A Massively Parallel Evolutionary Markov Chain Monte Carlo Algorithm for Sampling Complicated Multimodal State Spaces

Extended Abstract

Wendy K. Tam Cho
Departments of Political Science, Statistics, Math, Law, Asian American Studies, and National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
wendycho@illinois.edu

Yan Y. Liu
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
yanliu@illinois.edu

ABSTRACT

We develop an Evolutionary Markov Chain Monte Carlo (EMCMC) algorithm for sampling from large multi-modal state spaces. Our algorithm combines the advantages of evolutionary algorithms (EAs) as optimization heuristics and the theoretical convergence properties of Markov Chain Monte Carlo (MCMC) algorithms for sampling from unknown distributions. We harness massive computational power with a parallel EA framework that guides a large set of Markov chains. Our algorithm has applications in many different fields of science. We demonstrate its effectiveness with an application to political redistricting.

KEYWORDS

Optimization, Evolutionary Algorithms, Markov Chain Monte Carlo

ACM Reference Format:

Wendy K. Tam Cho and Yan Y. Liu. 2018. A Massively Parallel Evolutionary Markov Chain Monte Carlo Algorithm for Sampling Complicated Multimodal State Spaces: Extended Abstract. In *Proceedings of (SC18)*. ACM, New York, NY, USA, 3 pages.

1 INTRODUCTION

MCMC originated in statistical physics [7] and has migrated to applications in many disciplines for sampling from unknown distributions. Theoretically, MCMC ensures sampling from unknown distributions, but this theoretical result is asymptotic. In large applications, the time required for convergence can be prohibitive.

Markov transitions are commonly devised as small random changes in the current state for two reasons. First, small changes are conceptually and operationally simple—easy to define and relatively simple to implement. Second, they likely result in a large Metropolis-Hastings (MH) ratio which leads to a fluid Markov chain. However, because these are small moves in a very large state space, the chain converges slowly and is likely to become trapped in localized regions. Performance is compromised for conceptual ease. Large moves may provide more efficient and effective state space traversal, leading to faster convergence of the chain. However, how to devise effective large movements is not straightforward. Simply "large" movements often result in small MH ratios, which lead to rejected proposals, and a non-fluid and ineffective Markov chain.

EA and MCMC algorithms have the same goal—to effectively and efficiently traverse the state space. Insights from the optimization heuristics literature can improve the performance of MCMC algorithms. The main task in marrying these two literatures is to fit the mechanics of the optimization search within the theoretical MCMC framework that enables sampling.

2 GUIDING MARKOV CHAINS WITH EAs

We focus on EAs that evolve an initial population through mutation and crossover to produce subsequent populations [2]. EAs enable a diversified search, which improves the mixing time of a Markov chain and lend themselves easily to parallelization, which enables greater computing power for large applications.

EAs and MCMC are not simple to harmonize. The moves that define a Markov chain need to satisfy particular criteria; and, incorporating the mechanics of EA operators introduces constraints on the movement of the chain, which may compromise convergence to the stationary distribution.

Adaptation of EAs to an MCMC framework have been proposed for Bayesian mixture models, C_p model sampling, and change point problems [3–5]. We build on this literature to propose a new EMCMC algorithm that adapts an EA for a very large spatially encoded application to an MCMC framework.

2.1 Adapting EAs and MCMC

We wish to sample from a distribution

$$f(x) = \frac{\exp\{-H(x)\}}{Z} \tag{1}$$

where $x = (\beta_1, \beta_2, \ldots, \beta_k)$ is a k-dimensional vector, and $Z = \sum_{i=1}^{N} \exp\{-H(x_i)\}$ is a normalizing constant. In an EA, $H(x)$ is the fitness function, x is a chromosome, and the β_i are the alleles. For each chromosome, we can assign a weight that is proportional to its Boltzmann probability.

