A POPULATION MONTE CARLO SCHEME FOR COMPUTATIONAL INFERENCE IN HIGH DIMENSIONAL SPACES

Eugenia Koblents, Joaquín Míguez

Universidad Carlos III de Madrid, Department of Signal Theory and Communications
Avda. de la Universidad 30, 28911 Leganés, Spain, E-mail: ekoblents,jmiguez@tsc.uc3m.es

ABSTRACT

In this paper we address the Monte Carlo approximation of integrals with respect to probability distributions in high-dimensional spaces. In particular, we investigate the population Monte Carlo (PMC) scheme, which is based on an iterative importance sampling (IS) approach. Both IS and PMC suffer from the well known problem of degeneracy of the importance weights (IWs), which is closely related to the curse-of-dimensionality, and limits their applicability in large-scale practical problems. In this paper we investigate a novel PMC scheme that consists in performing nonlinear transformations of the IWs in order to smooth their variations and avoid degeneracy. We apply the modified IS scheme to the well-known mixture-PMC (MPMC) algorithm, which constructs the importance functions as mixtures of kernels. We present numerical results that show how the modified version of MPMC clearly outperforms the original scheme.

Index Terms— Importance sampling, population Monte Carlo, mixture-PMC, degeneracy of importance weights

1. INTRODUCTION

Computational inference in high-dimensional spaces is a challenging problem with a broad scope. Various techniques based on the Monte Carlo methodology [1] have been successfully applied to a large variety of complex problems. In this work we concentrate on the importance sampling (IS) approach and its iterative extension, the population Monte Carlo (PMC) scheme [2].

IS allows to perform inference on a target probability density function (pdf) of interest, based on samples generated from a proposal pdf, or importance function, and their associated importance weights (IWs). This method is very sensitive to the selection of the proposal pdf and usually reveals a poor performance unless the dimension of the probability space is low [3]. The IWs present huge variations, leading to a low number of "effective" samples [4], i.e., those with non-negligible IWs. As a consequence, an extremely large number of samples is required in order to obtain a set of representative ones, even in simple and low-dimensional problems [3, 5].

The PMC method iteratively performs IS, adapting the importance function according to the previous samples and weights, so that it "approaches" the static target pdf along the iterations. As based on IS, PMC is also prone to suffer from the degeneracy of the IWs, and the availability of a low number of effective samples heavily hinders the update of the proposal, leading to numerical problems.

E. K. acknowledges the support of *Ministerio de Educación, Cultura y Deporte* of Spain (*Programa de Formación de Profesorado Universitario*, ref. AP2008-00469). This work has been partially supported by *Ministerio de Economía y Competitividad* of Spain (program Consolider-Ingenio 2010 CSD2008-00010 COMONSENS, project DEIPRO TEC2009-14504-C02-01 and project COMPREHENSION TEC2012-38883-C02-01).

Nevertheless, the IS approach and PMC present a set of advantages compared to the widely established Markov chain Monte Carlo (MCMC) methodology [1], such as the possibility of developing parallel implementations, the sample independency and the inexistence of a convergence period. Therefore, our work is directed towards alleviating the inefficiency of IS in high-dimensional problems. In [5], a simple modification of the classical IS scheme was proposed, which consists in applying nonlinear transformations to the IWs in order to reduce their variations and obtain a sufficient number of effective samples. Asymptotic convergence results were established for the approximation error of integrals with respect to the random measure constructed with the modified weights.

In this paper, we apply this novel IS scheme with transformed IWs (TIWs) to the recently introduced mixture-PMC (MPMC) algorithm [6], which constructs the sequence of importance functions as mixtures of kernels. In [6], a set of rules are put forward for selecting the parameters of the importance function that minimize the Kullback-Leibler distance (KLD) between the target and the proposal pdf at each iteration of the PMC scheme. However, this method suffers from severe numerical problems unless applied to signal spaces of low dimension, hence, the authors of [6] propose an additional Rao-Blackwellization scheme to improve its robustness.

