LANGEVIN-BASED STRATEGY FOR EFFICIENT PROPOSAL ADAPTATION IN POPULATION MONTE CARLO

Víctor Elvira[1], Émilie Chouzenoux[2]
[1]IMT Lille Douai & CRIStAL, victor.elvira@imt-lille-douai.fr
[2] Center for Visual Computing, Inria Saclay, CentraleSupélec, emilie.chouzenoux@centralesupelec.fr

INTRODUCTION

- Vector of unknowns $\mathbf{x} \in \mathcal{D} \subseteq \mathbb{R}^n$ and vector of observed data $\mathbf{y} \in \mathbb{R}^d$
- Posterior pdf

$$\tilde{\pi}(\mathbf{x}|\mathbf{y}) = \frac{\ell(\mathbf{y}|\mathbf{x})g(\mathbf{x})}{Z(\mathbf{y})} \propto \pi(\mathbf{x}|\mathbf{y}) = \ell(\mathbf{y}|\mathbf{x})g(\mathbf{x}) \quad (1)$$

where $\ell(\mathbf{y}|\mathbf{x})$ is the likelihood function, $g(\mathbf{x})$ is the prior pdf, and $Z(\mathbf{y})$ is the normalization factor

- **Goal**: Estimate a particular moment f of $\mathbf{x}$:

$$I = \frac{1}{Z}\int_{\mathcal{D}} f(\mathbf{x})\pi(\mathbf{x})d\mathbf{x} \quad (2)$$

IMPORTANCE SAMPLING

- Set of N samples drawn from a proposal pdf, $q(\mathbf{x})$
- **1. Sampling:**

$$\mathbf{x}_n \sim q(\mathbf{x}) \qquad n = 1, ..., N$$

- **2. Weighting:**

$$w_n = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}, \quad n = 1, \ldots, N \quad (3)$$

- Estimators:

$$\hat{I}_{UIS} = \frac{1}{NZ}\sum_{n=1}^{N} w_n f(\mathbf{x}_n), \qquad \hat{I}_{SNIS} = \frac{1}{\sum_{j=1}^{N} w_j}\sum_{n=1}^{N} w_n f(\mathbf{x}_n) \quad (4)$$

- **Variance** related to the discrepancy between $\pi(\mathbf{x})|f(\mathbf{x})|$ and $q(\mathbf{x})$.
- Finding a good proposal pdf, $q(\mathbf{x})$, is **critical** and **challenging**

MULTIPLE IMPORTANCE SAMPLING

- Available set of N proposal pdfs, $\{q_1(\mathbf{x}), \ldots, q_N(\mathbf{x})\}$
- Several sampling and weighting schemes are valid [2], e.g.,
- **1. Sampling:** exactly one sample is drawn from each of them, i.e.,

$$\mathbf{x}_n \sim q_n(\mathbf{x}), \qquad n = 1, ..., N$$

- **2. Weighting:** At least two possible strategies:
  - **Option 1**: *Standard* MIS (s-MIS): $w_n = \frac{\pi(\mathbf{x}_n)}{q_n(\mathbf{x}_n)}, \quad n = 1, \ldots, N$
  - **Option 2**: *Deterministic mixture MIS* (DM-MIS):

$$w_n = \frac{\pi(\mathbf{x}_n)}{\psi(\mathbf{x}_n)} = \frac{\pi(\mathbf{x}_n)}{\frac{1}{N}\sum_{j=1}^{N} q_j(\mathbf{x}_n)}, \quad n = 1, \ldots, N, \quad (5)$$

POPULATION MONTE CARLO [1]

- Population Monte Carlo (PMC) is an **adaptive** IS algorithm [2].
- DM-PMC: extension in [3] to the standard PMC.

1. **[Initialization]**: Set $\sigma > 0$, $(N, K, T) \in \mathbb{N}^+$, $\{\boldsymbol{\nu}_n\}_{n=1}^N$. For $n = 1, \ldots, N$, select the initial parameters $\boldsymbol{\mu}_n^{(1)} \in \mathbb{R}^{d_x}$. Fix the adaptation matrix $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{d_x}$.
2. **[For $t = 1$ to T]**:
   2.1 Draw one sample from each proposal pdf,

$$\mathbf{x}_n^{(t)} \sim q_n^{(t)}(\mathbf{x}; \boldsymbol{\mu}_n^{(t)}, \boldsymbol{\Sigma}, \boldsymbol{\nu}_n), \qquad n = 1, \ldots, N \quad (6)$$

   2.2 Compute the importance weights,

$$w_n^{(t)} = \frac{\pi(\mathbf{x}_n^{(t)})}{q_n^{(t)}(\mathbf{x}_n^{(t)})} \quad (7)$$

   2.3 Resample N location parameters $\{\boldsymbol{\mu}_n^{(t+1)}\}_{n=1}^N$ from the set of N weighted samples of iteration t.
3. **[Output, $t = T$]**: Return the pairs $\{\mathbf{x}_n^{(t)}, w_n^{(t)}\}$, for $n = 1, \ldots, N$ and $t = 1, \ldots, T$.

