Frugal approaches to Sequential Monte Carlo (SMC) simulation of Baysesian posterior distributions

Mathias Rousset [1,2]

Joint work with Patrick Héas[1,2] and Fred Cérou[1,2].

[1]SimSmart project-team, Inria Rennes Bretagne Atlantique

[2]IRMAR, Université de Rennes 1

2024

1) The problem of sampling target distributions. Real world examples.

What is a 'target distribution' ?

- We will call a 'target distribution' a probability distribution of the form:

$$\eta(d\theta) := e^{-V(\theta)}\pi(d\theta)/Z.$$

  given explicit up to a normalization constant Z (standard Monte Carlo jargon/terminology).
- $\pi(d\theta)$ a reference probability distribution that can be cheaply and exactly Monte Carlo simulated. Example: normal i.i.d. .
- $V(\theta) \in]-\infty, +\infty]$ a given computable function given as a black-box. This means that one is given a numerical routine evaluating $\theta \mapsto V(\theta)$ and perhaps $\theta \mapsto \nabla_\theta V$ (or higher).

What is a 'target distribution' ?

$$\eta(d\theta) := e^{-V(\theta)}\pi(d\theta)/Z.$$

with $Z := \int e^{-V(\theta)}\pi(d\theta)$. E.g. π i.i.d. normal seq. .

Problem (Sampling target)

- Numerically estimate the *normalization* Z.
- Monte Carlo simulate a sample $(\Theta_1, \ldots, \Theta_N)$ with

$$\frac{1}{N}\sum_{n=1}^{N}\delta_{\Theta_n}(d\theta) \simeq \eta(d\theta)$$

.

Cartoon (to have in mind)

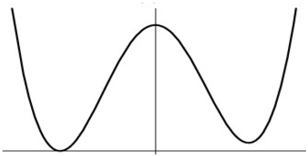


Figure: Double-well potential function $\theta \mapsto V(\theta)$

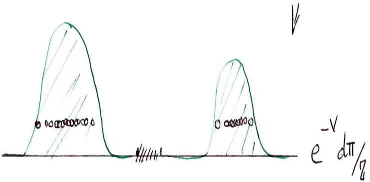


Figure: Target distribution $\eta \propto e^{-V}\pi$ (π uniform) with MC sample

Difficulties

- Cost: $V(\theta)$ evaluation is computationally intensive.
- High dimension: $\theta \in \mathbb{R}^d$ with $d \gg 1$.
- Peaky multimodality of η . Similar to non-convex optimization:

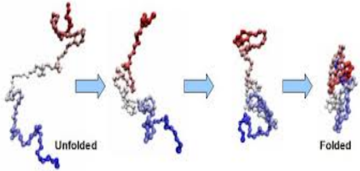
Lemma

If V has unique global π-essential minimum θ_*:

$$\lim_{\beta \to +\infty} e^{-\beta V(\theta)} \pi(d\theta)/Z_\beta = \delta_{\theta_*}(d\theta)$$

Pb: where is θ_* ?

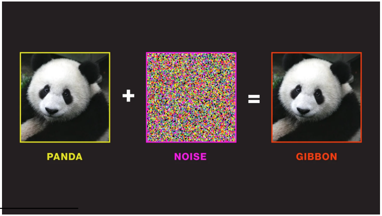
Example 1: low temperature equilibrium mechanics



- $\theta \in \mathbb{R}^{3M}$ positions of M atoms in space.
- Target = Equilibrium distribution of a (non-quantum, thermostatted) mechanical system.
- $\pi(\mathrm{d}\theta) = \mathrm{d}\theta$ is phase-space/uniform measure of indep. atomic positions.
- target = Gibbs distribution = $\eta(\mathrm{d}\theta) = \mathrm{e}^{-\beta V(\theta)} d\theta / Z_\beta$, V is Hamiltonian/interaction energy, β is inverse temperature.

Example 2 (rare event): robustness of Machine Learning

- $\theta :=$ image.
- Problem: study the robustness of a Deep Neural Network (DNN) classifier[1].
- If $f : \{\text{images}\} \to [0,1]^{\{\text{animals}\}}$ is a DNN classifier, we want to study the failure of recognizing a Panda.
- Model: $\pi(d\theta) = N(\theta_{\text{panda}}, \varepsilon \mathrm{Id})$ = small variance Gaussian distribution centered at a Panda image.



