THE UNIVERSITY of EDINBURGH

MCMC-driven (and optimization-based) adaptive importance samplers

Víctor Elvira
School of Mathematics
University of Edinburgh

Joint work with Deniz Akyildiz (Imperial, UK), Émilie Chouzenoux (INRIA Saclay, France), Luca Martino (Univ. Rey Juan Carlos of Madrid, Spain), and Christian P. Robert (Univ. Paris-Dauphine)

MCM 2023, Paris (France)
June 30, 2023

Outline

Importance Sampling: Basics

- Goal: $I = \int h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x}$
- **Importance sampling** (IS) approximates I if we cannot sample from $\widetilde{\pi}(\mathbf{x}) = \frac{\pi(\mathbf{x})}{Z}$ (or we do not want to)

Importance Sampling: Basics

- Goal: $I = \int h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x}$
- **Importance sampling** (IS) approximates I if we cannot sample from $\widetilde{\pi}(\mathbf{x}) = \frac{\pi(\mathbf{x})}{Z}$ (or we do not want to)
- Steps:
  1. **Sampling:** simulate N samples from the proposal $q(x)$

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

Importance Sampling: Basics

- Goal: $I = \int h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x}$
- **Importance sampling** (IS) approximates I if we cannot sample from $\widetilde{\pi}(\mathbf{x}) = \frac{\pi(\mathbf{x})}{Z}$ (or we do not want to)
- Steps:
  1. **Sampling:** simulate N samples from the proposal $q(x)$
  $$\mathbf{x}_n \sim q(\mathbf{x}), \qquad n = 1, ..., N.$$
  2. **Weighting:**
  $$w_n = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}, \qquad n = 1, ..., N.$$

Importance Sampling: Basics

- Goal: $I = \int h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x}$
- **Importance sampling** (IS) approximates I if we cannot sample from $\widetilde{\pi}(\mathbf{x}) = \frac{\pi(\mathbf{x})}{Z}$ (or we do not want to)
- Steps:
  1. **Sampling:** simulate N samples from the proposal $q(x)$
  $$\mathbf{x}_n \sim q(\mathbf{x}), \qquad n = 1, ..., N.$$
  2. **Weighting:**
  $$w_n = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}, \qquad n = 1, ..., N.$$
- Estimators of $I = \int h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x}$:

$$\widehat{I} = \frac{1}{NZ}\sum_{n=1}^{N} w_n h(\mathbf{x}_n), \quad \text{(UIS)}$$

$$\widetilde{I} = \sum_{n=1}^{N} \bar{w}_n h(\mathbf{x}_n), \quad \text{(SNIS)}$$

with $\bar{w}_n = \frac{w_n}{\sum_{i=1}^{N} w_i}$.

Importance Sampling: Basics

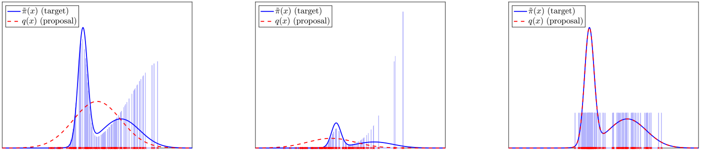
- Goal: $I = \int h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x}$
- **Importance sampling** (IS) approximates I if we cannot sample from $\widetilde{\pi}(\mathbf{x}) = \frac{\pi(\mathbf{x})}{Z}$ (or we do not want to)
- Steps:
  1. **Sampling:** simulate N samples from the proposal $q(x)$
  $$\mathbf{x}_n \sim q(\mathbf{x}), \qquad n = 1, ..., N.$$
  2. **Weighting:**
  $$w_n = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}, \qquad n = 1, ..., N.$$
- Estimators of $I = \int h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x}$:
$$\widehat{I} = \frac{1}{NZ}\sum_{n=1}^{N} w_n h(\mathbf{x}_n), \quad \text{(UIS)}$$
$$\widetilde{I} = \sum_{n=1}^{N} \bar{w}_n h(\mathbf{x}_n), \quad \text{(SNIS)}$$
with $\bar{w}_n = \frac{w_n}{\sum_{i=1}^{N} w_i}$.
- Estimator of Z: $\widehat{Z} = \frac{1}{N}\sum_{n=1}^{N} w_n$

The variance in IS and the need of better proposals

- A good proposal $q(\mathbf{x})$ is key for the efficiency of IS.
- Variance of the UIS estimator:

$$\mathrm{Var}_{\widetilde{\pi}(\mathbf{x})}(\widehat{I}) = \frac{1}{N}\int \frac{h^2(\mathbf{x})\widetilde{\pi}^2(\mathbf{x})}{q(\mathbf{x})}d\mathbf{x} - \frac{I^2}{N}$$

- optimal UIS proposal: $q(\mathbf{x}) \propto |h(\mathbf{x})|\widetilde{\pi}(\mathbf{x})$
- for a generic $h(\mathbf{x})$, $q(\mathbf{x})$ should be as *close* as possible to $\widetilde{\pi}(\mathbf{x})$

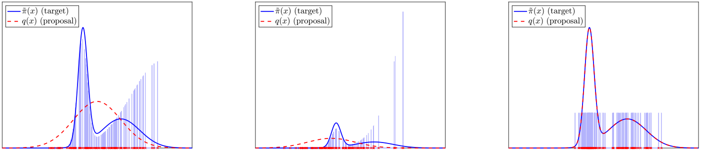


The variance in IS and the need of better proposals

- A good proposal $q(\mathbf{x})$ is key for the efficiency of IS.
- Variance of the UIS estimator:

$$\mathrm{Var}_{\widetilde{\pi}(\mathbf{x})}(\widehat{I}) = \frac{1}{N}\int \frac{h^2(\mathbf{x})\widetilde{\pi}^2(\mathbf{x})}{q(\mathbf{x})}d\mathbf{x} - \frac{I^2}{N}$$

- optimal UIS proposal: $q(\mathbf{x}) \propto |h(\mathbf{x})|\widetilde{\pi}(\mathbf{x})$
- for a generic $h(\mathbf{x})$, $q(\mathbf{x})$ should be as *close* as possible to $\widetilde{\pi}(\mathbf{x})$



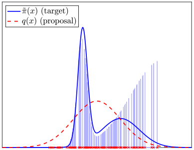
- Very difficult to find a good $q(\mathbf{x})$ a priori:
  - $\widetilde{\pi}(\mathbf{x})$ can be only evaluated (up to a normalizing constant)
  - $\widetilde{\pi}(\mathbf{x})$ may be multimodal, skewed, heavy tailed

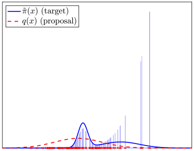
The variance in IS and the need of better proposals

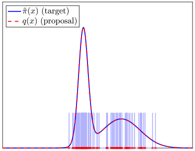
- A good proposal $q(\mathbf{x})$ is key for the efficiency of IS.
- Variance of the UIS estimator:

$$\mathrm{Var}_{\widetilde{\pi}(\mathbf{x})}(\widehat{I}) = \frac{1}{N}\int \frac{h^2(\mathbf{x})\widetilde{\pi}^2(\mathbf{x})}{q(\mathbf{x})}d\mathbf{x} - \frac{I^2}{N}$$

- optimal UIS proposal: $q(\mathbf{x}) \propto |h(\mathbf{x})|\widetilde{\pi}(\mathbf{x})$
- for a generic $h(\mathbf{x})$, $q(\mathbf{x})$ should be as *close* as possible to $\widetilde{\pi}(\mathbf{x})$







- Very difficult to find a good $q(\mathbf{x})$ a priori:
  - $\widetilde{\pi}(\mathbf{x})$ can be only evaluated (up to a normalizing constant)
  - $\widetilde{\pi}(\mathbf{x})$ may be multimodal, skewed, heavy tailed
- Use **multiple** proposals (**MIS**) and **explore** the space (**AIS**).