Here we introduce a MPMC algorithm that computes TIWs at each iteration and demonstrate numerically (for a Gaussian mixture target model) that this modification leads to a large performance improvement compared to the original method of [6]. The new algorithm avoids numerical problems and leads to a smooth and stable convergence toward the target pdf with high probability.

The rest of the paper is organized as follows. In Section 2, we give a formal statement of the class of problems we address and then review IS and the PMC method. In Section 3 we describe the nonlinear IS and its application in the PMC framework. In Section 4, we particularize the methodology to derive a MPMC algorithm with TIWs and present numerical results. In Section 5 we discuss the contributions of this paper and provide some concluding remarks.

2. BACKGROUND

2.1. Problem statement

Let $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_K]^\top$ be a vector of K unobserved real random variables with pdf $\pi(\boldsymbol{\theta})$, termed the *target* pdf. The Monte Carlo framework allows to approximate integrals with respect to $\pi(\boldsymbol{\theta})$ using a random grid of M points, $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^{M}$, in the space of the random vector $\boldsymbol{\theta}$. However, the generation of useful samples that represent the probability measure $\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ adequately when K is large is normally a very difficult task. The goal of this work is to devise and assess an efficient computational inference (Monte Carlo) methodology, based on the IS principle, for the approximation of $\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$

and its moments, i.e., integrals of the form $(f,\pi)=\int f(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$, where $f:\mathbb{R}^K\to\mathbb{R}$ is a real, integrable function of $\boldsymbol{\theta}$.

2.2. Importance sampling

The main application of statistical Monte Carlo methods is the approximation of (f,π) by means of empirical sums of the form

$$(f,\hat{\pi}^M)=\frac{1}{M}\sum_{i=1}^{M}f(\boldsymbol{\theta}^{(i)}),\ \text{where}\ \hat{\pi}^M(d\boldsymbol{\theta})=\frac{1}{M}\sum_{i=1}^{M}\delta_{\boldsymbol{\theta}^{(i)}}(d\boldsymbol{\theta}),$$

$\delta_{\boldsymbol{\theta}^{(i)}}(d\boldsymbol{\theta})$ is the unit delta measure located at $\boldsymbol{\theta}=\boldsymbol{\theta}^{(i)}$, and $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^M$ is a random i.i.d. (independent and identically distributed) sample drawn from $\pi(\boldsymbol{\theta})$. It is straightforward to analyze the convergence of $(f,\hat{\pi}^M)$ towards (f,π) [1].

However, in many practical cases it is not possible to sample from $\pi(\boldsymbol{\theta})$ directly. A common approach to overcome this difficulty is to apply an IS procedure [1]. The key idea is to draw the samples $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^M$ from a (simpler) proposal pdf, or importance function, $q(\boldsymbol{\theta})$, and then compute normalized IWs as

$$w^{(i)}\propto\pi(\boldsymbol{\theta}^{(i)})/q(\boldsymbol{\theta}^{(i)}),\ \ \text{with}\ \ \sum_{i=1}^{M}w^{(i)}=1.$$

The integral (f,π) is then approximated by the weighted sum

$$(f,\pi^M)=\sum_{i=1}^{M}w^{(i)}f(\boldsymbol{\theta}^{(i)}),\ \text{where}\ \pi^M(d\boldsymbol{\theta})=\sum_{i=1}^{M}w^{(i)}\delta_{\boldsymbol{\theta}^{(i)}}(d\boldsymbol{\theta}).$$

The efficiency of an IS algorithm depends heavily on the choice of the proposal, $q(\boldsymbol{\theta})$. However, in order to ensure the asymptotic convergence of the approximation (f,π^M), when M is large enough, it is sufficient to select $q(\boldsymbol{\theta})$ such that $q(\boldsymbol{\theta})>0$ whenever $\pi(\boldsymbol{\theta})>0$, and guarantee that $q(\boldsymbol{\theta})$ has heavier tails than $\pi(\boldsymbol{\theta})$ [1].