[1] Furon Tit R. *Efficient Statistical Assessment of Neural Network*

Example 2 (rare event): robustness of Machine Learning

- Misclassification event $\theta_{\text{panda}} + \text{noise} \in \mathrm{Miss}$ defined by:

$$\mathrm{argmax}(f(\underbrace{\theta_{\text{panda}} + \text{noise}}_{\sim \pi(d\ \theta)})) \neq \text{panda}$$

- Target distribution is random model conditioned by misclassification ($V(\theta) = 0$ if $\theta \in \mathrm{Miss}$ else $= +\infty$):

$$\mathrm{e}^{-V(\theta)} \pi(d\theta)/Z = 1_{\theta \in \mathrm{Miss}} \pi(d\theta)/p$$

- Normalisation is the rare event probability:

$$p = \int 1_{\theta \in \mathrm{Miss}} \pi(d\theta) = \mathbb{P}(\text{Misclassification})$$

Example 3: Bayesian statistics

- Reference distribution on parameter space with explicit easily simulated prior probability $\pi(\mathrm{d}\theta)$.
- Explicit parametric model of partial noisy observations. E.g. d-dimensional :

$$Y = T(\theta) + N(0, \varepsilon \mathrm{Id}_d) \in \mathbb{R}^d$$

- For some real observations t_{obs}, the target is the "posterior" distribution defined by Bayes formula :

$$\mathrm{Law}_\pi(\Theta \mid Y = t_{\mathrm{obs}}) = \mathrm{e}^{-|t_{\mathrm{obs}} - T(\theta)|^2/2\varepsilon} \pi(\mathrm{d}\theta)/Z =: \eta(\mathrm{d}\theta)$$

Example 3: Bayesian statistics

- Monte Carlo sampling the posterior distribution yields estimation of a typical true parameter.
- Importantly Sampling the posterior distribution yields uncertainty quantification on the statistical inference.

Example 3: Bayesian conductivity estimation problem

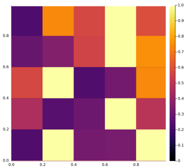
- Temperature field $T^*_\theta : \Omega \to [0, +\infty[$ solution of the elliptic parametric Partial Differential Equation :

$$\begin{cases} -\mathrm{div}_x(\kappa_\theta(x)\nabla_x T^*_\theta(x)) = f_0 & x \text{ in } \Omega \\ T^*_\theta = 0 & \text{at boundary } \partial\Omega \end{cases}$$

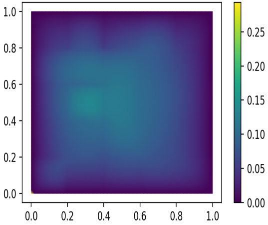
- Model:
    - $\Omega :=$ domain of the plane.
    - $\kappa_\theta(x) = \sum_{q=1}^{Q} \theta_q 1_{x\in\Omega_q} > 0 :=$ uncertain thermal conductivities at $x \in \Omega$. $\Omega_q \subset \Omega$ are sub-domains ("blocks").
    - $f_0 :=$ (constant) heating source in Ω.

Example 3: Bayesian conductivity estimation problem

- Let $\pi_{\text{prior}}(d\theta_1 \dots d\theta_Q)$ denotes the manufacturer's prior probability distribution modeling this uncertainty (e.g.: i.i.d. log-normal).



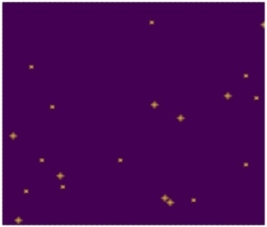
(a) Typical π_{prior}



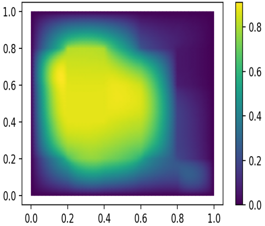
(b) Typical under π_{prior}

Example 3: Bayesian conductivity estimation problem

- Observation of an atypically large temperature :



(a) Observation points $x_1, \ldots, x_J$ with normal errors $N(0, \varepsilon)$.