Outline

Adaptive Importance Sampling: Basics

- A set of N proposals $\{q_{n,j}(\mathbf{x}|\boldsymbol{\theta}_{n,j})\}_{n=1}^{N}$ is adapted over $j = 1, ..., J$ iterations

$$\{q_{n,1}(\mathbf{x}|\boldsymbol{\theta}_{n,1})\}_{n=1}^{N} \rightarrow \{q_{n,2}(\mathbf{x}|\boldsymbol{\theta}_{n,2})\}_{n=1}^{N} \rightarrow ... \rightarrow \{q_{n,J}(\mathbf{x}|\boldsymbol{\theta}_{n,J})\}_{n=1}^{N}$$

[1] M. F. Bugallo et al. "Adaptive importance sampling: The past, the present, and the future". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 60–79.

Adaptive Importance Sampling: Basics

- A set of N proposals $\{q_{n,j}(\mathbf{x}|\boldsymbol{\theta}_{n,j})\}_{n=1}^{N}$ is adapted over $j = 1, ..., J$ iterations

$$\{q_{n,1}(\mathbf{x}|\boldsymbol{\theta}_{n,1})\}_{n=1}^{N} \rightarrow \{q_{n,2}(\mathbf{x}|\boldsymbol{\theta}_{n,2})\}_{n=1}^{N} \rightarrow ... \rightarrow \{q_{n,J}(\mathbf{x}|\boldsymbol{\theta}_{n,J})\}_{n=1}^{N}$$

- AIS can be summarized to the process of adapting those parameters:

$$\{\boldsymbol{\theta}_{n,1}\}_{n=1}^{N} \rightarrow \{\boldsymbol{\theta}_{n,2}\}_{n=1}^{N} \rightarrow ... \rightarrow \{\boldsymbol{\theta}_{n,J}\}_{n=1}^{N}$$

  - abundant literature of AIS algorithms[1]

[1] M. F. Bugallo et al. "Adaptive importance sampling: The past, the present, and the future". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 60–79.

Adaptive Importance Sampling: Basics

- A set of N proposals $\{q_{n,j}(\mathbf{x}|\boldsymbol{\theta}_{n,j})\}_{n=1}^{N}$ is adapted over $j = 1, ..., J$ iterations

$$\{q_{n,1}(\mathbf{x}|\boldsymbol{\theta}_{n,1})\}_{n=1}^{N} \rightarrow \{q_{n,2}(\mathbf{x}|\boldsymbol{\theta}_{n,2})\}_{n=1}^{N} \rightarrow ... \rightarrow \{q_{n,J}(\mathbf{x}|\boldsymbol{\theta}_{n,J})\}_{n=1}^{N}$$

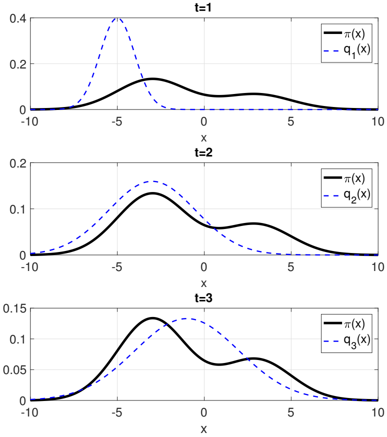
- AIS can be summarized to the process of adapting those parameters:

$$\{\boldsymbol{\theta}_{n,1}\}_{n=1}^{N} \rightarrow \{\boldsymbol{\theta}_{n,2}\}_{n=1}^{N} \rightarrow ... \rightarrow \{\boldsymbol{\theta}_{n,J}\}_{n=1}^{N}$$

  - abundant literature of AIS algorithms[1]

[1] M. F. Bugallo et al. "Adaptive importance sampling: The past, the present, and the future". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 60–79.

Adaptive Importance Sampling



Adaptive Importance Sampling: Generic Algorithm

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, and initial parameters $\{\boldsymbol{\theta}_{n,1}\}_{n=1}^{N}$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate K samples per proposal

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\theta}_{n,j}), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

Adaptive Importance Sampling: Generic Algorithm

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, and initial parameters $\{\ \boldsymbol{\theta}_{n,1}\}_{n=1}^{N}$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate K samples per proposal

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\theta}_{n,j}), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

2. **Weighting:**

$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\varphi_{n,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

Adaptive Importance Sampling: Generic Algorithm

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, and initial parameters $\{ \boldsymbol{\theta}_{n,1}\}_{n=1}^{N}$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate K samples per proposal

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\theta}_{n,j}), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

2. **Weighting:**

$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\varphi_{n,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

3. **Adaptation of the parameters:**

$$\{\boldsymbol{\theta}_{n,j}\}_{n=1}^{N} \xrightarrow{\text{Adapt}} \{\boldsymbol{\theta}_{n,j+1}\}_{n=1}^{N}$$

Adaptive Importance Sampling: Generic Algorithm

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, and initial parameters $\{\boldsymbol{\theta}_{n,1}\}_{n=1}^{N}$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate K samples per proposal

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\theta}_{n,j}), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

2. **Weighting:**

$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\varphi_{n,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

3. **Adaptation of the parameters:**

$$\{\boldsymbol{\theta}_{n,j}\}_{n=1}^{N} \xrightarrow{\text{Adapt}} \{\boldsymbol{\theta}_{n,j+1}\}_{n=1}^{N}$$

Outputs: NKT weighted samples, $\{\mathbf{x}_{n,j}^{(k)}, w_{n,j}^{(k)}\}_{n=1,k=1,j=1}^{N,K,J}$

Adaptive Importance Sampling: Generic Algorithm

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, and initial parameters $\{\boldsymbol{\theta}_{n,1}\}_{n=1}^{N}$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate K samples per proposal

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\theta}_{n,j}), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

2. **Weighting:**

$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\varphi_{n,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

3. **Adaptation of the parameters:**

$$\{\boldsymbol{\theta}_{n,j}\}_{n=1}^{N} \xrightarrow{\text{Adapt}} \{\boldsymbol{\theta}_{n,j+1}\}_{n=1}^{N}$$

Outputs: NKT weighted samples, $\{\mathbf{x}_{n,j}^{(k)}, w_{n,j}^{(k)}\}_{n=1,k=1,j=1}^{N,K,J}$

Questions:

1. Weighting scheme?[2]

[2] V. E., L. Martino, D. Luengo, and M. F. Bugallo. "Generalized Multiple Importance Sampling". In: *Statistical Science* 34.1 (2019), pp. 129–155.

Adaptive Importance Sampling: Generic Algorithm

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, and initial parameters $\{\boldsymbol{\theta}_{n,1}\}_{n=1}^{N}$
For $j = 1, \ldots, J$:

1. **Sampling:** Simulate K samples per proposal
$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\theta}_{n,j}), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$
2. **Weighting:**
$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\varphi_{n,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$
3. **Adaptation of the parameters:**
$$\{\boldsymbol{\theta}_{n,j}\}_{n=1}^{N} \xrightarrow{\text{Adapt}} \{\boldsymbol{\theta}_{n,j+1}\}_{n=1}^{N}$$

Outputs: NKT weighted samples, $\{\mathbf{x}_{n,j}^{(k)}, w_{n,j}^{(k)}\}_{n=1,k=1,j=1}^{N,K,J}$

Questions:

1. Weighting scheme?
2. Adaptive procedure of $\boldsymbol{\theta}_{n,j}$?[3]

[3] M. F. Bugallo et al. "Adaptive importance sampling: The past, the present, and the future". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 60–79.