2.3. Population Monte Carlo algorithm

The PMC method [2] is an iterative IS scheme that generates a sequence of proposal pdfs $q_\ell(\boldsymbol{\theta})$, $\ell=1,\dots,L$, such that every new proposal is "closer" to the target density $\pi(\boldsymbol{\theta})$ than the previous importance function. Such scheme demands the ability to learn about the target $\pi(\boldsymbol{\theta})$, given the set of samples and weights at the $(\ell-1)$-th iteration ($\ell\geq 2$), in order to produce the new proposal $q_\ell(\boldsymbol{\theta})$ for the ℓ-th iteration. The PMC algorithm is outlined in Table 1.

Table 1. Generic PMC algorithm

Iteration ($\ell=1,\dots,L$):

1. Select a proposal pdf $q_\ell(\boldsymbol{\theta})$.
2. Draw a collection of M i.i.d. samples $\{\boldsymbol{\theta}_\ell^{(i)}\}_{i=1}^M$ from $q_\ell(\boldsymbol{\theta})$.
3. Compute normalized IWs $w_\ell^{(i)}\propto\pi(\boldsymbol{\theta}_\ell^{(i)})/q_\ell(\boldsymbol{\theta}_\ell^{(i)})$, $i=1,\dots,M$.

When the dimension K is large, the probability of generating representative samples from $q_\ell(\boldsymbol{\theta})$ is very low, unless the proposal fits well the target pdf, leading to extreme variations of the normalized IWs. Also, the so called effective sample size (ESS), approximated as $M_\ell^{eff}=[\sum_{i=1}^M(w_\ell^{(i)})^2]^{-1}$, and its normalized version (NESS) $M_\ell^{neff}=M_\ell^{eff}/M$, take very low values, leading to severe numerical problems [2, 5].

3. NONLINEAR POPULATION MONTE CARLO

In this section we describe the nonlinear IS method, in which the IWs undergo a nonlinear transformation. Then, we briefly discuss the use of a specific type of transformation and, finally, introduce the general PMC algorithm with TIWs.

3.1. Nonlinear IS

In [5] a modification of the standard IS approach was introduced, in which the TIW $\bar{w}^{(i)}$ associated to a sample $\boldsymbol{\theta}^{(i)}$ is computed as a nonlinear transformation of the standard unnormalized IW $w^{(i)*}$. To be specific, one chooses a transformation function $\varphi:(\mathbb{R}^+)^M\times\{1,\dots,M\}\to\mathbb{R}^+$ and then computes the TIWs as

$$\bar{w}^{(i)}\propto\varphi^M(w^{(i)*}),\ i=1,\dots,M,$$

where $\varphi^M(w^{(i)*})$ is shorthand for $\varphi(\{w^{(j)*},j=1,\dots,M\},i)$. The TIWs are normalized to yield $\sum_{i=1}^M\bar{w}^{(i)}=1$. We may obtain an approximation of (f,π) based on the set of TIWs as

$$(f,\bar{\pi}^M)=\sum_{i=1}^{M}\bar{w}^{(i)}f(\boldsymbol{\theta}^{(i)}),\ \text{where}\ \bar{\pi}^M(d\boldsymbol{\theta})=\sum_{i=1}^{M}\bar{w}^{(i)}\delta_{\boldsymbol{\theta}^{(i)}}(d\boldsymbol{\theta}).$$

Some basic results for the asymptotic convergence of $(f,\bar{\pi}^M)$ towards (f,π) can also be found in [5].