(b) Atypical temperature = typical posterior

- Goal: Monte Carlo sample the Bayesian posterior distribution (e.g. maximum temperature).

2) Basic Monte Carlo samplers for target distributions

2.1) Markov Chain Monte Carlo

Markov Chain Monte Carlo

Definition (MCMC)

A Markov chain simulated with $\Theta_{n+1} = F(\Theta_n, U_{n+1})$ with $(U_n)_n$ i.i.d. having $\eta = e^{-V}\pi/Z$ as a unique invariant distribution : $\mathrm{Law}(\Theta_n) = \eta \Rightarrow \mathrm{Law}(\Theta_{n+1}) = \eta$. Moreover computing $F(\theta, u)$ only requires evaluation of $V(\theta)$ and/or $\nabla V(\theta)$.

- Ex.: F given by Metropolis rejection algorithm .
- Ex.: Discretization of $d\Theta_t = -\nabla_{\Theta_t} V dt + \sqrt{2} dW_t$.
- Ex.: Velocity jump Piecewise Deterministic Markov Process.

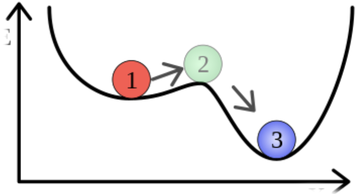
Mixing problems of MCMC

- How fast (if true !) one has:

$$\mathrm{Law}(\Theta_n) \xrightarrow[n \to +\infty]{\text{speed ?}} \underbrace{\mathrm{Law}(\Theta_\infty)}_{=\eta(\mathrm{d}\theta)=\mathrm{e}^{-\beta V(\theta)}\pi(\mathrm{d}\theta)/Z}$$

- Problem 1: η is multi-modal with unknown localization of modes (e.g. V has multiple 'peaked' local minima). The chain is typically stuck in modes ('meta-stability'):

$\theta \mapsto V(\theta)$:



Mixing problems of MCMC

- Problem 2: strong anisotropy. Typical in high dimension. E.g. the spectrum of $\mathrm{Hess}_\theta V$ is badly conditioned, extremal eigenvalues satisfy:

$$\frac{\lambda_{\min}(\mathrm{Hess}_\theta V)}{\lambda_{\max}(\mathrm{Hess}_\theta V)} \ll 1$$

MCMC rejection rate (or stability if unadjusted) constrains the fastest and most local scales:

Figure: Horizontal: slow variable; vertical: fast variable

2.2) Importance Sampling

Basic Importance Sampling (IS):

Method (Basic IS = weighting)

- Sample N i.i.d. 'clones' $(\Theta_1, \ldots, \Theta_N)$ with distribution the reference (easy to sample) π.
- Compute weights and estimate the normalisation Z as:

$$Z^N := \frac{1}{N} \sum_{n=1}^{N} \underbrace{\mathrm{e}^{-V(\Theta_n)}}_{=:W_n}$$

- Estimate the without bias non-normalized target with:

$$\gamma^N(\mathrm{d}\theta) := \frac{1}{N} \sum_{n=1}^{N} W_n \delta_{\Theta_n}(\mathrm{d}\theta).$$

Basic Importance Sampling (IS):

- Unbiasedness.
- By the law of large number, a.s.,

$$Z^N \xrightarrow[N\to+\infty]{a.s.} Z := \int \mathrm{e}^{-V(\theta)}\pi(\mathrm{d}\theta) = \text{normalization}$$

- and

$$\gamma^N(\varphi) \xrightarrow[N\to+\infty]{a.s.} \gamma(\varphi) := \int \varphi(\theta)\mathrm{e}^{-V(\theta)}\pi(\mathrm{d}\theta) = \text{un-normalized tar}$$

  for all integrable φ.
- TCL etc...

Weight problems of IS

- In the case of a rare event: $V : \mathbb{R}^d \to \{0, +\infty\}$, clones get a weight 0 or 1 so that

$$\mathbb{E}[\sharp \{n : W_n = 1\}] = Np$$

- where p is the probability of the rare event. tiny fraction (given by p) of clones involved, small effective sample size.
- Nasty in high dimension (High dimensional geometry $\Rightarrow$ probabilities tends to be singular with each other).

