all Metropolis-Hastings proposals to estimate mean values*, Preprint. (Stats), ("Weighted-mean", not "Rao-Blackwellized")
- Neal (2011). *MCMC using ensembles of states for problems with fast and slow variables such as Gaussian process regression*, Preprint. (Stats), ("Ensemble MCMC", "Computational Short-cut")

A Complicated Landscape

- Ceperly et al. (1977). *Monte Carlo simulation of a many fermion study*, Phys. Rev. B. (Physics)
- Neal (2003). *Markov chain sampling for non-linear state space models using embedded hidden Markov models*, Preprint. (Stats)
- Frenkel (2004). *Speed-up of Monte Carlo simulations by sampling of rejected states*, PNAS. (Physics), ("Waste-Recycling", "Coarse-grained moves", "Macro-steps", "Superdetailed balance")
- Tjelmeland (2004). *Using all Metropolis-Hastings proposals to estimate mean values*, Preprint. (Stats), ("Weighted-mean", not "Rao-Blackwellized")
- Neal (2011). *MCMC using ensembles of states for problems with fast and slow variables such as Gaussian process regression*, Preprint. (Stats), ("Ensemble MCMC", "Computational Short-cut")
- Calderhead (2014). *A general construction for parallelizing Metropolis-Hastings algorithms*, PNAS. (Stats), "Generalized Metropolis-Hastings", simulates "Finite-state Markov Chain" over proposals)

A Complicated Landscape

- Ceperly et al. (1977). *Monte Carlo simulation of a many fermion study*, Phys. Rev. B. (Physics)
- Neal (2003). *Markov chain sampling for non-linear state space models using embedded hidden Markov models*, Preprint. (Stats)
- Frenkel (2004). *Speed-up of Monte Carlo simulations by sampling of rejected states*, PNAS. (Physics), ("Waste-Recycling", "Coarse-grained moves", "Macro-steps", "Superdetailed balance")
- Tjelmeland (2004). *Using all Metropolis-Hastings proposals to estimate mean values*, Preprint. (Stats), ("Weighted-mean", not "Rao-Blackwellized")
- Neal (2011). *MCMC using ensembles of states for problems with fast and slow variables such as Gaussian process regression*, Preprint. (Stats), ("Ensemble MCMC", "Computational Short-cut")
- Calderhead (2014). *A general construction for parallelizing Metropolis-Hastings algorithms*, PNAS. (Stats), "Generalized Metropolis-Hastings", simulates "Finite-state Markov Chain" over proposals)
- Delmas and Jourdain (2016). *Does Waste Recycling Really Improve the Multi-Proposal Metropolis–Hastings algorithm? an Analysis Based on Control Variates*, J. Appl. Prob. (Probability), ("Multi-proposal Metropolis-Hastings")

A Complicated Landscape

- Ceperly et al. (1977). *Monte Carlo simulation of a many fermion study*, Phys. Rev. B. (Physics)
- Neal (2003). *Markov chain sampling for non-linear state space models using embedded hidden Markov models*, Preprint. (Stats)
- Frenkel (2004). *Speed-up of Monte Carlo simulations by sampling of rejected states*, PNAS. (Physics), ("Waste-Recycling", "Coarse-grained moves", "Macro-steps", "Superdetailed balance")
- Tjelmeland (2004). *Using all Metropolis-Hastings proposals to estimate mean values*, Preprint. (Stats), ("Weighted-mean", not "Rao-Blackwellized")
- Neal (2011). *MCMC using ensembles of states for problems with fast and slow variables such as Gaussian process regression*, Preprint. (Stats), ("Ensemble MCMC", "Computational Short-cut")
- Calderhead (2014). *A general construction for parallelizing Metropolis-Hastings algorithms*, PNAS. (Stats), "Generalized Metropolis-Hastings", simulates "Finite-state Markov Chain" over proposals)
- Delmas and Jourdain (2016). *Does Waste Recycling Really Improve the Multi-Proposal Metropolis–Hastings algorithm? an Analysis Based on Control Variates*, J. Appl. Prob. (Probability), ("Multi-proposal Metropolis-Hastings")
- Yang et al. (2018). *On parallelizable Markov chain Monte Carlo algorithms with waste-recycling*, Stat. Comp. (Stats), ("Locally-weighted parallel MCMC" as "Rao-Blackwellization")