3.2. Clipping transformation

The nonlinearity φ^M may be constructed in multiple ways. In this paper we only investigate a "clipping" transformation. Consider a permutation $i_1,\dots,i_M$ of the indices in $\{1,...,M\}$ such that $w^{(i_1)*}\geq\dots\geq w^{(i_M)*}$ and choose an integer $1<M_T<M$. The unnormalized TIWs $\bar{w}^{(i)*}$ are computed from the IWs $w^{(i)*}$ as

$$\bar{w}^{(i)*}=\varphi^M(w^{(i)*})=\min(w^{(i)*},w^{(i_{M_T})*}),\quad i=1,\dots,M,$$

where the threshold value $w^{(i_{M_T})*}$ corresponds to the M_T-th highest IW. This transformation leads to flat TIWs in the region of interest of $\boldsymbol{\theta}$, and guarantees a baseline ESS of M_T.

3.3. Nonlinear PMC

The nonlinear IS approach may be directly extended to the iterative PMC scheme. Thus, a nonlinear transformation φ_ℓ^M is applied to the standard IWs at each iteration ℓ. The generic nonlinear PMC algorithm is outlined in Table 2. An arbitrary proposal $q_1(\boldsymbol{\theta})$ is assumed for the first iteration of the algorithm.

Table 2. Generic nonlinear PMC algorithm

Iteration ($\ell=1,\dots,L$):

1. Given a set of samples and TIWs at iteration $\ell-1$, $\{\boldsymbol{\theta}_{\ell-1}^{(i)},\bar{w}_{\ell-1}^{(i)}\}_{i=1}^M$, select a proposal pdf $q_\ell(\boldsymbol{\theta})$ ($\ell\geq 2$).
2. Draw a collection of M i.i.d. samples $\{\boldsymbol{\theta}_\ell^{(i)}\}_{i=1}^M$ from $q_\ell(\boldsymbol{\theta})$.
3. For $i=1,\dots,M$, compute unnormalized IWs $w_\ell^{(i)*}\propto\pi(\boldsymbol{\theta}_\ell^{(i)})/q_\ell(\boldsymbol{\theta}_\ell^{(i)})$.
4. For $i=1,\dots,M$, compute TIWs $\bar{w}_\ell^{(i)*}=\varphi_\ell^M(w_\ell^{(i)*})$ and normalize them $\bar{w}_\ell^{(i)}=\bar{w}_\ell^{(i)*}/\sum_{j=1}^M\bar{w}_\ell^{(j)*}$.

This modification of the algorithm mitigates the sensitivity of the conventional IS to the selection of the importance function. The normalized TIWs present a lower variance than the standard IWs, increasing the ESS and allowing for a more robust proposal update.

4. EXAMPLE: NONLINEAR MIXTURE-PMC

In this section we demonstrate how the computation of TIWs can dramatically improve the performance of conventional PMC algorithms. In particular, we describe the recently proposed mixture-PMC (MPMC) method of [6], explain how to enhance it using nonlinear IS and finally show comparative computer simulation results.

4.1. Nonlinear mixture-PMC algorithm

We apply the general nonlinear PMC technique of Section 3.3 to extend the MPMC algorithm [6]. The latter is based on constructing the sequence of proposal pdfs as mixtures of kernels of the form

$$q_\ell(\boldsymbol{\theta}) = \sum_{d=1}^{D} \alpha_{\ell,d} q_{\ell,d}(\boldsymbol{\theta}; \boldsymbol{\beta}_{\ell,d}), \tag{1}$$

where the mixture weights $\alpha_{\ell,d}$ and the kernel parameters $\boldsymbol{\beta}_{\ell,d}$ of each component are adapted along the iterations in order to minimize the KLD between the target and the proposal pdf [7].

The MPMC update mechanism is similar to the EM algorithm [7] with the E-step replaced by IS computations, and is outlined in Table 3. In [6] an additional Rao-Blackwellization (RB) scheme is employed, in order to mitigate the numerical accuracy problems that arise when carrying out computations in high dimensional spaces. The plain and RB schemes only differ in step 3.