3) Sequential Monte Carlo for target distributions

Flow of targets

Idea (Flow of target distributions)

Instead of trying to sample directly $\eta(d\theta) = e^{-V(\theta)}\pi(d\theta)/Z$ *one constructs a continuous flow of target distributions:*

$$\beta \mapsto \eta_\beta(d\theta) := e^{-V(\theta,\beta)}\pi(d\theta)/Z_\beta,$$

with $\beta \mapsto V(\theta,\beta)$ *continuous and:*

$$V(\theta,1) = V(\theta), \qquad V(\theta,0) = 0.$$

E.g. here tempering:

$$\mathrm{e}^{-\beta V(\theta)}\pi(d\theta)/Z_\beta$$

Main steps

- Use Importance Sampling (clone weighting) to modify the target current distribution.
- Duplicate clones with large weights and kill those with small.
- Use Markov Chain MC to de-correlate clones (while sustaining target distribution).

Main SMC algorithm (target sampling pb.)

Scheme (SMC for target distributions)

$\beta_0 = 0$. N clones $(\Theta_0^1, \dots, \Theta_N^1)$i.i.d. with dist. π. Pick $\beta_1 < \dots < \beta_{i_{max}}$. Iterate on $i = 1 \dots i_{\max}$:

- (i) **Weights (Importance Sampling):** update the 'importance sampling weight' of each replica $n \in [1, N]$ to next target : $\mathrm{e}^{-\beta_i V} \mathrm{d}\pi$:

$$W_i^n := W_{i-1}^n \times \mathrm{e}^{(\beta_{i-1} - \beta_i) V(\Theta^n_{\beta_{i-1}})}$$

- (ii) **Selection** of clones according to weights (see after).
- (iii) **Mutation (MCMC)**: modify ('mutate') each clones with Markov Chain leaving invariant the new target $\mathrm{e}^{-\beta_i V} \mathrm{d}\pi / Z_{\beta_i} =: \eta_{\beta_i}$.

Mitigation of problems

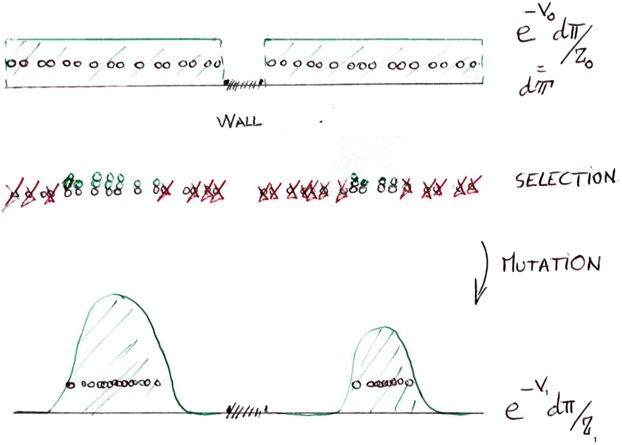


Figure: One SMC step between V_0 and V_1. V_1 must be close enough to V_0, and mutation mixes only locally.

Main SMC algorithm (target sampling pb.)

Estimators:

- Target measures $\eta_\beta = \frac{1}{Z_\beta} \mathrm{e}^{-\beta V(\theta)} \pi(d\theta)$ are estimated by empirical sample

$$\eta_{\beta_i} \simeq \eta_{\beta_i}^N := \frac{1}{N} \sum_{n=1}^{N} \delta_{\Theta_{\beta_i}^n}.$$

- Normalizations are estimated by the product of average weights over replicas

$$Z_{\beta_i} \simeq Z_{\beta_i}^N := \prod_{i'=1}^{i} \frac{1}{N} \sum_{n=1}^{N} \mathrm{W}_{\beta_i}^n \simeq \prod_{i'=1}^{i} \frac{Z_{\beta_{i'}}}{Z_{\beta_{i'-1}}}$$

Selection schemes

Definition (Selection or re-sampling scheme)

- **Require:** N clones with non-negative weights $(W^1, \ldots, W^N) \in \mathbb{R}_+^N$.
- **Do:**
  - Draw branching (cloning) numbers $(B_1, \ldots, B_N) \in \mathbb{N}^N$ for each clone:

$$\mathbb{E}\left[B^n \mid (W^1, \ldots, W^N)\right] = N \times \frac{W^n}{\sum_{m=1}^N W^m}$$

    with constraint $B_1 + \ldots + B_N = N$.
  - Kill any clone n if $B_n = 0$. If $B_n > 1$, $B_n - 1$ clones of n.