LANGEVIN DIFFUSION SCHEME

- **Langevin diffusion**: continuous-time Markov process with stationary distribution $\tilde{\pi}$ [5]
- **Unadjusted Langevin algorithm**: Discretized Langevin diffusion

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \frac{\theta^{(t)}}{2}\mathbf{A}^{1/2}(\mathbf{x}^{(t)})\nabla \log(\pi(\mathbf{x}^{(t)})) + \sqrt{\theta^{(t)}}\mathbf{A}^{1/2}(\mathbf{x}^{(t)})\boldsymbol{\omega}^{(t)} \quad (8)$$

with $\theta^{(t)}$ positive stepsize, $\mathbf{A}(\mathbf{x}^{(t)})$ SDP scaling matrix and $(\boldsymbol{\omega}^{(t)})_{t\in\mathbb{N}}$ zero-mean Gaussian i.i.d. noise.

- **Our contribution: Adapt jointly the means and the covariances** of the proposals in DM-PMC, making use of the **ULA strategy**.
  - **Scaling matrix**: Infer the local curvature of the target via second-order information $\Rightarrow$ Increased complexity but assessed performance in the context of MCMC [5].
  - **Stepsize tuning**: Backtracking scheme promotes local increase of the target value.

PROPOSED ALGORITHM: SL-PMC

1. **[Initialization]**: Set $\sigma > 0$, $(N, K, T) \in \mathbb{N}^+$, $\{\boldsymbol{\nu}_n\}_{n=1}^N$. For $n = 1, \ldots, N$, select the initial adaptive parameters $\boldsymbol{\mu}_n^{(1)} \in \mathbb{R}^{d_x}$ and $\boldsymbol{\Sigma}_n^{(1)} = \sigma^2 \mathbf{I}_{d_x}$.
2. **[For $t = 1$ to T]**:
   2.1 Draw K samples from each proposal pdf,

$$\mathbf{x}_{n,k}^{(t)} \sim q_n^{(t)}(\mathbf{x}; \boldsymbol{\mu}_n^{(t)}, \boldsymbol{\Sigma}_n^{(t)}, \boldsymbol{\nu}_n), \qquad n = 1, \ldots, N, \qquad k = 1, \ldots, K. \quad (9)$$

   2.2 Compute the importance weights,

$$w_{n,k}^{(t)} = \frac{\pi(\mathbf{x}_{n,k}^{(t)})}{\frac{1}{N}\sum_{i=1}^{N} q_i^{(t)}(\mathbf{x}_{n,k}^{(t)})} \quad (10)$$

   2.3 Resample N location parameters $\{\tilde{\boldsymbol{\mu}}_n^{(t+1)}\}_{n=1}^N$ from the set of NK weighted samples of iteration t using the **local resampling** strategy [3].
   2.4 **Adapt the proposal parameters** $\{(\boldsymbol{\mu}_n^{(t+1)}, \boldsymbol{\Sigma}_n^{(t+1)})\}_{n=1}^N$ according to

$$\begin{cases} \boldsymbol{\mu}_n^{(t+1)} & = \tilde{\boldsymbol{\mu}}_n^{(t+1)} + \theta_n^{(t+1)}\mathbf{A}(\tilde{\boldsymbol{\mu}}_n^{(t+1)})\nabla \log \pi(\tilde{\boldsymbol{\mu}}_n^{(t+1)}) \\ \boldsymbol{\Sigma}_n^{(t+1)} & = \left(\theta_n^{(t+1)}\mathbf{A}(\tilde{\boldsymbol{\mu}}_n^{(t+1)})\right)^{1/2} \end{cases}$$

3. **[Output, $t = T$]**: Return the pairs $\{\mathbf{x}_{n,k}^{(t)}, w_{n,k}^{(t)}\}$, for $n = 1, \ldots, N$, $k = 1, \ldots, K$ and $t = 1, \ldots, T$.

NUMERICAL RESULTS

Example 1.

- **Target pdf:** mixture of 5 Gaussians, i.e.,

$$\pi(\mathbf{x}) = \frac{1}{5}\sum_{n=1}^{5}\mathcal{N}(\mathbf{x}; \boldsymbol{\nu}_i, \boldsymbol{\Sigma}_i), \quad \mathbf{x} \in \mathbb{R}^2, \quad (11)$$

- **Goal:** Estimate the first and second moments, i.e., $\mathrm{E}_{\tilde{\pi}}[\mathbf{X}] = \int \mathbf{x}\pi(\mathbf{x})d\mathbf{x}$ and $\mathrm{E}_{\tilde{\pi}}[\mathbf{X}^2] = \int \mathbf{x}\pi(\mathbf{x})d\mathbf{x}$.
- For GR-PMC and LR-PMC, we set proposal covariances $\boldsymbol{\Sigma}_n = \sigma^2 \mathbf{I}_{d_x}$, with $\sigma \in \{1, 3, 5\}$, while we take $\sigma = 5$ in SL-PMC. All methods are run with $N = 50$ proposals (randomly initialized in the square $[-4, 4] \times [-4, 4]$), $T = 20$ iterations, and $K = 20$ samples per proposal and iteration.