A Complicated Landscape

- Ceperly et al. (1977). *Monte Carlo simulation of a many fermion study*, Phys. Rev. B. (Physics)
- Neal (2003). *Markov chain sampling for non-linear state space models using embedded hidden Markov models*, Preprint. (Stats)
- Frenkel (2004). *Speed-up of Monte Carlo simulations by sampling of rejected states*, PNAS. (Physics), ("Waste-Recycling", "Coarse-grained moves", "Macro-steps", "Superdetailed balance")
- Tjelmeland (2004). *Using all Metropolis-Hastings proposals to estimate mean values*, Preprint. (Stats), ("Weighted-mean", not "Rao-Blackwellized")
- Neal (2011). *MCMC using ensembles of states for problems with fast and slow variables such as Gaussian process regression*, Preprint. (Stats), ("Ensemble MCMC", "Computational Short-cut")
- Calderhead (2014). *A general construction for parallelizing Metropolis-Hastings algorithms*, PNAS. (Stats), "Generalized Metropolis-Hastings", simulates "Finite-state Markov Chain" over proposals)
- Delmas and Jourdain (2016). *Does Waste Recycling Really Improve the Multi-Proposal Metropolis–Hastings algorithm? an Analysis Based on Control Variates*, J. Appl. Prob. (Probability), ("Multi-proposal Metropolis-Hastings")
- Yang et al. (2018). *On parallelizable Markov chain Monte Carlo algorithms with waste-recycling*, Stat. Comp. (Stats), ("Locally-weighted parallel MCMC" as "Rao-Blackwellization")
- Luo and Tjelmeland (2019). *A multiple-try Metropolis-Hastings algorithm with tailored proposals*, Comp. Stat. (Stats), ("Multiple-try" but not MTM)

A Complicated Landscape

- Ceperly et al. (1977). *Monte Carlo simulation of a many fermion study*, Phys. Rev. B. (Physics)
- Neal (2003). *Markov chain sampling for non-linear state space models using embedded hidden Markov models*, Preprint. (Stats)
- Frenkel (2004). *Speed-up of Monte Carlo simulations by sampling of rejected states*, PNAS. (Physics), ("Waste-Recycling", "Coarse-grained moves", "Macro-steps", "Superdetailed balance")
- Tjelmeland (2004). *Using all Metropolis-Hastings proposals to estimate mean values*, Preprint. (Stats), ("Weighted-mean", not "Rao-Blackwellized")
- Neal (2011). *MCMC using ensembles of states for problems with fast and slow variables such as Gaussian process regression*, Preprint. (Stats), ("Ensemble MCMC", "Computational Short-cut")
- Calderhead (2014). *A general construction for parallelizing Metropolis-Hastings algorithms*, PNAS. (Stats), "Generalized Metropolis-Hastings", simulates "Finite-state Markov Chain" over proposals)
- Delmas and Jourdain (2016). *Does Waste Recycling Really Improve the Multi-Proposal Metropolis–Hastings algorithm? an Analysis Based on Control Variates*, J. Appl. Prob. (Probability), ("Multi-proposal Metropolis-Hastings")
- Yang et al. (2018). *On parallelizable Markov chain Monte Carlo algorithms with waste-recycling*, Stat. Comp. (Stats), ("Locally-weighted parallel MCMC" as "Rao-Blackwellization")
- Luo and Tjelmeland (2019). *A multiple-try Metropolis-Hastings algorithm with tailored proposals*, Comp. Stat. (Stats), ("Multiple-try" but not MTM)
- Schwedes and Calderhead (2021). *Rao-Blackwellized parallel MCMC*, AISTATS. (ML), ("Parallel MCMC")