Adaptive Importance Sampling: Generic Algorithm

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^N$, and initial parameters $\{\boldsymbol{\theta}_{n,1}\}_{n=1}^N$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate K samples per proposal

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\theta}_{n,j}), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

2. **Weighting:**

$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\varphi_{n,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

3. **Adaptation of the parameters:**

$$\{\boldsymbol{\theta}_{n,j}\}_{n=1}^N \xrightarrow{\text{Adapt}} \{\boldsymbol{\theta}_{n,j+1}\}_{n=1}^N$$

Outputs: NKT weighted samples, $\{\mathbf{x}_{n,j}^{(k)}, w_{n,j}^{(k)}\}_{n=1,k=1,j=1}^{N,K,J}$

Questions:

1. Weighting scheme?
2. Adaptive procedure of $\boldsymbol{\theta}_{n,j}$?[3]
3. How do we build the estimators?

[3] M. F. Bugallo et al. "Adaptive importance sampling: The past, the present, and the future". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 60–79.

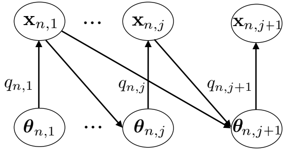
Classification of AIS methods

1. Adapt from previous samples via moment matching
2. Adapt from previous samples via resampling
3. Adapt from previous parameters (layered/hierarchical algorithms)

Classification of AIS methods (1)

1. Adapt from previous samples via moment matching

$$\{\mathbf{x}_{n,j}, w_{n,j}\}_{n=1}^{N} \xrightarrow{\text{moment matching}} \{\boldsymbol{\theta}_{n,j+1}\}_{n=1}^{N}$$



- M-PMC[5]: KL divergence minimization (MC estimate)
- APIS[6]: Multiple local approximations of target means
- **AMIS**[7]: Stochastic approximation of target moments
- CAIS[8]: Robust approximation of target moments

2. Adapt from previous samples via resampling
3. Adapt from previous parameters (layered/hierarchical algorithms)

---

[5] O. Cappé, R. Douc, A. Guillin, J. M. Marin, and C. P. Robert. "Adaptive importance sampling in general mixture classes". In: *Stat. Comput.* 18 (2008), pp. 447–459.

[6] L. Martino, V. E., D. Luengo, and J. Corander. "An adaptive population importance sampler: Learning from uncertainty". In: *IEEE Transactions on Signal Processing* 63.16 (2015), pp. 4422–4437.

[7] J. M. Cornuet, J. M. Marin, A. Mira, and C. P. Robert. "Adaptive Multiple Importance Sampling". In: *Scandinavian Journal of Statistics* 39.4 (Dec. 2012), pp. 798–812.

[8] Y. El-Laham, V. E., and M. F. Bugallo. "Robust covariance adaptation in adaptive importance sampling". In: *IEEE Signal Processing Letters* 25.7 (2018), pp. 1049–1053.

AMIS: Adaptive Multiple Importance Sampling (Cornuet et al., 2012)

Initialization: Choose J, K, q_1, $\boldsymbol{\mu}_1$ and $\mathbf{C}_1$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_1, \mathbf{C}_1)$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate K samples from q_j, i.e.,

$$\mathbf{x}_j^{(k)} \sim q_j(\mathbf{x}|\boldsymbol{\mu}_j, \mathbf{C}_j), \qquad k = 1, \ldots, K.$$

2. **Weighting:** (Re-)weight all current/past samples with a temporal mixture

$$w_\tau^{(k)} = \frac{\pi(\mathbf{x}_\tau^{(k)})}{\frac{1}{j}\sum_{i=1}^{j} q_i(\mathbf{x}_\tau^{(k)})}, \qquad k = 1, \ldots, K, \qquad \tau = 1, \ldots, j.$$

3. **Adaptation of the parameters:**[9][10] Update the mean and covariance with moment matching.

$$\{\mathbf{x}_\tau^{(k)}\}_{k=1,\tau=1}^{K,J} \xrightarrow{\text{mean estimate}} \{\boldsymbol{\mu}_{j+1}\},$$

$$\{\mathbf{x}_\tau^{(k)}\}_{k=1,\tau=1}^{K,J} \xrightarrow{\text{covariance estimate}} \{\mathbf{C}_{j+1}\},$$

Outputs: KJ weighted samples, $\{\mathbf{x}_j^{(k)}, w_j^{(k)}\}_{k=1,j=1}^{K,J}$

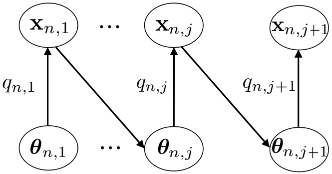
[9] J. M. Cornuet, J. M. Marin, A. Mira, and C. P. Robert. "Adaptive Multiple Importance Sampling". In: *Scandinavian Journal of Statistics* 39.4 (Dec. 2012), pp. 798–812.

[10] Y. El-Laham, V. E., and M. F. Bugallo. "Robust covariance adaptation in adaptive importance sampling". In: *IEEE Signal Processing Letters* 25.7 (2018), pp. 1049–1053.

Classification of AIS methods (2)

1. Adapt from previous samples via moment matching
2. Adapt from previous samples via resampling:

$$\{\mathbf{x}_{n,j}, w_{n,j}\}_{n=1}^{N} \xrightarrow{\text{resampling}} \{\boldsymbol{\theta}_{n,j+1}\}_{n=1}^{N}$$



- PMC[11]
- **DM-PMC and variants**[12][13]

3. Adapt from previous parameters (layered/hierarchical algorithms)

---

[11] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert. "Population Monte Carlo". In: *Journal of Comp. and Graphical Statistics* 13.4 (2004), pp. 907–929.

[12] V. E., L. Martino, D. Luengo, and M. F. Bugallo. "Improving population Monte Carlo: Alternative weighting and resampling schemes". In: *Signal Processing* 131 (2017), pp. 77–91.

[13] V. E., L. Martino, D. Luengo, and M. F. Bugallo. "Population Monte Carlo schemes with reduced path degeneracy". In: *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE. 2017, pp. 1–5.

DM-PMC: Deterministic Mixture Population Monte Carlo (E. et al, 2017)

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, $\{\boldsymbol{\mu}_{n,1}\}_{n=1}^{N}$ and $\{\mathbf{C}_n\}_{n=1}^{N}$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$

[14] V. E., L. Martino, D. Luengo, and M. F. Bugallo. "Improving population Monte Carlo: Alternative weighting and resampling schemes". In: *Signal Processing* 131 (2017), pp. 77–91.

[15] V. E., L. Martino, D. Luengo, and M. F. Bugallo. "Population Monte Carlo schemes with reduced path degeneracy". In: *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE. 2017, pp. 1–5.

DM-PMC: Deterministic Mixture Population Monte Carlo (E. et al, 2017)

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, $\{\boldsymbol{\mu}_{n,1}\}_{n=1}^{N}$ and $\{\mathbf{C}_n\}_{n=1}^{N}$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate NK samples as

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

[14] V. E., L. Martino, D. Luengo, and M. F. Bugallo. "Improving population Monte Carlo: Alternative weighting and resampling schemes". In: *Signal Processing* 131 (2017), pp. 77–91.

[15] V. E., L. Martino, D. Luengo, and M. F. Bugallo. "Population Monte Carlo schemes with reduced path degeneracy". In: *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE. 2017, pp. 1–5.