Table 3. Original mixture-PMC algorithm [6]

Iteration ($\ell = 1, \ldots, L$)**:**

1. Generate a sample $\{\boldsymbol{\theta}_\ell^{(i)}\}_{i=1}^M$ from the current mixture proposal $q_\ell(\boldsymbol{\theta})$ in Eq. (1).
2. For $i = 1, \ldots, M$, compute normalized IWs $w_\ell^{(i)} \propto w_\ell^{(i)*} = \pi(\boldsymbol{\theta}_\ell^{(i)})/q_\ell(\boldsymbol{\theta}_\ell^{(i)})$ and normalized mixture posterior probabilities $\rho_{\ell,d}^{(i)}$, which satisfy $\sum_{d=1}^D \rho_{\ell,d}^{(i)} = 1$, as
$$\rho_{\ell,d}^{(i)*} \propto \alpha_{\ell,d} q_{\ell,d}(\boldsymbol{\theta}_\ell^{(i)}; \boldsymbol{\beta}_{\ell,d}), \quad \rho_{\ell,d}^{(i)} = \rho_{\ell,d}^{(i)*} / \sum_{k=1}^{D} \rho_{\ell,k}^{(i)*}.$$
3. Update the weights and the parameters of each component as
$$\alpha_{\ell+1,d} = \sum_{i=1}^{M} w_\ell^{(i)} \xi_{\ell,d}^{(i)} \quad \text{and}$$
$$\boldsymbol{\beta}_{\ell+1,d} = \arg\max_{\boldsymbol{\beta}_{\ell,d}} \left[\sum_{i=1}^{M} w_\ell^{(i)} \xi_{\ell,d}^{(i)} \log q_{\ell,d}(\boldsymbol{\theta}_\ell^{(i)}; \boldsymbol{\beta}_{\ell,d}) \right].$$
Let $Z_\ell^{(i)} \in \{1, \ldots, M\}$ be the random index that identifies from which kernel the sample $\boldsymbol{\theta}_\ell^{(i)}$ has been drawn in step 1. In the plain MPMC scheme, $\xi_{\ell,d}^{(i)} = \mathbb{1}\{Z_\ell^{(i)} = d\}$, while in the RB-MPMC scheme $\xi_{\ell,d}^{(i)} = \rho_{\ell,d}^{(i)}$.

A nonlinear MPMC algorithm is easily obtained by additionally computing TIWs in step 2, i.e., $\bar{w}_\ell^{(i)*} = \varphi_\ell^M(w_\ell^{(i)*})$, $\bar{w}_\ell^{(i)} = \bar{w}_\ell^{(i)*} / \sum_{j=1}^M \bar{w}_\ell^{(j)*}$ and replacing $w_\ell^{(i)}$ by $\bar{w}_\ell^{(i)}$ in step 3.

4.2. Gaussian mixture importance function

Following [6], assume that the proposal pdf at iteration ℓ is a mixture of Gaussian kernels of the form

$$q_\ell(\boldsymbol{\theta}) = \sum_{d=1}^{D} \alpha_{\ell,d} \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\mu}_{\ell,d}, \boldsymbol{\Sigma}_{\ell,d}),$$

where $\boldsymbol{\mu}_{\ell,d}$ and $\boldsymbol{\Sigma}_{\ell,d}$ are the mean vector and the covariance matrix of each component, respectively (i.e., $\boldsymbol{\beta}_{\ell,d} = \{\boldsymbol{\mu}_{\ell,d}, \boldsymbol{\Sigma}_{\ell,d}\}$). The MPMC algorithm provides the update rules for the weights and parameters of each component in the Gaussian case as

$$\alpha_{\ell+1,d} = \sum_{i=1}^{M} w_\ell^{(i)} \xi_{\ell,d}^{(i)}, \quad \boldsymbol{\mu}_{\ell+1,d} = \frac{\sum_{i=1}^M w_\ell^{(i)} \xi_{\ell,d}^{(i)} \boldsymbol{\theta}_\ell^{(i)}}{\alpha_{\ell+1,d}} \quad \text{and}$$