A Complicated Landscape

- Interesting gap between 1977 and 2003
- Literature is interdisciplinary
- Non-negligible preprint count
- A large amount of redundant, contradictory terminology
- Much of literature focuses on weighted averages (and calls this Rao-Blackwellization)
- "Parallelizable" is often conflated with "Parallelized"
- I have probably not included your work

Efficient Multiproposal Structures

Multiproposal MCMC

A parallel MCMC algorithm builds a transition kernel $P(\boldsymbol{\theta}_0, \mathrm{d}\boldsymbol{\theta})$ by:

1. generating P proposals $\boldsymbol{\Theta}_{-0} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_P)$ from a joint distribution $Q(\boldsymbol{\theta}_0, \mathrm{d}\boldsymbol{\Theta}_{-0}) =: q(\boldsymbol{\theta}_0, \boldsymbol{\Theta}_{-0})\mathrm{d}\boldsymbol{\Theta}_{-0}$; and

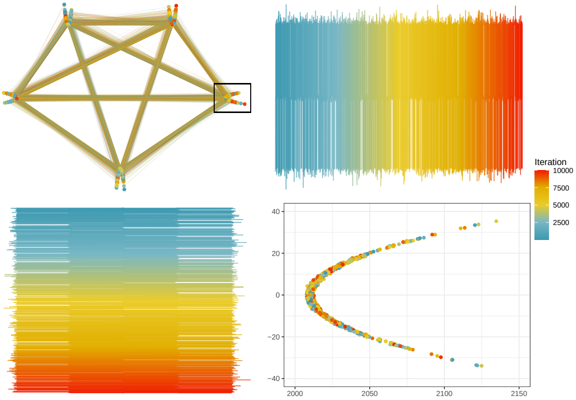
2. selecting the next state with probabilities

$$\pi_p = \frac{\pi(\boldsymbol{\theta}_p) q(\boldsymbol{\theta}_p, \boldsymbol{\Theta}_{-p})}{\sum_{p'=0}^{P} \pi(\boldsymbol{\theta}_{p'}) q(\boldsymbol{\theta}_{p'}, \boldsymbol{\Theta}_{-p'})}\,, \quad p \in \{0, 1, \ldots, P\}\,.$$

This kernel maintains detailed balance and leaves $\pi(\mathrm{d}\boldsymbol{\theta})$ invariant.

Multiproposal MCMC

PRO: using large numbers of proposals P helps overcome multimodality and non-linearity.



CON: requires $\mathcal{O}(P)$ target evaluations $\pi(\boldsymbol{\theta}_p)$ and proposal evaluations $q(\boldsymbol{\theta}_p, \boldsymbol{\Theta}_{-p})$, each of the latter being $\mathcal{O}(P)$.

Simplified Acceptance Probabilities

Can we somehow enforce $q(\boldsymbol{\theta}_p, \boldsymbol{\Theta}_{-p}) = q(\boldsymbol{\theta}_{p'}, \boldsymbol{\Theta}_{-p'})$, $\forall p, p' \in \{0, 1, \dots, P\}$, to obtain simplified acceptance probabilities

$$\pi_p = \frac{\pi(\boldsymbol{\theta}_p)}{\sum_{p'=0}^{P} \pi(\boldsymbol{\theta}_{p'})}\,, \quad p \in \{0, 1, \dots, P\}\,.$$

Such structured multiproposals would result in $\mathcal{O}(P^2)$ time savings and simpler implementation. I consider two such approaches in

- Holbrook (2023a). *Generating MCMC proposals by randomly rotating the regular simplex*, Journal of Multivariate Analysis.