In a **real-encoded EA**, $\beta_i \in \mathbb{R}$. Adapting a k-point or uniform mutation operator to MCMC can be accomplished in a fashion similar to the binary-encoded case [5]. A new population, $\mathbf{y} = \{x_1, \ldots, y_k, \ldots, x_n\}$, is proposed where x_k is replaced in the population with y_k. The proposal is accepted with probability, $\min(1, r_m)$, where

$$r_m = \frac{f(\mathbf{y})\,T(\mathbf{x}\,|\,\mathbf{y})}{f(\mathbf{x})\,T(\mathbf{y}\,|\,\mathbf{x})} = \exp\{-[H(y_k) - H(x_k)]/t_k\}\frac{T(\mathbf{x}\,|\,\mathbf{y})}{T(\mathbf{y}\,|\,\mathbf{x})} \tag{2}$$

is the MH ratio, and $T(\cdot\,|\,\cdot)$ is the transition probability between two populations. If the proposal is accepted, then $\mathbf{y}$ becomes the population. Otherwise, the population is unchanged.

SC18, November 2018, Dallas, TX USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00

For a symmetric mutation operator, $T(\mathbf{x} \mid \mathbf{y}) = T(\mathbf{y} \mid \mathbf{x})$, allowing the MH ratio to be computed simply from the fitness values of the chromosomes. While mutation operators are often easily adapted in this way, crossover operators are not as malleable, and must be carefully devised to preserve constraints.

2.2 Adaptive Direction Sampling (ADS)

For EA and MCMC, performance is predicated on random *and* purposeful moves. A Gibbs sampler is a sequence of conditional distributions along a set of directions that follows the local dynamics of the target distribution. However, it is difficult to construct univariate sampling directions that ensure rapid movement around the support of the distribution.

ADS generalizes the Gibbs sampler [8] so that

$$x_c^{(t+1)} = x_c^{(t)} + r(v^{(t)} + u^{(t)} x_c^{(t)}), \tag{3}$$

where $v^{(t)}$, an n-vector, and $u^{(t)}$, a scalar, are any functions of the current set, $S(t)$, excluding $x_c^{(t)}$. The snooker algorithm arises when $u^{(t)} = -1$ and $v^{(t)} = x_a^{(t)}$, where $x_a^{(t)}$ is a randomly chosen point from $S(t)$, excluding $x_c^{(t)}$. If $v^{(t)}$ is the difference between two points in the current set, parallel ADS emerges. If $v^{(t)}$ is a random coordinate direction with $u^{(t)} = 0$, we have the Gibbs sampler.

MCMC has been proposed for sampling the space of redistricting maps where n geographic units, $u_1, u_2, \ldots, u_n$ are partitioned into k districts, $d_1, d_2, \ldots, d_k$, that satisfy legal constraints. The problem is challenging for MCMC because the state space is both multimodal and very large. We demonstrate with a data set (with 141,197,991,025 possible solutions) that is small enough so that we can enumerate all possible solutions, but large enough to be non-trivial.

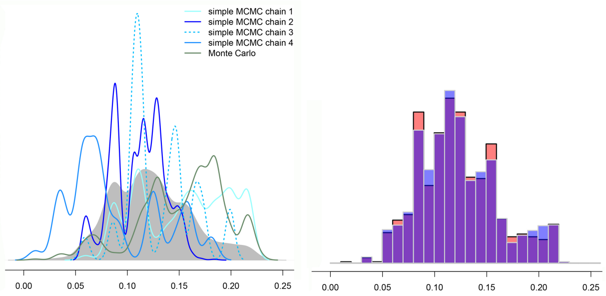

Figure 1: EMCMC uniformly samples, but simple MCMC and Monte Carlo algorithms do not uniformly sample in these large multimodal state spaces.

None of the blue lines provide a close outline of the gray area (Fig. 1), illustrating the difficulty of a simple MCMC implementation in traversing a multimodal state space. We propose a more intelligent EA-guided spatial neighborhood search [6]. Our mutation operator alters a small number of "mutable" alleles (where change does not result in an infeasible solution) from a randomly chosen chromosome. Our crossover, a variant of ADS, is a spatial path relinking operator that performs recombination while respecting spatial constraints in the neighborhood space [1]:

(1) Randomly choose two chromosomes, a source solution, x_s, and a target solution, x_t, $s \neq t$, from the population $\mathbf{x}$.