$$\boldsymbol{\Sigma}_{\ell+1,d} = \frac{\sum_{i=1}^M w_\ell^{(i)} \xi_{\ell,d}^{(i)} \left(\boldsymbol{\theta}_\ell^{(i)} - \boldsymbol{\mu}_{\ell+1,d}\right)\left(\boldsymbol{\theta}_\ell^{(i)} - \boldsymbol{\mu}_{\ell+1,d}\right)^\top}{\alpha_{\ell+1,d}},$$

where the $\xi_{\ell,d}^{(i)}$'s are as in step 3 of Table 3. In the nonlinear MPMC algorithm we compute $\alpha_{\ell+1,d}$, $\boldsymbol{\mu}_{\ell+1,d}$ and $\boldsymbol{\Sigma}_{\ell+1,d}$ in the same way but replacing the IWs $w_\ell^{(i)}$ by TIWs $\bar{w}_\ell^{(i)}$, $i = 1, \ldots, M$.

4.3. Simulation results

We present computer simulation results to compare the original and nonlinear MPMC algorithms for a target density consisting of a Gaussian mixture in a 10-dimensional space. In particular, let

$$\begin{aligned}\pi(\boldsymbol{\theta}) = 0.35\mathcal{N}(\boldsymbol{\theta}; -2\mathbf{1}_{10}, 0.5\mathbf{I}_{10}) &+ 0.4\mathcal{N}(\boldsymbol{\theta}; 0.5\mathbf{1}_{10}, 0.25\mathbf{I}_{10}) \\ &+ 0.25\mathcal{N}(\boldsymbol{\theta}; 2\mathbf{1}_{10}, 0.5\mathbf{I}_{10}),\end{aligned}$$

where $\mathbf{1}_{10} = [1, \ldots, 1]^\top$ and $\mathbf{I}_{10}$ is the 10×10 identity matrix. For the nonlinear MPMC method, we use the clipping transformation of Section 3.2 in order to compute the TIWs and set $M_T = \sqrt{M}$ (to guarantee that $\lim_{M\to\infty} M_T/M = 0$ and ensure the asymptotic convergence of the nonlinear IS method, see [5]).

4.3.1. IS vs nonlinear IS

If we restrict our attention to the first iteration of the PMC schemes, then we can carry out a comparison of the standard IS method (with conventional IWs) and the nonlinear IS technique (with TIWs). We consider an importance function given by the 10-dimensional Gaussian pdf $q(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}; \mathbf{0}_{10}, 10\mathbf{I}_{10})$, where $\mathbf{0}_{10} = [0, \ldots, 0]^\top$, which corresponds to a vague prior knowledge. Both the marginal target and proposal are represented in Fig. 1 (left).

We compute an approximation of the mean of $\pi(\boldsymbol{\theta})$, i.e., $\hat{\boldsymbol{\theta}} = \int \boldsymbol{\theta}\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$, based on a set of M samples $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^M$ from $q(\boldsymbol{\theta})$, with standard IWs as $\hat{\boldsymbol{\theta}}^M = \sum_{i=1}^M w^{(i)}\boldsymbol{\theta}^{(i)}$ and with TIWs as $\bar{\boldsymbol{\theta}}^M = \sum_{i=1}^M \bar{w}^{(i)}\boldsymbol{\theta}^{(i)}$. In Fig. 1 (central) we depict the approximation error obtained by standard IS ($\|\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^M\|^2$) and nonlinear IS ($\|\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}^M\|^2$), as a function of the number of samples M, averaged over 10^4 independent simulation runs. The exact Monte Carlo error (approximating $\hat{\boldsymbol{\theta}}$ with samples generated from $\pi(\boldsymbol{\theta})$) is also depicted for comparison. It can be clearly observed that the approximation error obtained with nonlinear IS is far below the one obtained with standard IS. Thus, the number of samples needed with nonlinear IS to obtain a given approximation error is much lesser. Correspondingly, Fig. 1 (right) shows that the average ESS obtained with standard IS increases more slowly than that of nonlinear IS.