DM-PMC: Deterministic Mixture Population Monte Carlo (E. et al, 2017)

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, $\{\boldsymbol{\mu}_{n,1}\}_{n=1}^{N}$ and $\{\mathbf{C}_n\}_{n=1}^{N}$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$

For $j = 1, \ldots, J$:

1. **Sampling:** Simulate NK samples as

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

2. **Weighting:** Weight the NK samples with the DM weights as

$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

[14] V. E., L. Martino, D. Luengo, and M. F. Bugallo. "Improving population Monte Carlo: Alternative weighting and resampling schemes". In: *Signal Processing* 131 (2017), pp. 77–91.

[15] V. E., L. Martino, D. Luengo, and M. F. Bugallo. "Population Monte Carlo schemes with reduced path degeneracy". In: *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE. 2017, pp. 1–5.

DM-PMC: Deterministic Mixture Population Monte Carlo (E. et al, 2017)

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, $\{\boldsymbol{\mu}_{n,1}\}_{n=1}^{N}$ and $\{\mathbf{C}_n\}_{n=1}^{N}$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate NK samples as

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

2. **Weighting:** Weight the NK samples with the DM weights as

$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

3. **Adaptation of the parameters:**[14][15] Update the means via resampling
   - from the population of the NK samples to N location parameters

$$\{\mathbf{x}_{n,j}^{(k)}, w_{n,j}^{(k)}\}_{n=1,k=1}^{N,K} \xrightarrow{\text{Resampling}} \{\boldsymbol{\mu}_{n,j+1}\}_{n=1}^{N}$$

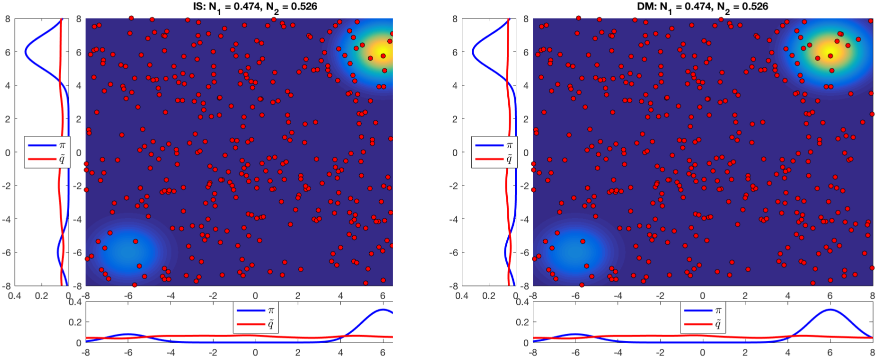
Outputs: NKJ weighted samples, $\{\mathbf{x}_{n,j}^{(k)}, w_{n,j}^{(k)}\}_{n=1,k=1,j=1}^{N,K,J}$

---

[14] V. E., L. Martino, D. Luengo, and M. F. Bugallo. "Improving population Monte Carlo: Alternative weighting and resampling schemes". In: *Signal Processing* 131 (2017), pp. 77–91.

[15] V. E., L. Martino, D. Luengo, and M. F. Bugallo. "Population Monte Carlo schemes with reduced path degeneracy". In: *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE. 2017, pp. 1–5.

DM-PMC: DM weights in resampling (iteration $j = 1$)

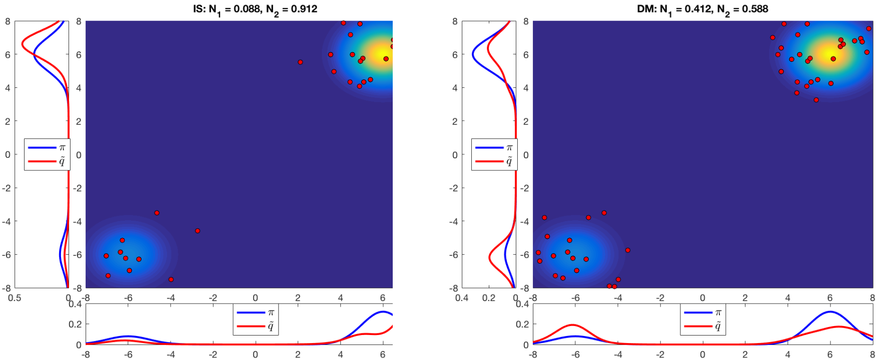


(a) Standard IS weights $w_{n,j} = \frac{\pi(\mathbf{x}_{n,j})}{q_{n,j}(\mathbf{x}_{n,j})}$.

(b) DM weights $w_{n,j} = \frac{\pi(\mathbf{x}_{n,j})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j})}$.

Figure: The red circles represent the means $\boldsymbol{\mu}_{n,j}$ of the IS proposals ($N = 500$) at iteration $j = 1$. Covariance of IS proposal $\mathbf{C}_n = \sigma_0^2 \mathbf{I}$, with $\sigma_0^2 = 0.5$.

DM-PMC: DM weights in resampling (iteration $j = 2$)

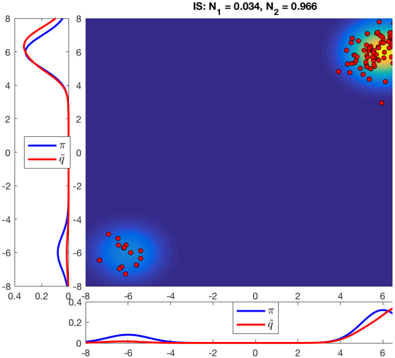


(a) Standard IS weights $w_{n,j} = \frac{\pi(\mathbf{x}_{n,j})}{q_{n,j}(\mathbf{x}_{n,j})}$.

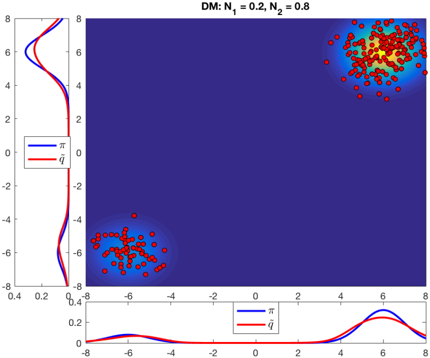
(b) DM weights $w_{n,j} = \frac{\pi(\mathbf{x}_{n,j})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j})}$.

Figure: The red circles represent the means $\boldsymbol{\mu}_{n,j}$ of the IS proposals ($N = 500$) at iteration $j = 2$. Covariance of IS proposal $\mathbf{C}_n = \sigma_0^2 \mathbf{I}$, with $\sigma_0^2 = 0.5$.

DM-PMC: DM weights in resampling (iteration $j = 3$)



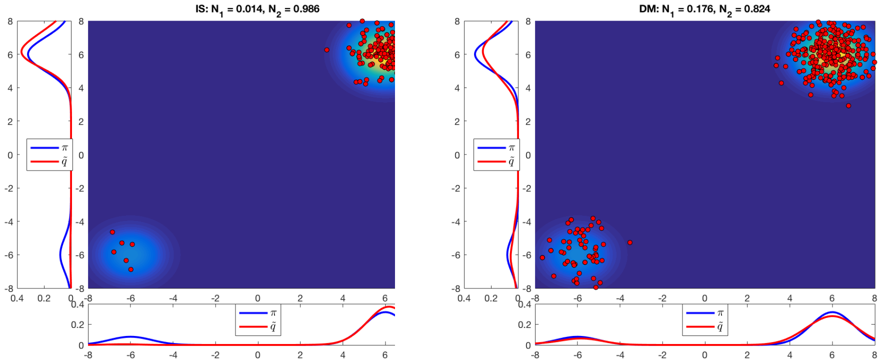
(a) Standard IS weights $w_{n,j} = \frac{\pi(\mathbf{x}_{n,j})}{q_{n,j}(\mathbf{x}_{n,j})}$.



(b) DM weights $w_{n,j} = \frac{\pi(\mathbf{x}_{n,j})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j})}$.