The Simplicial Sampler (elegant and expensive)

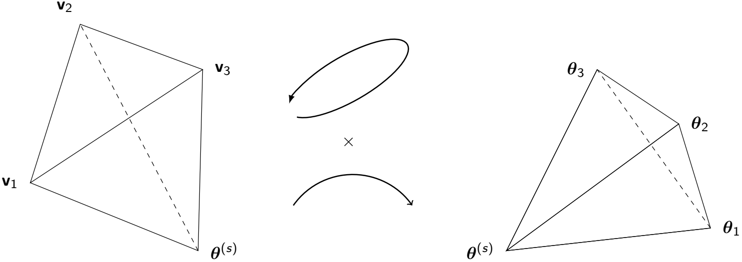
Let state space be $\mathbb{R}^D$ and $\mathbf{v}_1, \dots, \mathbf{v}_D \in \mathbb{R}^D$ satisfy

$$||\mathbf{v}_d - \mathbf{v}_{d'}||_2 = \lambda > 0\,, \quad d \neq d' \in \{1, \dots, D\}.$$

Then the simplicial sampler follows the following steps:

1. Sample $D \times D$ orthonormal matrix $\mathbf{Q}$ according to Haar distribution $\mathbf{Q} \sim \mathcal{H}(\mathcal{O}_D)$.
2. Rotate and translate the simplicial vertices $(\mathbf{0}, \mathbf{v}_1, \dots, \mathbf{v}_D) \longmapsto \mathbf{Q}(\mathbf{0}, \mathbf{v}_1, \dots, \mathbf{v}_D) + \boldsymbol{\theta}^{(s)} =: (\mathbf{0}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_D)$.
3. Draw a single sample $\boldsymbol{\theta}_d$ from $(\boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_D)$ with probability proportional to $\pi(\boldsymbol{\theta}_d)$.
4. Set $\boldsymbol{\theta}^{(s+1)} = \boldsymbol{\theta}_d$.

The Simplicial Sampler (elegant and expensive)



A simplicial sampling multiproposal for $D = 3$. Proposal set is obtained by rotating three simplex vertices about current state $\boldsymbol{\theta}^{(s)}$.

PRO: saves $\mathcal{O}(D^2)$ time for D evaluations $q(\boldsymbol{\theta}_d, \boldsymbol{\Theta}_{-d})$.
CON: $P = D$ and cost is $\mathcal{O}(D^3)$.

Tjelmeland Correction (a free lunch)

Tjelmeland (2004) suggests the two-step multiproposal

1. $\bar{\boldsymbol{\theta}} \sim N_D(\boldsymbol{\theta}^{(s)}, \boldsymbol{\Sigma})$;
2. $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_P \overset{iid}{\sim} N_D(\bar{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$.

Why? No satisfactory explanation. But it turns out that this structure leads to the desired equality (Holbrook 2023a):

$$q(\boldsymbol{\theta}_p, \boldsymbol{\Theta}_{-p}) = q(\boldsymbol{\theta}_{p'}, \boldsymbol{\Theta}_{-p'}), \forall p, p' \in \{0, 1, \dots, P\}.$$

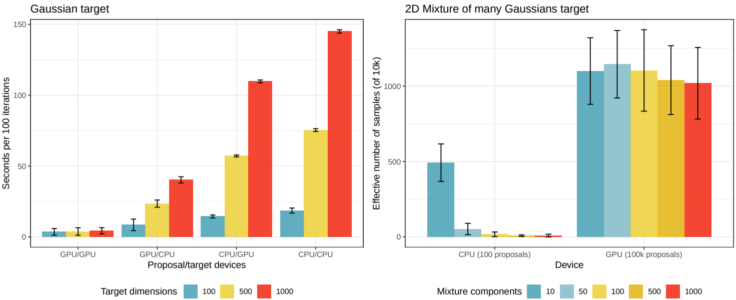
As promised, the resulting acceptance probabilities are:

$$\pi_p = \frac{\pi(\boldsymbol{\theta}_p)}{\sum_{p'=0}^{P} \pi(\boldsymbol{\theta}_{p'})}, \quad p \in \{0, 1, \dots, P\}.$$

Only the $\mathcal{O}(P)$ target evaluations remain in our way.