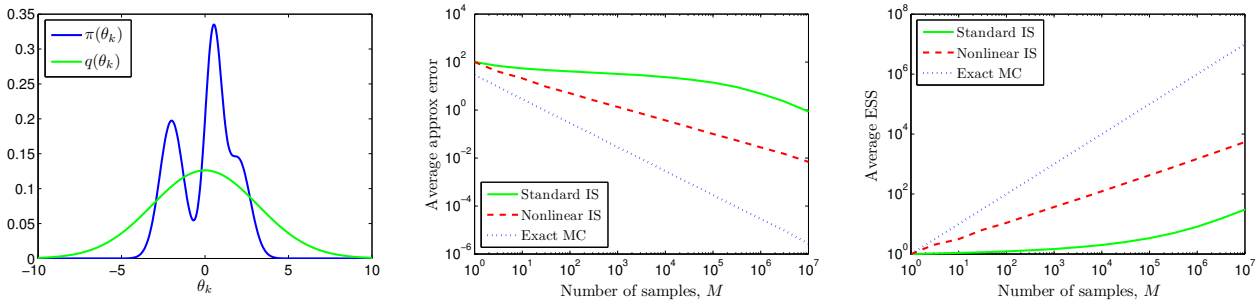

Fig. 1. Performance of standard IS vs nonlinear IS. Marginal target and marginal proposal pdf (*left*). Average approximation error (*central*) and ESS (*right*) vs M with standard IS, nonlinear IS and exact Monte Carlo sampling (labeled as MC).

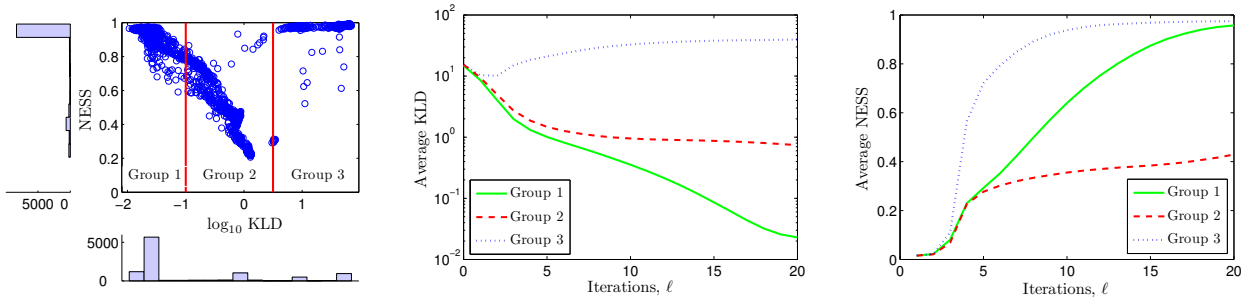

Fig. 2. Performance of nonlinear RB-MPMC. Final KLD vs final NESS obtained at each simulation run with the nonlinear RB-MPMC (*left*). The obtained outcomes can be classified into three groups according to the final KLD, the threshold values are represented in solid red lines. Evolution of the KLD (*central*) and NESS (*right*) along the iterations with the nonlinear RB-MPMC in each of the three groups.

4.3.2. MPMC vs nonlinear MPMC

We have performed 10^4 independent simulation runs of the following algorithms: MPMC, RB-MPMC, nonlinear MPMC and nonlinear RB-MPMC. In all the simulations, we considered an initial proposal pdf $q_1(\boldsymbol{\theta})$ composed of $D = 5$ equally weighted Gaussian components with covariance matrix $10\mathbf{I}_{10}$ and random mean vectors. The number of iterations has been set to $L = 20$ and the number of samples per iteration to $M = 5000$. At each iteration of all PMC schemes we compute an approximation of the KLD between the target and the ℓ-th proposal by Monte Carlo simulation.