Figure: The red circles represent the means $\boldsymbol{\mu}_{n,j}$ of the IS proposals ($N = 500$) at iteration $j = 3$. Covariance of IS proposal $\mathbf{C}_n = \sigma_0^2\mathbf{I}$, with $\sigma_0^2 = 0.5$.

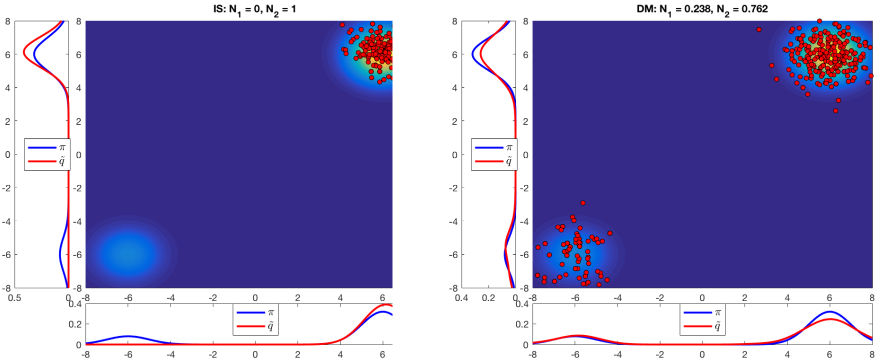
DM-PMC: DM weights in resampling (iteration $j = 4$)



(a) Standard IS weights $w_{n,j} = \frac{\pi(\mathbf{x}_{n,j})}{q_{n,j}(\mathbf{x}_{n,j})}$. (b) DM weights $w_{n,j} = \frac{\pi(\mathbf{x}_{n,j})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j})}$.

Figure: The red circles represent the means $\boldsymbol{\mu}_{n,j}$ of the IS proposals ($N = 500$) at iteration $j = 4$. Covariance of IS proposal $\mathbf{C}_n = \sigma_0^2\mathbf{I}$, with $\sigma_0^2 = 0.5$.

DM-PMC: DM weights in resampling (iteration $j = 5$)

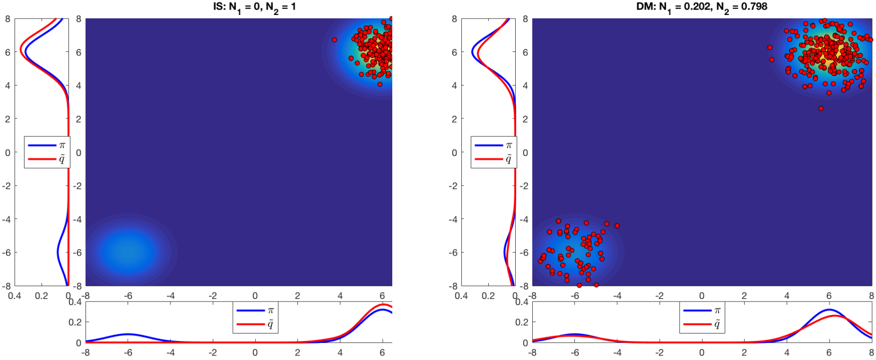


(a) Standard IS weights $w_{n,j} = \frac{\pi(\mathbf{x}_{n,j})}{q_{n,j}(\mathbf{x}_{n,j})}$.

(b) DM weights $w_{n,j} = \frac{\pi(\mathbf{x}_{n,j})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j})}$.

Figure: The red circles represent the means $\boldsymbol{\mu}_{n,j}$ of the IS proposals ($N = 500$) at iteration $j = 5$. Covariance of IS proposal $\mathbf{C}_n = \sigma_0^2\mathbf{I}$, with $\sigma_0^2 = 0.5$.

DM-PMC: DM weights in resampling (iteration $j = 6$)



(a) Standard IS weights $w_{n,j} = \frac{\pi(\mathbf{x}_{n,j})}{q_{n,j}(\mathbf{x}_{n,j})}$.

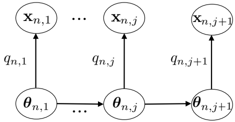
(b) DM weights $w_{n,j} = \frac{\pi(\mathbf{x}_{n,j})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j})}$.

Figure: The red circles represent the means $\boldsymbol{\mu}_{n,j}$ of the IS proposals ($N = 500$) at iteration $j = 6$. Covariance of IS proposal $\mathbf{C}_n = \sigma_0^2\mathbf{I}$, with $\sigma_0^2 = 0.5$.

Outline

Classification of AIS methods (3)

3. Adapt from previous parameters (layered/hierarchical algorithms)



- GAPIS[16]: gradient updates
- LAIS[17]: location parameters sampled via MCMC on $\widetilde{\pi}$
- O-PMC[18]: langevin dynamics and second-order information
- GRAMIS[19]: first and second order information of the target
- Purely optimization-based: variational inference[20][21], χ^2-div min.[22]

[16] V. E., L. Martino, D. Luengo, and J. Corander. "A gradient adaptive population importance sampler". In: *IEEE ICASSP*. IEEE. 2015, pp. 4075–4079.

[17] L. Martino, V. E., D. Luengo, and J. Corander. "Layered adaptive importance sampling". In: *Statistics and Computing* 27 (2017), pp. 599–623.

[18] V. E. and E. Chouzenoux. "Optimized population monte carlo". In: *IEEE Transactions on Signal Processing* 70 (2022), pp. 2489–2501.

[19] V. E., E. Chouzenoux, Ö. D. Akyildiz, and L. Martino. "Gradient-based Adaptive Importance Samplers". In: *Journal of the Franklin Institute (to appear in)* (2023).

[20] Y. El-Laham, P. M. Djurić, and M. F. Bugallo. "A variational adaptive population importance sampler". In: *IEEE ICASSP*. IEEE. 2019, pp. 5052–5056.

[21] G. Jerfel et al. "Variational refinement for importance sampling using the forward kullback-leibler divergence". In: *Uncertainty in Artificial Intelligence*. PMLR. 2021, pp. 1819–1829.

[22] E. K. Ryu and S. P. Boyd. "Adaptive importance sampling via stochastic convex programming". In: *arXiv preprint arXiv:1412.4845* (2014).

LAIS: Layered Adaptive Importance Sampling (Martino et al., 2017)

Initialization: Choose J, N, K, $\{q_{n,1}(\boldsymbol{\theta}_{n,1})\}_{n=1}^{N}$, $\boldsymbol{\theta}_{n,1} = [\boldsymbol{\mu}_{n,1}\ \mathbf{C}_n]$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$

[23] L. Martino, V. E., D. Luengo, and J. Corander. "Layered adaptive importance sampling". In: *Statistics and Computing* 27 (2017), pp. 599–623.

LAIS: Layered Adaptive Importance Sampling (Martino et al., 2017)

Initialization: Choose J, N, K, $\{q_{n,1}(\boldsymbol{\theta}_{n,1})\}_{n=1}^{N}$, $\boldsymbol{\theta}_{n,1} = [\boldsymbol{\mu}_{n,1}\ \mathbf{C}_n]$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$
For $j = 1, \ldots, J$:

1. **Sampling:** Simulate NK samples as

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

[23] L. Martino, V. E., D. Luengo, and J. Corander. "Layered adaptive importance sampling". In: *Statistics and Computing* 27 (2017), pp. 599–623.