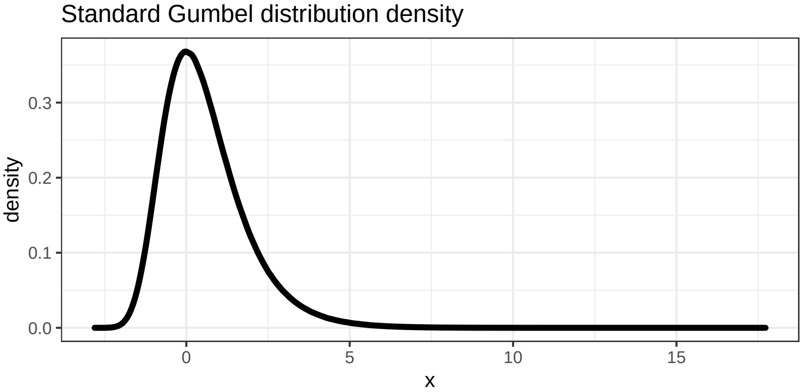
Parallelizing Parallel MCMC

Parallelizing Target Evaluations: CPU vs GPU



- Glatt-Holtz et al. (2022). *Parallel MCMC algorithms: theoretical foundations, algorithm design, case studies*, Preprint.

The Gumbel Distribution



If $z \sim \mathit{Gumbel}(0, 1)$, then it has density and distribution functions

$$g(z) = \exp\big(-z - \exp(-z)\big) \quad \text{and} \quad G(z) = \exp\big(-\exp(-z)\big)\,.$$

Gumbel-Max Trick

We wish to sample from the discrete distribution $\hat{p} \sim \mathit{Discrete}(\boldsymbol{\pi})$ for $\hat{p} \in \{0, 1, \ldots, P\}$ and we only know $\boldsymbol{\pi}^* = c\boldsymbol{\pi}$ for some $c > 0$.

Define $\boldsymbol{\lambda}^* = \log \boldsymbol{\pi}^* = \log \boldsymbol{\pi} + \log c$ and suppose $z_0, z_1, \ldots, z_P \overset{iid}{\sim} \mathit{Gumbel}(0, 1)$.

Finally, define $\alpha_p^* := \lambda_p^* + z_p$ and $\hat{p} = \arg\max_{p=0,\ldots,P} \alpha_p^*$.

Then the following holds (Papandreou and Yuille, 2011):

$$\Pr(\hat{p} = p) = \pi_p\,, \quad p = 0, 1, \ldots, P.$$

Data: Initial Markov chain state $\boldsymbol{\theta}^{(0)}$; total length of Markov chain S; total number of proposals per iteration P.
Result: A Markov chain $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(S)}$.

for $s \in \{1, \ldots, S\}$ **do**

$\quad \boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}^{(s-1)}$;

$\quad \bar{\boldsymbol{\theta}} \leftarrow \mathit{Normal}_D(\boldsymbol{\theta}_0, \boldsymbol{\Sigma})$;

$\quad z_0 \leftarrow \mathit{Gumbel}(0, 1)$;

$\quad$ **for** $p \in \{1, \ldots, P\}$ **do**

$\quad\quad \boldsymbol{\theta}_p \leftarrow \mathit{Normal}_D(\bar{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$;

$\quad\quad z_p \leftarrow \mathit{Gumbel}(0, 1)$;

$\quad$ **end**

$\quad \hat{p} \leftarrow \arg\min_{p=0,\ldots,P} \Big(f(p) := -\big(z_p + \log \pi(\boldsymbol{\theta}_p)\big) \Big)$;

$\quad \boldsymbol{\theta}^{(s)} \leftarrow \boldsymbol{\theta}_{\hat{p}}$;