Example of selection scheme: systematic/wheel resampling

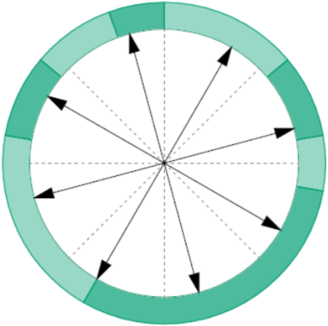


Figure: $N = 8$, Green: weights of old clones, **Black arrow**: new clones

Unbiasedness and Adaptivity

Lemma (SMC unbiasedness)

Non-normalized estimators are unbiased: for each i:

$$\mathbb{E}\left[\underbrace{Z^N_{\beta_i}\frac{1}{N}\sum_{n=1}^{N}\delta_{\Theta^n_{\beta_i}}}_{=:\gamma_i^N}\right] = \mathrm{e}^{-\beta_i V}d\pi$$

Proof.

The process $i \mapsto \int Q_{i\to i_{max}}(\theta, d\theta')\gamma_i^N(d\theta)$ is a martingale. Q is the Feynman-Kac semi-groups of the Markov Chains steps weighted by the Importance weights for $N = 1$. $\square$

Unbiasedness and Adaptivity

- Law Large Number, Central Limit Theorem, ... (see .e.g. Pierre Del Moral).
- In practice use adaptivity to tune the ladder $(\beta_i)_{i \geqslant 1}$ (e.g. to obtain exactly a $90\% \times N$ effective sample size after selection).
- Use adaptivity to tune the Markov Chain MC mutations.
- Adaptivity $\rightarrow O(1/N)$ bias and open maths questions.

Computational cost problem

- Partially solves the multi-modality and high dimension problem.
- Pb.: Very large number of evaluation of $\theta \mapsto V(\theta)/\nabla_\theta V$.
- E.g.: typical: $N = 100$ clones $\times$ 100 levels, $\times$ 10 MCMC steps $= 10^5$ evaluations + memory.

4) Model reduction with reduced basis

Model problem: thermal block problem

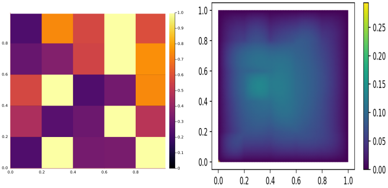
- Temperature field $T^*_\theta : \Omega \to [0, +\infty[$ solution of the parametric Elliptic Partial Differential Equation :

$$\begin{cases} -\mathrm{div}_x(\kappa_\theta(x)\nabla_x T^*_\theta(x)) = f_0 & x \text{ in } \Omega \\ T^*_\theta = 0 & \text{at boundary } \partial\Omega \end{cases}$$

- Model:
  - $\Omega :=$ domain of the plane.
  - $\kappa_\theta(x) = \sum_{q=1}^{Q} \theta_q 1_{x\in\Omega_q} > 0 :=$ uncertain thermal conductivity at $x \in \Omega$. $\Omega_q \subset \Omega$ given sub-domains ("blocks").
  - $f_0 :=$ (constant) heating source in Ω.

Model problem: thermal block problem

- Numerical Solution "low insulation".



- Numerical Solution "high insulation".

