Sequential Core-Set Monte Carlo

Boyan Beronov[1], Christian Weilbach[1], Frank Wood[1], Trevor Campbell[2]

37th Conference on Uncertainty in Artificial Intelligence

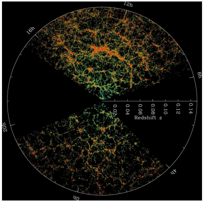
Online, July 2021

[1]Department of Computer Science, [2]Department of Statistics
University of British Columbia, Vancouver, Canada




Motivation—Scalable incremental Bayesian inference

(*Simultaneous localization & mapping*)



(*Sloan Digital Sky Survey*)

Necessity of streaming memory

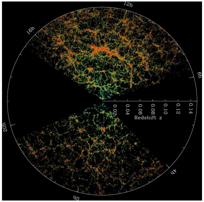
- **online** inference, **continual** learning
- **large**-**scale** data (limited memory)
- analysis by **synthesis** (dictionary)
- model interpretation by **examples**

Problems of previous approaches

- repeated access to **all data**
- **heuristic** approximations
- **model-specific** methods
- **opaque** representations

Motivation—Scalable incremental Bayesian inference

(*Simultaneous localization & mapping*)



(*Sloan Digital Sky Survey*)

Necessity of streaming memory

- **online** inference, **continual** learning
- **large**-scale data (limited memory)
- analysis by **synthesis** (dictionary)
- model interpretation by **examples**

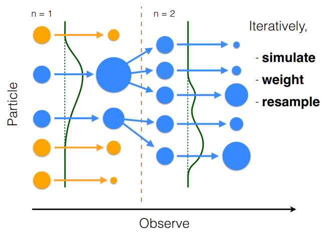
Problems of previous approaches

- repeated access to **all data**
- **heuristic** approximations
- **model-specific** methods
- **opaque** representations

Our Contribution

__Online__ variant of sequential Monte Carlo for __generic__ exchangeable Bayesian models with __optimization__-based __streaming__ memory compression

Sequential Monte Carlo (SMC) with rejuvenation (MCMC)



(*SMC*)

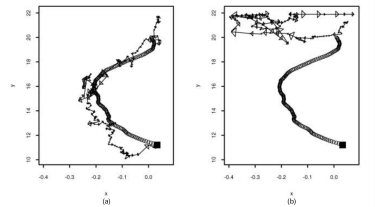


Fig. 2. Comparing the performances of (a) the resample–move method and (b) a standard SIR filter: ○, the ship's true trajectory; ■, final position; ———, the filter's output

(Gilks and Berzuini, 2001)

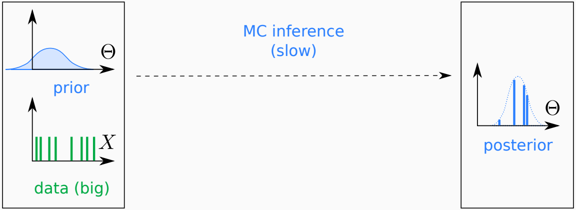
Sequential importance resampling

- particle population
- recursive updates via *reweighting* & *resampling*

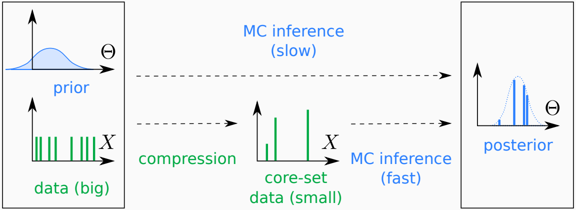
Variance reduction

- interleaved MCMC moves (*rejuvenation*/*resample-move*/*mutation*)
- crucial for approximation quality
- **computation** & **memory** $\propto$ **observations** !