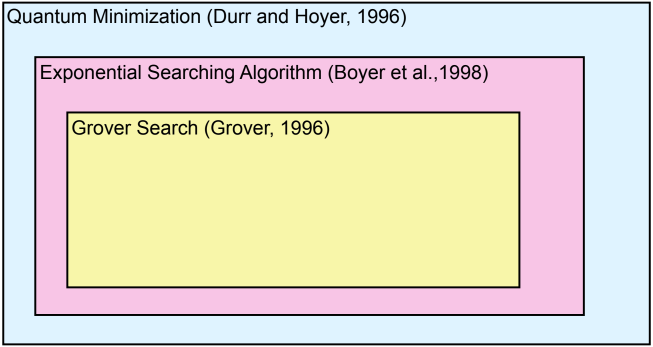
end

return $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(S)}$.

Quantum Parallel MCMC

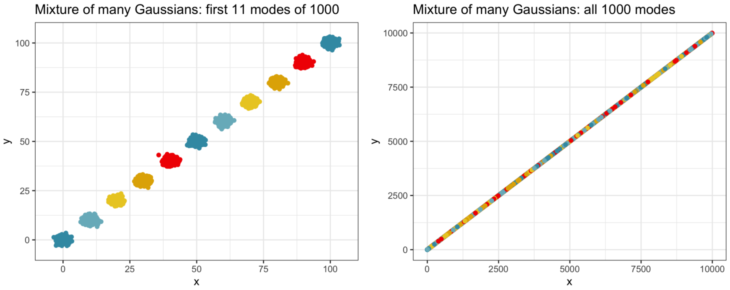
Use a quantum circuit to obtain

$$\hat{p} = \underset{p=0,\dots,P}{\arg\min} \Big(f(p) := -\big(z_p + \log \pi(\boldsymbol{\theta}_p)\big) \Big)$$



- Holbrook (2023b). *A quantum parallel Markov chain Monte Carlo*, JCGS.

QPMCMC: Racing to an ESS of 100



Proposals	MCMC iterations	Target evaluations	Speedup	Efficiency gain
1,000	249,398 (200,998, 311,998)	24,988,963 (20,149,132, 31,265,011)	9.98 (9.98, 9.98)	1
5,000	14,398 (12,998, 16,998)	3,314,560 (2,989,418, 3,916,281)	21.72 (21.70, 21.74)	7.58 (6.25, 9.71)
10,000	5,998 (4,998, 6,998)	1,993,484 (1,662,592, 2,330,842)	30 (29.96, 30.26)	12.87 (8.64, 18.80)

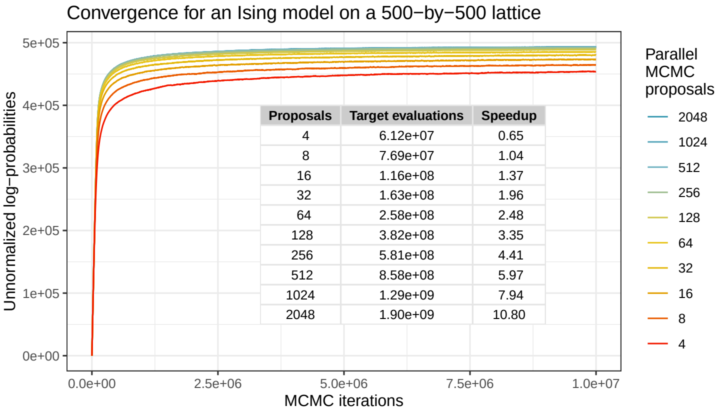
Ising Model Target

Consider the Ising-type lattice model over configurations $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_D)$ consisting of D individual spins $\theta_d \in \{-1, 1\}$

$$\pi(\boldsymbol{\theta}|\rho) \propto \exp\left(\rho \sum_{(d,d')\in\mathcal{E}} \theta_d \theta_{d'}\right) .$$

No need for Tjelmeland corrections when we use uniform proposals on $\{-1, 1\}^D$. The following results are based on single-flip proposals (although not necessary).