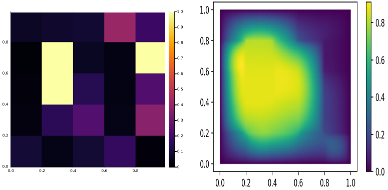


Figure: Left: conductivities (dark=small), Right: Temperature field

Model Reduction: Reduced Basis

- Assume K 'exact' solutions (snapshots) are given $T^*_{\theta(1)}, \ldots, T^*_{\theta(K)}$ for K different $\theta \in \mathbb{R}^Q$ ($Q :=$ number of blocks).
- How can one compute quickly an approximate solution T_θ for any value of θ ?
- Variational formulation of the PDE

$$T^*_\theta := \underset{T}{argmin} \frac{1}{2} \int_\Omega \kappa_\theta \left|\nabla T\right|^2 - f_0 T$$

Same thing as in linear algebra (A symmetric):

$$T = A^{-1} F_0 = \underset{T}{argmin} \frac{1}{2} \langle T, AT \rangle - \langle F_0, T \rangle$$

Model Reduction: Reduced Basis

Lemma (Reduced Basis)

Assume that the quantities $\int_{\Omega_q} \nabla T^*_{\theta(k)} \cdot \nabla T^*_{\theta(k')}$ *and* $\int_{\Omega_q} f_0 T^*_{\theta(k)}$ *are given for* $1 \leqslant q \leqslant Q$ *blocks and* $1 \leqslant k, k' \leqslant K$ *snapshots, then solving for any* $\theta \in \mathbb{R}^Q$ *the finite dimensional Reduced basis (RB) solution:*

$$T_\theta^{(K)} = \underset{T \in \operatorname{span}(T^*_{\theta(1)}, \dots, T^*_{\theta(K)})}{\text{argmin}} \frac{1}{2} \int_\Omega \kappa_\theta |\nabla T|^2 + f_0 T$$

requires $\mathrm{O}(K^3 + QK^2)$ *operations.*

Reduced Basis: A posteriori error computation

Lemma (RB a posteriori error)

Assume that the quantities $\int \mathrm{div}(1_{\Omega_q} \nabla T^*_{\theta(k)})\mathrm{div}(1_{\Omega_{q'}} \nabla T^*_{\theta(k')})$ *and* $\int_{\Omega_q} \nabla f_0 \cdot \nabla T^*_{\theta(k)}$ *are given for* $1 \leqslant q, q' \leqslant Q$ *blocks and* $1 \leqslant k, k' \leqslant K$ *snapshots, then computing for any* $\theta \in \mathbb{R}^Q$ *the following Reduced basis (RB) a posteriori error estimate:*

$$E^{(K)}(\theta) := \|\mathrm{div}(\kappa_\theta \nabla (T^*_\theta - T_\theta))\|_{\mathbb{L}^2(\Omega)}$$

requires only $\mathrm{O}(Q^2 K^2)$ *operations, and in particular does not require the exact solution* T^*_θ*.*

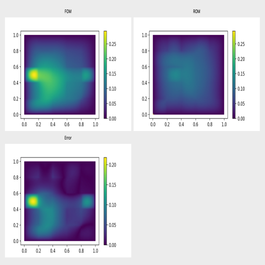
Reduced Basis: greedy-like strategies

Remark (RB greedy improvement)

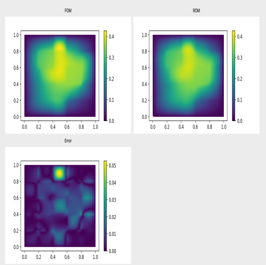
- Example: $\theta(K+1) := argmax_{\theta \in \mathcal{S}} \, E^{(K)}(\theta)$.

Model problem: thermal block problem

- Numerical Non-typical (for θ different from $\theta(k)$, $k = 1 \ldots K$) Solutions:



(a) 7 snapshots



(b) 65 snapshots

Figure: UpLeft: exact T_θ^*, UpRight: reduced T_θ BottomLeft: error $E(\theta)$

$\rightarrow$ snapshots must be well-adapted .

5) Frugal Sampling of Posterior (or Rare Event) distributions

Bayesian conductivity estimation problem

- Let $\pi(d\theta) = \pi_{\text{prior}}(\mathrm{d}\theta_1 \ldots \mathrm{d}\theta_Q)$ denotes the manufacturer's prior probability distribution modeling this uncertainty (e.g.: i.i.d. log-normal).