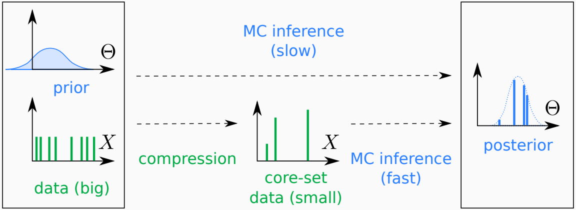
Bayesian core-sets—Faster approximate inference



Bayesian core-sets—Faster approximate inference



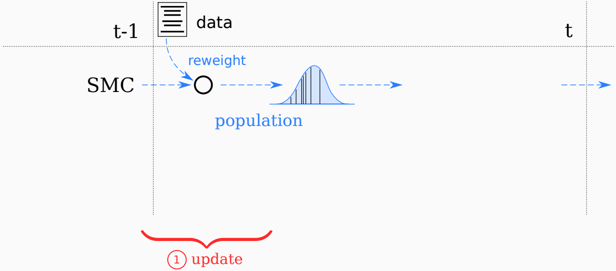
Bayesian core-sets—Faster approximate inference



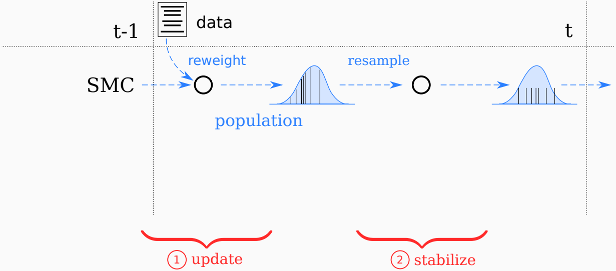
Complementary belief representations

- Θ: MC particles
- X: Bayesian core-sets (Campbell and Broderick, 2017)

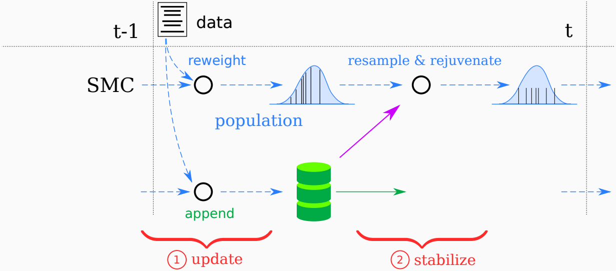
Dataflow in sequential core-set Monte Carlo (SCMC)



Dataflow in sequential core-set Monte Carlo (SCMC)



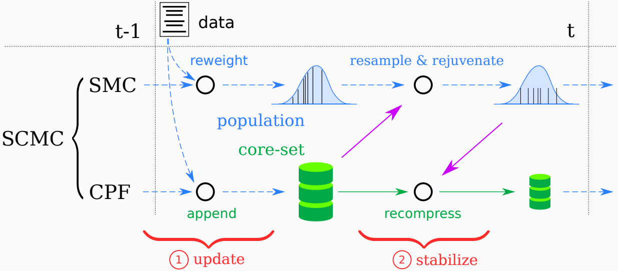
Dataflow in sequential core-set Monte Carlo (SCMC)



Filtering step in SCMC

- *rejuvenation* in sequential Monte Carlo (SMC) leverages core-set memory

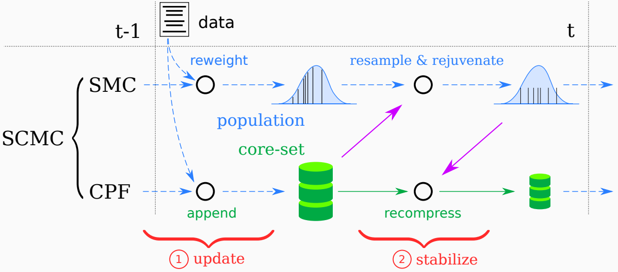
Dataflow in sequential core-set Monte Carlo (SCMC)



Filtering step in SCMC

- *rejuvenation* in sequential Monte Carlo (SMC) leverages core-set memory
- *recompression* in core-set projection filter (CPF) leverages MC belief

Dataflow in sequential core-set Monte Carlo (SCMC)



Filtering step in SCMC

- *rejuvenation* in sequential Monte Carlo (SMC) leverages core-set memory
- *recompression* in core-set projection filter (CPF) leverages MC belief

Intuition

- **SMC** $\approx$ '*incrementalized* **MCMC**'
- **Core-set** compression $\approx$ sparse **VI** (Campbell and Beronov, 2019)
- **SCMC** $\approx$ 'sparse **SMC**' $\approx$ '*incrementalized* & sparse **MCMC** & **VI**'