Ising Model Target



Proposals	Target evaluations	Speedup
4	6.12e+07	0.65
8	7.69e+07	1.04
16	1.16e+08	1.37
32	1.63e+08	1.96
64	2.58e+08	2.48
128	3.82e+08	3.35
256	5.81e+08	4.41
512	8.58e+08	5.97
1024	1.29e+09	7.94
2048	1.90e+09	10.80

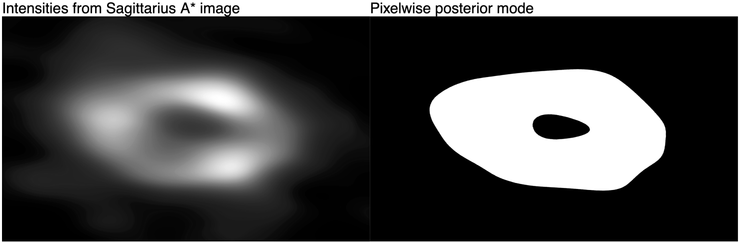
Bayesian Image Segmentation

Following Hurn (1997), y_d are intensity values associated with individual pixels.

$$
\begin{aligned}
y_d|(\mu_\ell, \sigma^2, \theta_d) &\overset{ind}{\sim} \text{Normal}(\mu_\ell, \sigma^2)\,, \quad y_d \in [0, 255]\,,\\
\theta_d &= \ell\,, \quad d \in \{1, \ldots, D\}\,,\\
\mu_\ell &\overset{iid}{\sim} \text{Uniform}(0, 255)\,, \quad \ell \in \{-1, 1\}\,,\\
\frac{1}{\sigma^2} &\sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)\\
\boldsymbol{\theta} &\sim \text{Ising}(\rho)\,, \quad \rho = 1.2\,.
\end{aligned}
$$

Bayesian Image Segmentation

Segmenting a 4,076-by-4,076 intensity map. Using 1,024 proposals, QPMCMC requires less than 10% the evaluations required by a conventional computer.



Future Directions

Theoretical Challenges

Glatt-Holtz et al. (2022) develop foundations for multiproposal MCMC, incorporating:

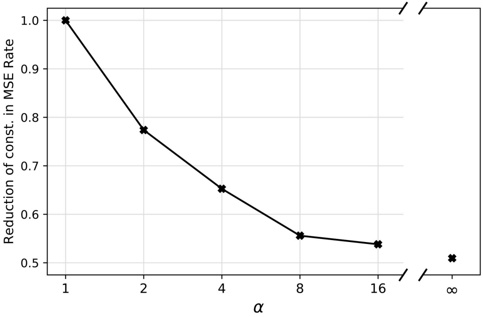
- general state space representation (Tierney, 1998);
- involutions on extended phase spaces (Nekludov et al., 2020; Glatt-Holtz et al., 2020; Andreiu et al., 2020);
- proposal cloud resampling;
- Metropolis-Hastings and Barker/Boltzmann acceptances.

We still lack:

- Optimal tuning guidances (D,P);
- Error bounds for biased kernels;
- nonreversible multiproposal MCMC.

Bizarre Benefits of Bias

Schwedes and Calderhead (2021) estimate the relative reduction in MSE for Monte Carlo estimators as a function of α, where $\alpha \times P$ is the number of proposal cloud resampling iterations.