Fig. 2 (left) depicts the final KLD in logarithmic scale versus the final NESS obtained at each simulation run of the nonlinear RB-MPMC, together with the corresponding histograms. We observe that most of the simulation runs end up with a low KLD (below 10^{-1}) and a NESS close to 1. The outcomes of this type correspond to exact matching of the final proposal to the target pdf and are classified into Group 1. Outcomes with a final KLD between 10^{-1} and $10^{0.5}$ belong to Group 2 and correspond to solutions in which some of the modes are grouped into one. The outcomes with a final KLD above $10^{0.5}$ belong to Group 3 and correspond to solutions where some of the modes are ignored. These threshold values are also represented in Fig. 2 (left) in solid red line. Additionally, we define Group 4 containing simulation runs which ended with non-proper solutions or numerical errors. Fig. 2 (central and right plots) shows the evolution along the iterations of the average KLD and NESS, respectively, of the outcomes belonging to Groups 1, 2 and 3.

In Table 4 the percentage of results obtained in each of these groups, for each of the tested algorithms are shown. We observe that the nonlinear (NL) MPMC schemes (both the plain and RB versions) clearly outperform the original MPMC, which obtains outcomes in Group 4 in most of the cases. The nonlinear RB-MPMC obtains $\approx 70\%$ of outcomes in Group 1 and presents no numerical problems.

Table 4. Percentage of simulation runs belonging to each group.

	Group 1	Group 2	Group 3	Group 4
MPMC	0 %	0 %	1 %	99 %
RB-MPMC	0 %	0.07 %	4.34 %	95.59 %
NL MPMC	14.65 %	45.51 %	34.73 %	5.11 %
NL RB-MPMC	**69.96 %**	**14.64 %**	**15.40 %**	**0 %**

5. SUMMARY OF CONTRIBUTIONS

We have investigated the combination of nonlinear IS and PMC schemes, termed nonlinear PMC and originally proposed in [5]. While in [5] the proposal pdf's were assumed to be simple Gaussians and the method was applied only to low-dimensional problems (where weight degeneracy was due to the need to evaluate very narrow likelihood functions), in this paper we have applied the methodology: (a) to enhance the MPMC algorithm of [6], where the proposal densities are mixtures of kernels built by way of a sophisticated method that involves the minimization of a KLD, and (b) to solve a computational inference problem in a higher-dimensional space (where weight degeneracy is due to the curse of dimensionality). We have shown, through computer simulations, that the resulting nonlinear MPMC algorithms drastically outperform their conventional MPMC counterparts, in terms of both estimation accuracy and robustness to numerical precision issues.

6. REFERENCES

[1] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 2004.

[2] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert, "Population Monte Carlo," *Computational and Graphical Statistics*, vol. 13, no. 4, pp. 907–929, 2004.

[3] T. Bengtsson, P. Bickel, and B. Li, "Curse of dimensionaluty revisited: Collapse of particle filter in very large scale systems," *Probability and statistics: Essay in honour of David A. Freedman*, vol. 2, pp. 316–334, 2008.

[4] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and Bayesian missing data problems," *Journal of the American Statistical Association*, vol. 9, pp. 278–288, 1994.

[5] E. Koblents and J. Míguez, "A population Monte Carlo scheme with transformed weights and its application to stochastic kinetic models," *arXiv preprint arXiv:1208.5600*, 2012.

[6] O. Cappé, R. Douc, A. Guillin, J. M. Marin, and C. P. Robert, "Adaptive importance sampling in general mixture classes," *Statistics and Computing*, vol. 18, no. 4, pp. 447–459, 2008.

[7] C.M. Bishop et al., *Pattern recognition and machine learning*, vol. 4, springer New York, 2006.