Approximate rejuvenation in SCMC

Rejuvenation target

prior	$p_0(\theta)$		
likelihood	$p(x \mid \theta)$		
exact MCMC target	$p(\theta \mid x_1, \ldots, x_N)$	$\propto$	$p_0(\theta) \cdot \prod_{j=1}^{N} p(x_j \mid \theta)$
approximate MCMC target		$\approx$	$p_0(\theta) \cdot \prod_{j=1}^{C} p(\bar{x}_j \mid \theta)^{\bar{w}_j}$

Approximate rejuvenation in SCMC

Rejuvenation target

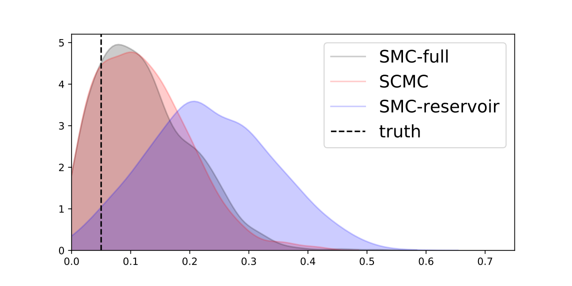
prior	$p_0(\theta)$		
likelihood	$p(x \mid \theta)$		
exact MCMC target	$p(\theta \mid x_1, \ldots, x_N)$	$\propto$	$p_0(\theta) \cdot \prod_{j=1}^{N} p(x_j \mid \theta)$
approximate MCMC target		$\approx$	$p_0(\theta) \cdot \prod_{j=1}^{C} p(\bar{x}_j \mid \theta)^{\bar{w}_j}$

Example MCMC method—Metropolis-Hastings
acceptance ratio $\alpha(\cdot, \theta)$ for proposal $g(\cdot \mid \theta)$

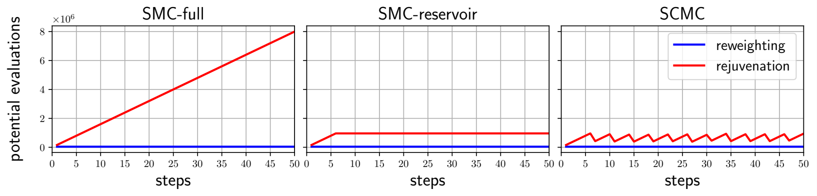
$$\alpha(\hat{\theta}, \theta) := \min\left(1, \frac{g(\theta \mid \hat{\theta})}{g(\hat{\theta} \mid \theta)} \cdot \frac{p(\hat{\theta} \mid x_1, \ldots, x_N)}{p(\theta \mid x_1, \ldots, x_N)}\right) \approx \min\left(1, \frac{g(\theta \mid \hat{\theta})}{g(\hat{\theta} \mid \theta)} \cdot \frac{p_0(\hat{\theta})}{p_0(\theta)} \cdot \prod_{j=1}^{C} \frac{p(\bar{x}_j \mid \hat{\theta})^{\bar{w}_j}}{p(\bar{x}_j \mid \theta)^{\bar{w}_j}}\right)$$

Posterior approximation quality

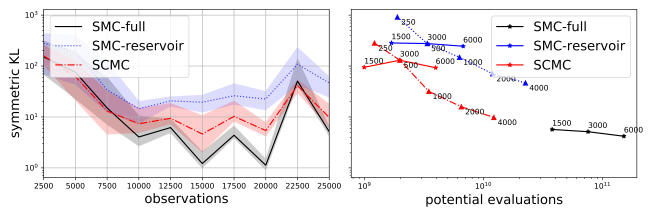
Qualitative improvement: AR(1) system identification (1 parameter)



Computation cost reduction: Normal-inverse-Wishart (27 parameters)



Error & variance reduction: Bayesian logistic regression (11 parameters)



Related methods—Online Bayesian inference

Related methods—Online Bayesian inference

(Extended/Unscented/Ensemble) Kalman filters

(locally) Gaussian models

Sparse online Gaussian processes

(Csató and Opper, 2002)