LAIS: Layered Adaptive Importance Sampling (Martino et al., 2017)

Initialization: Choose J, N, K, $\{q_{n,1}(\boldsymbol{\theta}_{n,1})\}_{n=1}^{N}$, $\boldsymbol{\theta}_{n,1} = [\boldsymbol{\mu}_{n,1}\ \mathbf{C}_n]$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate NK samples as
$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$
2. **Weighting:** Weight the NK samples with DM weights as
$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

[23] L. Martino, V. E., D. Luengo, and J. Corander. "Layered adaptive importance sampling". In: *Statistics and Computing* 27 (2017), pp. 599–623.

LAIS: Layered Adaptive Importance Sampling (Martino et al., 2017)

Initialization: Choose J, N, K, $\{q_{n,1}(\boldsymbol{\theta}_{n,1})\}_{n=1}^{N}$, $\boldsymbol{\theta}_{n,1} = [\boldsymbol{\mu}_{n,1}\ \mathbf{C}_n]$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$

For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate NK samples as

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

2. **Weighting:** Weight the NK samples with DM weights as

$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

3. **Adaptation of the parameters:**[23] Update the means with one/several independent/dependent MCMC steps (upper layer)

$$\{\boldsymbol{\mu}_{n,j}\}_{n=1}^{N} \xrightarrow{\text{MCMC on } \tilde{\pi}} \{\boldsymbol{\mu}_{n,j+1}\}_{n=1}^{N},$$

Outputs: NKJ weighted samples, $\{\mathbf{x}_{n,j}^{(k)}, w_{n,j}^{(k)}\}_{n=1,k=1,j=1}^{N,K,J}$

[23] L. Martino, V. E., D. Luengo, and J. Corander. "Layered adaptive importance sampling". In: *Statistics and Computing* 27 (2017), pp. 599–623.

LAIS: Layered Adaptive Importance Sampling (Martino et al., 2017)

Initialization: Choose J, N, K, $\{q_{n,1}(\boldsymbol{\theta}_{n,1})\}_{n=1}^{N}$, $\boldsymbol{\theta}_{n,1} = [\boldsymbol{\mu}_{n,1}\ \mathbf{C}_n]$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$

For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate NK samples as
$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$
2. **Weighting:** Weight the NK samples with DM weights as
$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$
3. **Adaptation of the parameters:**[23] Update the means with one/several independent/dependent MCMC steps (upper layer)
$$\{\boldsymbol{\mu}_{n,j}\}_{n=1}^{N} \xrightarrow{\text{MCMC on } \tilde{\pi}} \{\boldsymbol{\mu}_{n,j+1}\}_{n=1}^{N},$$

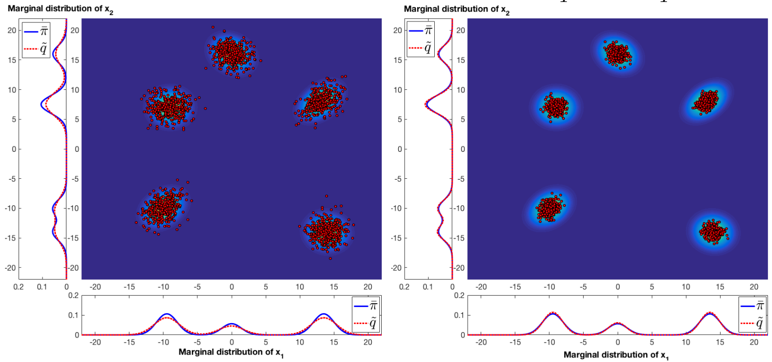
Outputs: NKJ weighted samples, $\{\mathbf{x}_{n,j}^{(k)}, w_{n,j}^{(k)}\}_{n=1,k=1,j=1}^{N,K,J}$

- LAIS can be seen as a mixture sampling from a KDE of $\tilde{\pi}$

[23] L. Martino, V. E., D. Luengo, and J. Corander. "Layered adaptive importance sampling". In: *Statistics and Computing* 27 (2017), pp. 599–623.

LAIS and AT-LAIS

- Anti-tempered LAIS (AT-LAIS)[24] runs the MCMC on $\tilde{\pi}^{\beta}$, $\beta > 1$, so the stationary distribution of the AIS samples is closer to $\tilde{\pi}$.
  - The red circles represent the means $\boldsymbol{\mu}$ of the IS proposals ($N = 2000$) in the last iteration. Covariance of IS proposal $\mathbf{C} = \sigma_T^2 \mathbf{I}$, with $\sigma_T^2 = 1$.



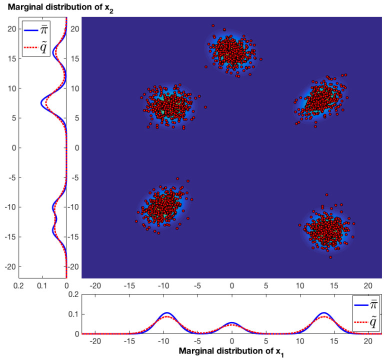
(a) LAIS ($\beta = 1$) (b) AT-LAIS ($\beta = 4$)

[24] L. Martino, V. E., and D. Luengo. "Anti-tempered layered adaptive importance sampling". In: *2017 22nd International Conference on Digital Signal Processing (DSP)*. IEEE. 2017, pp. 1–5.

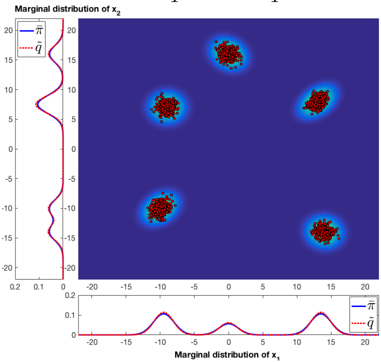
[25] A. Mousavi, R. Monsefi, and V. E. "Hamiltonian adaptive importance sampling". In: *IEEE Signal Processing Letters* 28 (2021), pp. 713–717.

LAIS and AT-LAIS

- Anti-tempered LAIS (AT-LAIS)[24] runs the MCMC on $\tilde{\pi}^{\beta}$, $\beta > 1$, so the stationary distribution of the AIS samples is closer to $\tilde{\pi}$.
  - The red circles represent the means $\boldsymbol{\mu}$ of the IS proposals ($N = 2000$) in the last iteration. Covariance of IS proposal $\mathbf{C} = \sigma_T^2 \mathbf{I}$, with $\sigma_T^2 = 1$.



(c) LAIS ($\beta = 1$)



(d) AT-LAIS ($\beta = 4$)

- Hamiltonian AIS (HAIS) version[25]

[24] L. Martino, V. E., and D. Luengo. "Anti-tempered layered adaptive importance sampling". In: *2017 22nd International Conference on Digital Signal Processing (DSP)*. IEEE. 2017, pp. 1–5.

[25] A. Mousavi, R. Monsefi, and V. E. "Hamiltonian adaptive importance sampling". In: *IEEE Signal Processing Letters* 28 (2021), pp. 713–717.

O-PMC: Optimized Population Monte Carlo (E. and Chouzenoux, 2022)

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, $\{\boldsymbol{\mu}_{n,1}\}_{n=1}^{N}$ and $\{\mathbf{C}_n\}_{n=1}^{N}$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$

[26] V. E. and E. Chouzenoux. "Optimized population monte carlo". In: *IEEE Transactions on Signal Processing* 70 (2022), pp. 2489–2501.

O-PMC: Optimized Population Monte Carlo (E. and Chouzenoux, 2022)

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, $\{\boldsymbol{\mu}_{n,1}\}_{n=1}^{N}$ and $\{\mathbf{C}_n\}_{n=1}^{N}$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$
For $j = 1, \dots, J$**:**

1. **Sampling:** Simulate NK samples as

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \dots, K, \qquad n = 1, \dots, N.$$

[26] V. E. and E. Chouzenoux. "Optimized population monte carlo". In: *IEEE Transactions on Signal Processing* 70 (2022), pp. 2489–2501.