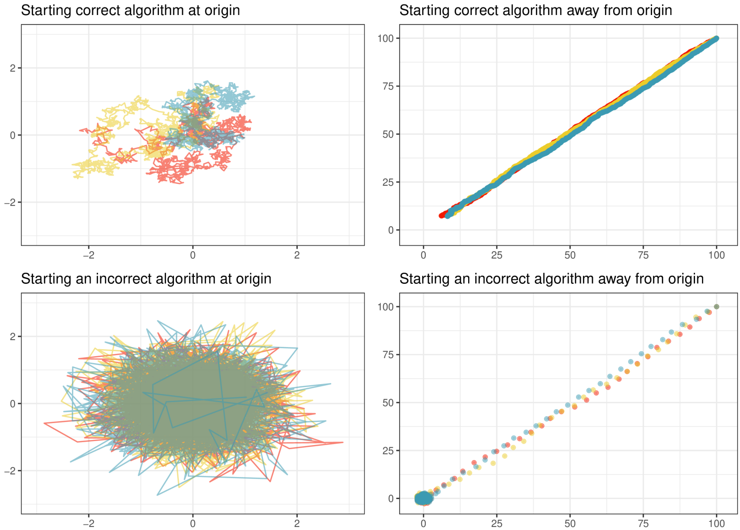


This is using the naive multiproposal

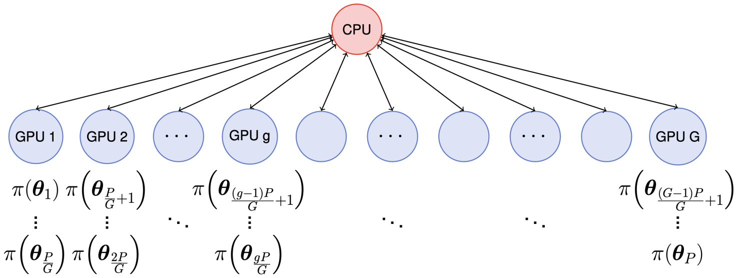
$$\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_P \overset{iid}{\sim} N_D(0, \boldsymbol{\Sigma}) \, .$$

with full acceptance probabilities $\pi_p \propto \pi(\boldsymbol{\theta}_p) \prod_{p' \neq p} q(\boldsymbol{\theta}_p, \boldsymbol{\theta}_{p'})$.

Bizarre Benefits of Bias



Tjelmeland Correction Reduces Communication



If models are multimodal and parallelizable:

- Bayesian inversion of nonlinear PDEs;
- Hawkes processes (temporal, spatiotemporal, multivariate).

Acknowledgments

Much of my work on parallel MCMC is joint with:

- Xiang Ji (Tulane)
- Justin Krometis (Virginia Tech)
- Nathan Glatt-Holtz (Tulane)
- Cecilia Mondaini (Drexel)
- Marc Suchard (UCLA)

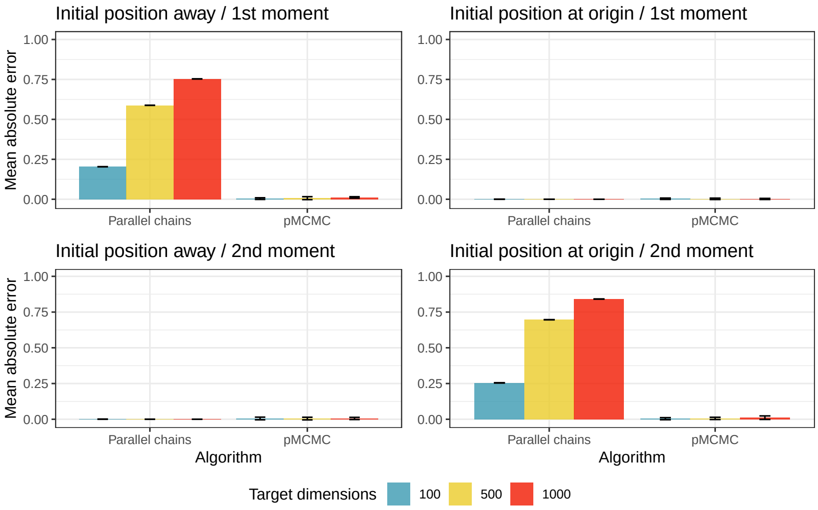
Acknowledgments

This work is supported by

- NSF grants DMS 2236854 and DMS 2152774;
- NIH grant K25 AI153816;
- and a gift from the Karen Toffler Charitable Trust.

Many Proposals vs Many Chains

Multivariate Gaussian Targets



Massively Multimodal Targets