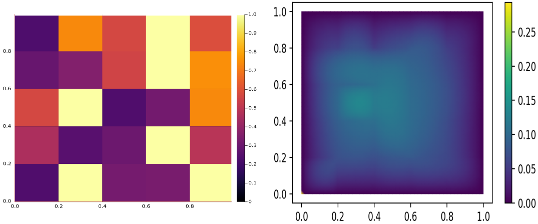
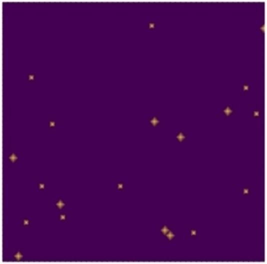


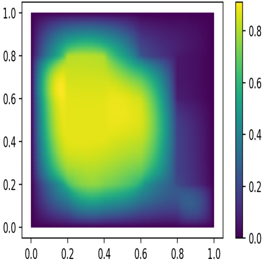
Figure: Typical under π_{prior}

Bayesian conductivity estimation problem

- Observation of an atypically large temperature :



(a) Observation points $x_1, \ldots, x_J$ with normal errors $N(0, \varepsilon)$.



(b) Atypical observed temperature $x \mapsto t_{\text{obs}}(x)$.

- Want: infer the posterior distribution of conductivities and e.g. $\sup_x T^*_\Theta$.

Bayesian sampling problem (atypical observation)

- Solution: Monte Carlo sample the posterior probability distribution (Bayes rule):

$$\eta^*_{\mathrm{posterior}}(\mathrm{d}\theta)$$
$$:= \mathrm{Law}_\pi(\Theta \mid (T_\Theta(x_j) = t_{obs}(x_j) + N(0,\varepsilon))_{j=1\ldots J})$$
$$= \frac{1}{Z(t_{\mathrm{obs}})} \mathrm{e}^{-\sum_{j=1}^{J} \frac{1}{2\varepsilon}(t_{\mathrm{obs}}(x_j) - T^*_\theta(x_j))^2} \pi(\mathrm{d}\theta)$$

- Atypical observation[2]: $\mathrm{Ent}(\eta^*_{\mathrm{posterior}} \mid \pi) \gg 1$ with perhaps log-likelihood with local minima.

[2] (Ent=relative entropy = Kullback-Leibler divergence)

Importance Sampling with reduced posterior

- Our methodology (Adaptive Reduced Tempering – ART) is based on the (Sequential) Monte Carlo sampling of the reduced posterior (frugal) flow:

$$\eta^{(k)}_{\beta,\mathrm{posterior}}(\mathrm{d}\theta) := \frac{1}{Z^{(k)}(t_{\mathrm{obs}})} \mathrm{e}^{-\sum_{j=1}^{J} \frac{\beta}{2\varepsilon}(t_{\mathrm{obs}}(x_j) - T^{(k)}_\theta(x_j))^2} \pi_{\mathrm{prior}}(\mathrm{d}\theta)$$

- $\beta \in [0,1]$ is the tempering parameter (interpolates between prior $\beta = 0$ and posterior $\beta = 1$).
- $T^{(k)}_\theta$ is computed for any θ with reduced basis based on k snapshots: $T^*_{\Theta(1)}, \dots, T^*_{\Theta(k)}$.

Importance Sampling with reduced posterior

- Key idea: Increase tempering β and number of snapshots k at the same time.
- For each k, one needs to find[3] an adapted $\hat\beta^{k+1}$:

find next $\hat\beta^{k+1}$ s.t.: $\mathrm{Ent}(\eta^*_{\hat\beta^{k+1},\mathrm{posterior}} \mid \eta^{(k)}_{\hat\beta^{k+1},\mathrm{posterior}}) \simeq \delta_0$(given

Lemma (meta) (Diaconis Chatterjee 2018)

The cost of Importance sampling of target η by proposal π is given by $\exp(\mathrm{Ent}(\eta \mid \pi))$.

- Practice: $\eta^*_{\beta,\mathrm{posterior}}(d\theta)$ too expensive. Replaced (thanks to a posteriori error estimation) by frugal pessimistic ansatz: $|T^*_\theta(x) - t_{\mathrm{obs}}(x)| \to \left|T^{(k)}_\theta(x) - t_{\mathrm{obs}}(x)\right| + E^{(k)}(\theta)$

[3] Ent = relative entropy = K.-L. divergence.