O-PMC: Optimized Population Monte Carlo (E. and Chouzenoux, 2022)

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^N$, $\{\boldsymbol{\mu}_{n,1}\}_{n=1}^N$ and $\{\mathbf{C}_n\}_{n=1}^N$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate NK samples as

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

2. **Weighting:** Weight the NK samples with the DM weights

$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\frac{1}{N}\sum_{i=1}^N q_{i,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

[26] V. E. and E. Chouzenoux. "Optimized population monte carlo". In: *IEEE Transactions on Signal Processing* 70 (2022), pp. 2489–2501.

O-PMC: Optimized Population Monte Carlo (E. and Chouzenoux, 2022)

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^N$, $\{\boldsymbol{\mu}_{n,1}\}_{n=1}^N$ and $\{\mathbf{C}_n\}_{n=1}^N$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate NK samples as

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

2. **Weighting:** Weight the NK samples with the DM weights

$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

3. **Adaptation of the parameters:**[26]

$$\{\mathbf{C}_{n,j}\}_{n=1}^N \xrightarrow{\text{Hessian of } -log(\widetilde{\pi}(\mathbf{x}))} \{\mathbf{C}_{n,j+1}\}_{n=1}^N$$

$$\{\mathbf{x}_{n,j}^{(k)}, w_{n,j}^{(k)}\}_{n=1,k=1}^{N,K} \xrightarrow{\text{Resampling}} \{\widetilde{\boldsymbol{\mu}}_{n,j}\}_{n=1}^N$$

$$\{\widetilde{\boldsymbol{\mu}}_{n,j}, \mathbf{C}_{n,j}\}_{n=1}^N \xrightarrow{\text{gradient step}} \{\boldsymbol{\mu}_{n,j+1}\}_{n=1}^N$$

Outputs: NKJ weighted samples, $\{\mathbf{x}_{n,j}^{(k)}, w_{n,j}^{(k)}\}_{n=1,k=1,j=1}^{N,K,J}$

[26] V. E. and E. Chouzenoux. "Optimized population monte carlo". In: *IEEE Transactions on Signal Processing* 70 (2022), pp. 2489–2501.

Outline

GRAMIS: Gradient Adaptive Multiple Importance Sampling (E. et al, 2023)

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, $\{\boldsymbol{\mu}_{n,1}\}_{n=1}^{N}$ and $\{\mathbf{C}_n\}_{n=1}^{N}$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$

[27] V. E., E. Chouzenoux, Ö. D. Akyildiz, and L. Martino. "Gradient-based Adaptive Importance Samplers". In: *Journal of the Franklin Institute (to appear in)* (2023).

GRAMIS: Gradient Adaptive Multiple Importance Sampling (E. et al, 2023)

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^N$, $\{\boldsymbol{\mu}_{n,1}\}_{n=1}^N$ and $\{\mathbf{C}_n\}_{n=1}^N$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$

For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate NK samples as

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

[27] V. E., E. Chouzenoux, Ö. D. Akyildiz, and L. Martino. "Gradient-based Adaptive Importance Samplers". In: *Journal of the Franklin Institute (to appear in)* (2023).

GRAMIS: Gradient Adaptive Multiple Importance Sampling (E. et al, 2023)

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^N$, $\{\boldsymbol{\mu}_{n,1}\}_{n=1}^N$ and $\{\mathbf{C}_n\}_{n=1}^N$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$
For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate NK samples as

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

2. **Weighting:** Weight the NK samples with the DM weights

$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\frac{1}{N}\sum_{i=1}^N q_{i,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

[27] V. E., E. Chouzenoux, Ö. D. Akyildiz, and L. Martino. "Gradient-based Adaptive Importance Samplers". In: *Journal of the Franklin Institute (to appear in)* (2023).

GRAMIS: Gradient Adaptive Multiple Importance Sampling (E. et al, 2023)

Initialization: Choose J, N, K, $\{q_{n,1}\}_{n=1}^{N}$, $\{\boldsymbol{\mu}_{n,1}\}_{n=1}^{N}$ and $\{\mathbf{C}_n\}_{n=1}^{N}$, e.g., $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{n,1}, \mathbf{C}_n)$

For $j = 1, \ldots, J$**:**

1. **Sampling:** Simulate NK samples as

$$\mathbf{x}_{n,j}^{(k)} \sim q_{n,j}(\mathbf{x}|\boldsymbol{\mu}_{n,j}, \mathbf{C}_n), \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

2. **Weighting:** Weight the NK samples with the DM weights

$$w_{n,j}^{(k)} = \frac{\pi(\mathbf{x}_{n,j}^{(k)})}{\frac{1}{N}\sum_{i=1}^{N} q_{i,j}(\mathbf{x}_{n,j}^{(k)})}, \qquad k = 1, \ldots, K, \qquad n = 1, \ldots, N.$$

3. **Adaptation of the parameters:**[27]

$$\{\mathbf{C}_{n,j}\}_{n=1}^{N} \xrightarrow{\text{Hessian of } -log(\widetilde{\pi}(\mathbf{x}))} \{\mathbf{C}_{n,j+1}\}_{n=1}^{N}$$

$$\{\boldsymbol{\mu}_{n,j}, \mathbf{C}_{n,j}\}_{n=1}^{N} \xrightarrow{\text{gradient step + repulsion among proposals}} \{\boldsymbol{\mu}_{n,j+1}\}_{n=1}^{N}$$

Outputs: NKJ weighted samples, $\{\mathbf{x}_{n,j}^{(k)}, w_{n,j}^{(k)}\}_{n=1,k=1,j=1}^{N,K,J}$

[27] V. E., E. Chouzenoux, Ö. D. Akyildiz, and L. Martino. "Gradient-based Adaptive Importance Samplers". In: *Journal of the Franklin Institute (to appear in)* (2023).

GRAMIS: Gradient Adaptive Multiple Importance Sampling [Elvira23]

- At the j-th iteration of the GRAMIS:
  - Mean adaptation of the n-th proposal:

$$\boldsymbol{\mu}_{n,j} = \boldsymbol{\mu}_{n,j-1} + \mathbf{C}_{n,j-1} \nabla \log\left(\pi(\boldsymbol{\mu}_{n,j-1})\right) + \sum_{i=1, i\neq n}^{N} \mathbf{r}_{n,i}^{(j-1)}, \qquad (1)$$

    with

$$\mathbf{r}_{n,i}^{(j-1)} = G_j \frac{m_n m_i}{\|\mathbf{d}_{n,i}^{(j-1)}\|^{d_x}} \mathbf{d}_{n,i}^{(j-1)}, \qquad (2)$$

    where $\|\cdot\|$ represents the norm operator, $\mathbf{d}_{n,i}^{(j-1)} = \boldsymbol{\mu}_{n,j-1} - \boldsymbol{\mu}_{i,j-1}$, and $m_n, m_i > 0$ are two positive terms that depend on the n-th and i-th proposals respectively.
  - Covariance adaptation of the n-th proposal:

$$\mathbf{C}_{n,j} = \begin{cases} \left(-\nabla^2 \log \pi(\boldsymbol{\mu}_{n,j})\right)^{-1}, & \text{if } \nabla^2 \log \pi(\boldsymbol{\mu}_{n,j}) \succ 0, \\ \mathbf{C}_{n,j-1}, & \text{otherwise.} \end{cases} \qquad (3)$$