Particle filters / Sequential Monte Carlo

non-adaptive/offline approximation of rejuvenation, or growing cost

(Kitagawa, 1998; Chopin, 2002; Del Moral, Doucet, and Jasra, 2006; Doucet and Johansen, 2009; Naesseth, Lindsten, and Schön, 2019; Gunawan et al., 2020)

Variational filters

non-adaptive, asymptotically inconsistent variational family

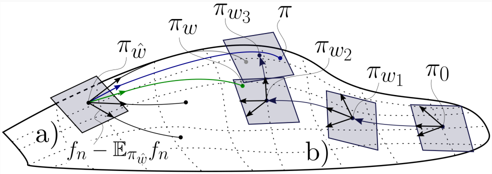
(Friston, 2008; Broderick et al., 2013; Nguyen et al., 2017; Marino, Cvitkovic, and Yue, 2018)

Information geometric projection filters

non-adaptive variational family

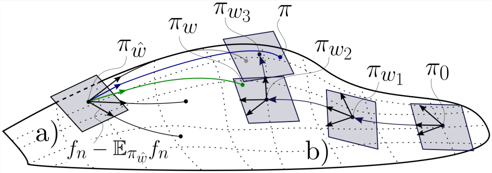
(Kulhavý, 1990; Kulhavy, 1992; Kulhavý, 1996; Brigo, Hanzon, and Le Gland, 1995; Brigo, Hanzon, and Le Gland, 1999; Armstrong and Brigo, 2016)

Bayesian statistics ↔ Information geometry



Bayesian statistics	↔	Information geometry
Hilbert space	↔	*exponential* tangent space
of centered *log*-likelihood functions		of tangent vectors
at weighting distribution		at point of statistical manifold
log density ratio $\log(p/q)$	↔	*logarithmic* map $\mathrm{Log}_q(p)$
tuple of observations	↔	coordinate frame
Fisher information metric	↔	Riemannian metric

Bayesian statistics $\leftrightarrow$ Information geometry



Bayesian statistics	$\leftrightarrow$	Information geometry
Hilbert space	$\leftrightarrow$	*exponential* tangent space
of centered *log*-likelihood functions		of tangent vectors
at weighting distribution		at point of statistical manifold
log density ratio $\log(p/q)$	$\leftrightarrow$	*logarithmic* map $\mathrm{Log}_q(p)$
tuple of observations	$\leftrightarrow$	coordinate frame
Fisher information metric	$\leftrightarrow$	Riemannian metric

SparseVI (Campbell and Beronov, 2019): approximate 'future' **Log**$_{\text{prior}}$ (posterior)

SCMC: approximate 'past' **Log**$_{\text{belief}}$ (prior)

Conclusion—Sequential Core-Set Monte Carlo

Contributions

- **bounded-resource** *data tempering* SMC via core-set rejuvenation
- generic & statistically sound **streaming** construction of Bayesian core-sets
- **generalization** of *projection filtering* to core-set posterior exponential family

Future work

- convergence theory, adaptive choice of core-set & population size
- distributed streaming inference via core-set messages
- generalization to non-exchangeable/non-stationary models
- applications: 'Bayes optimal subsampling' for variance reduction & active learning in SL, episodic memory in RL, evolutionary optimization ...

References

References i

Armstrong, John and Damiano Brigo (Mar. 2016). "Nonlinear Filtering via Stochastic PDE Projection on Mixture Manifolds in $$L2̂$$ L 2 Direct Metric". en. In: *Mathematics of Control, Signals, and Systems* 28.1. ISSN: 0932-4194, 1435-568X. DOI: 10.1007/s00498-015-0154-1.

Brigo, Damiano, Bernard Hanzon, and François Le Gland (1995). "A Differential Geometric Approach to Nonlinear Filtering: The Projection Filter". In: *Proceedings of the 34th Conference on Decision and Control, New Orleans 1995*. Vol. 4, pp. 4006–4011.

— (1999). "Approximate Nonlinear Filtering by Projection on Exponential Manifolds of Densities". In: *Bernoulli* 5.3, pp. 495–534.