Example 2. Let us consider the random variable (r.v.) $\mathbf{X} \in \mathbb{R}^{d_x}$, $d_x \geq 2$. This r.v. is a transformation from d_x-dimensional multivariate Gaussian $\bar{\mathbf{X}} \sim \mathcal{N}(\mathbf{x}; \mathbf{0}_{d_x}, \mathbf{C})$ with $\mathbf{C} = \mathrm{diag}(c^2, 1, ..., 1)$. The transformed r.v. is computed as $X_j = \bar{X}_j$, $j \in \{1, ..., d_x\} \setminus 2$, and $X_2 = \bar{X}_2 - b(\bar{X}_1^2 - c^2)$, with $c = 1$ and $b = 3$. This transformation leads to a complicated banana-shaped distribution with uncorrelated components. In all methods, we set $N = 50$, $K = 20$ and $T = 20$.

	GR-PMC			LR-PMC			SL-PMC
	$\sigma = 1$	$\sigma = 3$	$\sigma = 5$	$\sigma = 1$	$\sigma = 3$	$\sigma = 5$	
Z	0.6419	42.1047	0.0289	0.2807	0.1309	0.1522	0.0014
$\mathrm{E}_{\tilde{\pi}}[\mathbf{X}]$	41.3552	8.0010	0.3583	5.4810	1.6225	0.4860	0.0238
$\mathrm{E}_{\tilde{\pi}}[\mathbf{X}^2]$	12.0858	10.0200	0.5253	6.5815	2.1486	0.6844	0.0556

Table 1: **Example 1.** Relative MSE in the estimation of Z, $\mathrm{E}_{\tilde{\pi}}[\mathbf{X}]$, and $\mathrm{E}_{\tilde{\pi}}[\mathbf{X}^2]$.

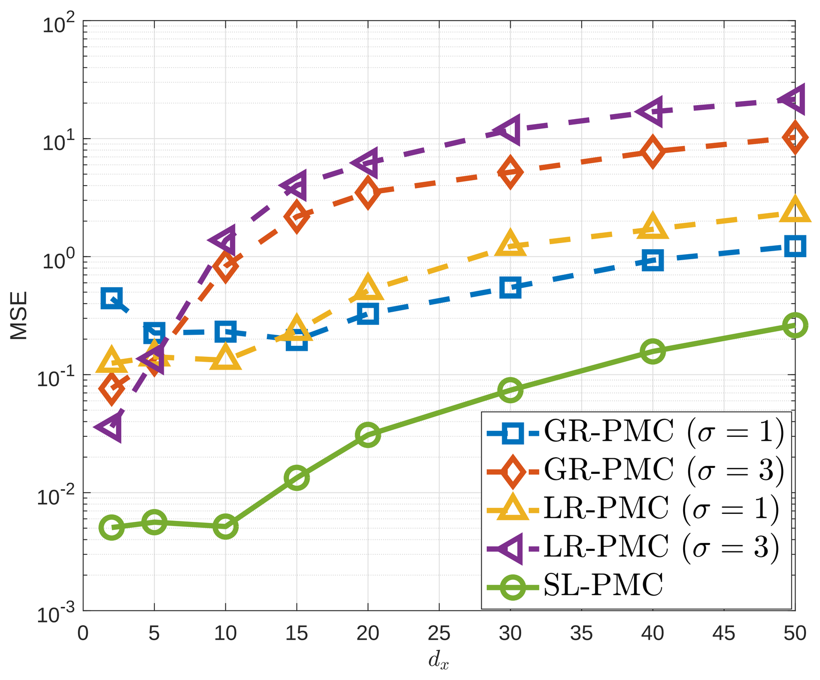

Figure 1: **Example 2.** MSE in the estimation of $\mathrm{E}_{\tilde{\pi}}[\mathbf{X}]$ of the banana-shaped distribution versus the dimension d_x.

CONCLUSION

- Joint **adaptation** of the means and the covariances of the proposals;
- Exploit the **geometry of the posterior**, with **limited parameter tuning**;
- **Superior performance**, especially in high-dimensional problems.

REFERENCES

[1] V. Elvira, L. Martino, D. Luengo, and M. Bugallo, "Generalized Multiple Importance Sampling", Statistical Science, Vol. 34, no. 1, pp. 129-155, 2019
[2] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert, "Population Monte Carlo", Journal of Computational and Graphical Statistics, pp. 907-929, 2004.
[3] V. Elvira, L. Martino, D. Luengo, and M. Bugallo, "Improving Population Monte Carlo: Alternative Weighting and Resampling Schemes", Signal Processing, vol. 131, pp. 77-91, February, 2017.
[4] G. O. Roberts and O. Stramer, "Langevin diffusions and Metropolis-Hastings algorithms", Methodol. Comput. Appl. Probab., vol. 4, no. 4, pp. 337–357, 2002.
[5] Y. Marnissi, E. Chouzenoux, A. Benazza-Benyahia, and J.C. Pesquet, "Majorize-Minimize adapted Metropolis-Hastings algorithm," Tech. Rep., 2018, https://hal.archives-ouvertes.fr/hal-01909153.