Repulsion term in GRAMIS

- Update of location parameter:

$$\boldsymbol{\mu}_{n,j} = \boldsymbol{\mu}_{n,j-1} + \mathbf{C}_{n,j-1}\nabla \log\left(\pi(\boldsymbol{\mu}_{n,j-1})\right) + \sum_{i=1, i\neq n}^{N} G_j \frac{m_n m_i}{\|\mathbf{d}_{n,i}^{(j-1)}\|^{d_x}} \mathbf{d}_{n,i}^{(j-1)}, \ \mathbf{d}_{n,j}^{(j-1)} = \boldsymbol{\mu}_{n,j-1} - \boldsymbol{\mu}_{i,j-1}$$

Repulsion term in GRAMIS

- Modified update of location parameter:

$$\boldsymbol{\mu}_{n,j} = \boldsymbol{\mu}_{n,j-1} + \gamma \nabla \log\left(\pi(\boldsymbol{\mu}_{n,j-1})\right) + \sum_{i=1, i \neq n}^{N} \frac{\gamma}{S_{d_x-1}(1) \|\mathbf{d}_{n,i}^{(j-1)}\|^{d_x}} \mathbf{d}_{n,i}^{(j-1)}, \quad \mathbf{d}_{n,i}^{(j-1)} = \boldsymbol{\mu}_{n,j-1} - \boldsymbol{\mu}_{i,j-1}$$

- Poisson field:

$$-\nabla \varphi^N(\boldsymbol{\mu}_{n,j-1}) = -\int \nabla_u G(\boldsymbol{u}, \boldsymbol{v}) \rho_{n,j-1}^N(\mathrm{d}\boldsymbol{v})$$
$$= \frac{1}{N-1} \sum_{i=1, j \neq n}^{N} \frac{1}{S_{d_x-1}(1)} \frac{\mathbf{d}_{n,i}^{(j-1)}}{\|\mathbf{d}_{n,i}^{(j-1)}\|^{d_x}}$$

  with potential function

$$G(\boldsymbol{u}, \boldsymbol{v}) = \frac{1}{(d_x - 2) S_{d_x-1}(1)} \frac{1}{\|\boldsymbol{u} - \boldsymbol{v}\|^{d_x-2}},$$

  and the empirical measure (of $N-1$ interpreted charges)

$$\rho_{n,j-1}^N(\mathrm{d}\boldsymbol{v}) = \frac{1}{N-1} \sum_{i=1, i \neq n}^{N} \delta_{\boldsymbol{\mu}_{i,j-1}}(\mathrm{d}\boldsymbol{v}),$$

- Summary: the n-th location parameter update balances (a) maximization of π due to $\nabla \log\left(\pi(\boldsymbol{\mu}_{n,j-1})\right)$, and (b) repulsion from all other location parameters due to $-\nabla \varphi_{t-1}^N(\boldsymbol{\mu}_{n,j-1})$

Outline

Mixture of generalized Gaussians

RMSE	AMIS[28]			LR-PMC[29]			O-PMC[30]			GRAMIS[31]		
	η=0.5	η=1	η=1.5	η=0.5	η=1	η=1.5	η=0.5	η=1	η=1.5	η=0.5	η=1	η=1.5
Z	0.969	0.969	0.969	0.012	0.515	0.610	0.009	0.640	0.626	$6.4 \cdot 10^{-4}$	$2.5 \cdot 10^{-6}$	$2.6 \cdot 10^{-3}$
$\mathrm{E}_{\widetilde{\pi}}[\mathbf{X}]$	56.1	56.1	56.1	1.234	50.49	55.2	0.623	56.1	55.5	0.025	$1.5 \cdot 10^{-4}$	0.110
$\mathrm{E}_{\widetilde{\pi}}[\mathbf{X}^2]$	67.95	67.93	68.03	1.12	55.6	65.8	1.029	67.9	66.6	1.296	0.002	0.188
$\chi^2\left(\widetilde{\pi}, \psi^{(T)}\right)$	980.1	980.1	980.1	30.6	556.7	616.8	7.84	639.9	630.3	0.778	0.005	4.34

Table: **Mixture of 5 generalized Gaussians.** RMSE of the IS estimators. In the case of GRAMIS, $G_1 = 1$ with exponential decay. The MSE results are obtained over 100 independent runs, with estimators using the weighted samples on the half last iterations.

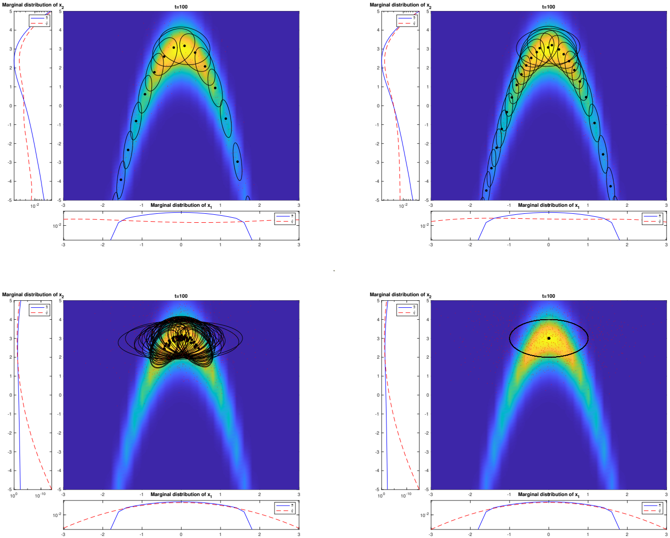
[28] J. M. Cornuet, J. M. Marin, A. Mira, and C. P. Robert. "Adaptive Multiple Importance Sampling". In: *Scandinavian Journal of Statistics* 39.4 (Dec. 2012), pp. 798–812.

[29] V. E., L. Martino, D. Luengo, and M. F. Bugallo. "Improving population Monte Carlo: Alternative weighting and resampling schemes". In: *Signal Processing* 131 (2017), pp. 77–91.

[30] V. E. and E. Chouzenoux. "Optimized population monte carlo". In: *IEEE Transactions on Signal Processing* 70 (2022), pp. 2489–2501.

[31] V. E., E. Chouzenoux, Ö. D. Akyildiz, and L. Martino. "Gradient-based Adaptive Importance Samplers". In: *Journal of the Franklin Institute (to appear in)* (2023).

GRAMIS in banana-shaped distribution (repulsion study)



[32] V. E., E. Chouzenoux, Ö. D. Akyildiz, and L. Martino. "Gradient-based Adaptive Importance Samplers". In: *Journal of the Franklin Institute (to appear in)* (2023).

Outline

Conclusion

- IS is a powerful methodology but requires the adaptation of the proposal(s).
- AIS methods have grown greatly in the last decade with new daptive mechanisms:
  - standard MCMC algorithms
  - HMC
  - Langevin dynamics
  - variational algorithms
  - interpretations from physical systems
  - ...
- We should probably drop the dichotomy between
  - optimization and sampling techniques
  - MCMC and IS

Thank you for your attention!

AIS Review. M. F. Bugallo, V. E., L. Martino, D. Luengo, J. Miguez, and P. M. Djuric, "Adaptive Importance Sampling: The Past, the Present, and the Future", IEEE Signal Processing Magazine, Vol. 34, No. 4, pp. 60-79, 2017.

DM-PMC. V. E., L. Martino, D. Luengo, and M. F. Bugallo, "Improving Population Monte Carlo: Alternative Weighting and Resampling Schemes", Signal Processing, Vol. 131, pp. 77-91, February, 2017.

O-PMC. V. E., E. Chouzenoux, "Optimized Population Monte Carlo", IEEE Transactions on Signal Processing, Vol. 70, pp. 2489 - 2501, 2022.

GRAMIS. V. E., E. Chouzenoux, O. D. Akyildiz, L. Martino, "Gradient-based adaptive importance samplers", to appear in Journal of the Franklin Institute, 2023.

email: victor.elvira@ed.ac.uk