Importance Sampling with reduced posterior

- The inverse temperature $\hat{\beta}_{k+1}$ can be interpreted as an approximate critical β for the matching between exact $\eta^*_{\beta,\mathrm{posterior}}$ and reduced $\eta^{(k)}_{\beta,\mathrm{posterior}}$.
- One can then use a learning function in order to sample a new snapshot $(\Theta(k+1), T^*_{\Theta(k+1)})$ with current clones:

SMC sample $\Theta(k+1)$ according to $\dfrac{1}{Z}\mathrm{e}^{\tau E^{(k)}(\theta)}\eta^{(k)}_{\hat{\beta}_{k+1},\mathrm{posterior}}(\mathrm{d}\theta)$

- The parameter $\tau > 0$ favors parameters θ with a large error $E^{(k)}(\theta)$ (under-represented in previous snapshots $\Theta(1)\dots\Theta(k)$).

Importance Sampling with reduced posterior

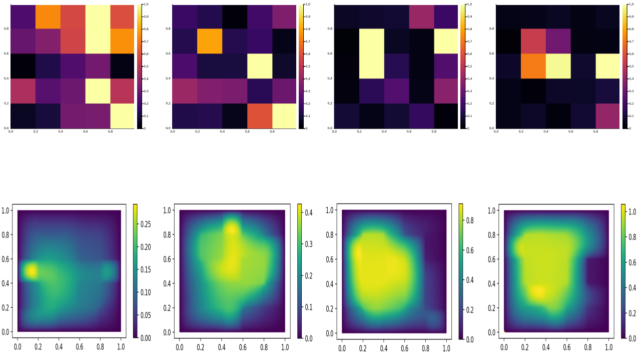
Pseudo-Code (Adaptive Reduced Tempering algorithm)

- Sample N i.i.d. parameters according to $\pi_{\text{prior}}(\mathrm{d}\theta)$
- While $\hat{\beta}_k < 1$, Sequential Monte Carlo sampling of flow:

$$\beta \mapsto \eta^{(k)}_{\beta,\text{posterior}}(\mathrm{d}\theta)$$

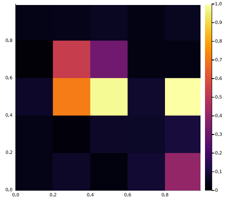
  until $\eta^{(k)}_{\hat{\beta}_{k+1},\text{posterior}}$ diverges from $\eta^{*}_{\hat{\beta}_{k+1},\text{posterior}}$ (use a posterior error $E^{(k)}$).
- Sample new snapshot $(\Theta(k+1), T^*(\Theta(k+1))$ within clones with distribution $\eta^{(k)}_{\hat{\beta}_{k+1},\text{posterior}}$.
- Compute $T^*_{\Theta(k+1)}$ and iterate $k \to k+1$

Numerical Results [C.,H.,R. 2023 and 2024])

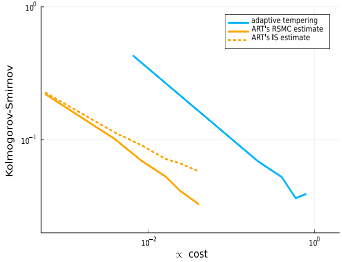


Numerical Results [C.,H.,R. 2023 and 2024])

- Kolmogorov-Smirnov (KS) distance (to "truth") of the distribution of maximum temperature $\sup_{x \in \Omega} T_\Theta(x)$:



(a) Typical a posterior temperature field



(b) KS distance versus computational log-cost for AdaptiveRedcuedTempering versus Non-Reduced

Numerical Results [C.,H.,R. 2023 and 2024])

- Computational gain here: ~ 10 for given precision
- In terms of exact eval.: 150 snapshots evaluation (Reduced Tempering) versus 10^5 snapshots (nonReduced)
- Potential gain here ~ 5000 if true model ∞ expensive.

References

SMC

- Del Moral Doucet Jasra *Sequential monte carlo samplers* 2006.
- A Beskos, A Jasra, N Kantas, A Thiery *On the convergence of adaptive sequential Monte Carlo methods* 2016.
- Del Moral, *Feynman-Kac Formulae*, 2004.
- N Chopin, O Papaspiliopoulos *An introduction to Sequential Monte Carlo* 2020

RB

- A Quarteroni, A Manzoni, F Negri *Reduced basis methods for partial differential equations: an introduction* - 2015