Broderick, Tamara et al. (2013). "Streaming Variational Bayes". In: *Advances in Neural Information Processing Systems 26*. Ed. by C. J. C. Burges et al. Curran Associates, Inc., pp. 1727–1735.

Campbell, Trevor and Boyan Beronov (2019). "Sparse Variational Inference: Bayesian Coresets from Scratch". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. Curran Associates, Inc., pp. 11461–11472.

Campbell, Trevor and Tamara Broderick (Oct. 2017). "Automated Scalable Bayesian Inference via Hilbert Coresets". en. In: *arXiv:1710.05053 [cs, stat]*. arXiv: 1710.05053 [cs, stat].

Chopin, N. (Aug. 2002). "A Sequential Particle Filter Method for Static Models". en. In: *Biometrika* 89.3, pp. 539–552. ISSN: 0006-3444, 1464-3510. DOI: 10.1093/biomet/89.3.539.

Csató, Lehel and Manfred Opper (2002). "Sparse On-Line Gaussian Processes". In: *Neural computation* 14.3, pp. 641–668.

References ii

Del Moral, Pierre, Arnaud Doucet, and Ajay Jasra (June 2006). "Sequential Monte Carlo Samplers". en. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.3, pp. 411–436. ISSN: 1369-7412, 1467-9868. DOI: 10.1111/j.1467-9868.2006.00553.x.

Doucet, Arnaud and Adam M. Johansen (2009). "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later". In: *Handbook of nonlinear filtering* 12.656-704, p. 3.

Friston, K.J. (July 2008). "Variational Filtering". en. In: *NeuroImage* 41.3, pp. 747–766. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2008.03.017.

Gilks, Walter R. and Carlo Berzuini (2001). "Following a Moving Target—Monte Carlo Inference for Dynamic Bayesian Models". en. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.1, pp. 127–146. ISSN: 1467-9868. DOI: 10.1111/1467-9868.00280.

Gunawan, David et al. (Mar. 2020). "Subsampling Sequential Monte Carlo for Static Bayesian Models". In: *arXiv:1805.03317 [stat]*. arXiv: 1805.03317 [stat].

Kitagawa, Genshiro (1998). "A Self-Organizing State-Space Model". In: *Journal of the American Statistical Association* 93.443, pp. 1203–1215. ISSN: 0162-1459. DOI: 10.2307/2669862.

Kulhavý, Rudolf (May 1990). "Recursive Nonlinear Estimation: A Geometric Approach". en. In: *Automatica* 26.3, pp. 545–555. ISSN: 0005-1098. DOI: 10.1016/0005-1098(90)90025-D.

Kulhavy, Rudolf (Mar. 1992). "Recursive Nonlinear Estimation: Geometry of a Space of Posterior Densities". en. In: *Automatica* 28.2, pp. 313–323. ISSN: 0005-1098. DOI: 10.1016/0005-1098(92)90118-Y.

Kulhavý, Rudolf (1996). *Recursive Nonlinear Estimation: A Geometric Approach*. Vol. 216. Springer.

References iii

Marino, Joseph, Milan Cvitkovic, and Yisong Yue (2018). "A General Method for Amortizing Variational Filtering". In: *Advances in Neural Information Processing Systems*, pp. 7857–7868.

Naesseth, Christian A., Fredrik Lindsten, and Thomas B. Schön (2019). "Elements of Sequential Monte Carlo". en. In: *Foundations and Trends® in Machine Learning* 12.3, pp. 187–306. ISSN: 1935-8237, 1935-8245. DOI: 10.1561/2200000074.

Nguyen, Cuong V. et al. (Oct. 2017). "Variational Continual Learning". en. In: *arXiv:1710.10628 [cs, stat]*. arXiv: 1710.10628 [cs, stat].

***Simultaneous localization & mapping*.** http://www.iri.upc.edu/research/webprojects/pau/img/silde_slam_ilaTRO2010_intel.jpg.

***Sloan Digital Sky Survey*.** https://dev.sdss.org/wp-content/uploads/2014/06/orangepie.jpg.

***SMC*.** https://www.robots.ox.ac.uk/~fwood/talks/2015/agi-keynote-2015-probabilistic-programming.pdf.