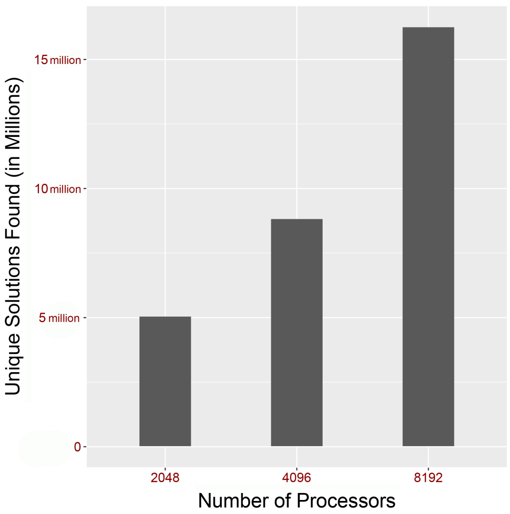

Figure 2: Scalability of Parallel EMCMC

(2) The relinking process is comprised of a "walk" between the two solutions. Each step in the path converts a random allele from its value in x_s to its value in x_t.
(3) If a feasible solution, y_k is found on the path, replace chromosome x_s with y_k to produce a new population, $\mathbf{y}$.

Our EMCMC implementation uniformly samples the state space (right plot in Fig. 1). In our small data set with known answers, there is substantial overlap between the true distribution (shown in red). and EMCMC sampled solutions (shown in blue).

In actual redistricting applications which are much larger, massive computing power is essential. Our directed crossover operator guides intelligent and efficient space traversal while parallel chains aid in reaching the different modes and harnessing more computing power. As we enlist more processors, our EMCMC, which uniformly samples, reaches more of the underlying space. Figure 2 shows the number of unique solutions found in 10 minutes of computing time on the Blue Waters supercomputer. With 8192 processors, we sample among almost 17 million unique solutions.

3 DISCUSSION

We have devised a massively parallel Evolutionary Markov Chain Monte Carlo (EMCMC) algorithm for astronomically large, idiosyncratic, and multimodal state spaces. We have demonstrated the applicability of our algorithm to the substantive problem of redistricting. The methods are general in nature and have applications to many fields of science.

ACKNOWLEDGMENTS. Our computational experiments used the Blue Waters sustained-petascale computing resources, which is supported by the NSF (Grants OCI–0725070 and ACI–1238993) and the state of Illinois under the projects, "Massively Parallel Evolutionary Markov Chain Monte Carlo for Sampling Complicated Multimodal State Spaces." Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications. Research support was provided by the National Science Foundation, Grant SES-1725418/1728902, "Collaborative Research: High-Performance Computational Standards for Redistricting."

REFERENCES

[1] Fred Glover, Manuel Laguna, and Rafael Marti. 2000. Fundamentals of Scatter Search and Path Relinking. *Control and Cybernetics* 29, 3 (2000), 653–684.

[2] David Goldberg. 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley Professional, Reading, MA.

[3] Kathryn Blackmond Laskey and James M. Myers. 2003. Population Markov Chain Monte Carlo. *Machine Learning* 50, 1–2 (2003), 175–196.

[4] Faming Liang and Wing Hung Wong. 2000. Evolutionary Monte Carlo Sampling: Applications to C_p Model Sampling and Change-point Problems. *Statistica Sinica* 10 (2000), 317–342.

[5] Faming Liang and Wing Hung Wong. 2001. Real-Parameter Evolutionary Monte Carlo With Applications to Bayesian Mixture Models. *Journal of the American Statistical Society* 96, 454 (2001), 653–666.

[6] Yan Y. Liu and Wendy K. Tam Cho. 2018. Spatially Explicit Evolutionary Computation for Large-scale Spatial Optimization. (2018). Technical Report.

[7] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. 1953. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* 21, 6 (1953), 1087–1092.

[8] G.O. Roberts and W.R. Gilks. 1994. Convergence of Adaptive Direction Sampling. *Journal of Multivariate Analysis* 49, 2 (1994), 287–298.