A Particle Method for Solving Fredholm Equations of the First Kind: Supplementary Material

Francesca R. Crucinio, Arnaud Doucet and Adam M. Johansen

A   Notation

For the convenience of the reader, we summarize the notation adopted in the following arguments. A slightly more technical presentation is adopted than within the main manuscript as a little care is required in order to obtain rigorous results.

We work on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ rich enough to allow the definition of the particle system introduced in Section 3 for all $N \in \mathbb{N}$. All expectations and probabilities which are not explicitly associated with some other measure are taken with respect to $\mathbb{P}$.

For any $\mathbb{H} \subseteq \mathbb{R}^d$, we consider the Borel σ-algebra $B(\mathbb{H})$ with respect to the Euclidean norm, and we endow any product space with the product Borel σ-algebra.

Let the Banach space of real-valued bounded measurable functions on $\mathbb{H}$, endowed with the supremum norm, $\|\varphi\|_\infty = \sup_{u \in \mathbb{H}} |\varphi(u)|$, be denoted by $\mathcal{B}_b(\mathbb{H})$.

Let $\mathcal{M}(\mathbb{H})$ be the Banach space of signed finite measures on $(\mathbb{H}, B(\mathbb{H}))$ endowed with the bounded Lipschitz norm (e.g. Dudley (2002, page 394))

$$\beta(\eta) := \sup_{\|\varphi\|_{BL} \leq 1} \left| \int_{\mathbb{H}} \eta(\mathrm{d}x)\varphi(x) \right|, \tag{25}$$

where $\| \cdot \|_{BL}$ denotes the bounded Lipschitz norm for bounded Lipschitz functions φ

$$\|\varphi\|_{BL} := \|\varphi\|_\infty + \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{\|x - y\|_2}.$$

For ease of notation, for every measure $\eta \in \mathcal{M}(\mathbb{H})$ and every $\varphi \in \mathcal{B}_b(\mathbb{H})$ we denote the integral of φ with respect to η by $\eta(\varphi) := \int_{\mathbb{H}} \eta(\mathrm{d}u)\varphi(u)$.

We denote by $\mathcal{M}^+(\mathbb{H}) \subset \mathcal{M}(\mathbb{H})$ the set of (unsigned) measures of nonzero mass and by $\mathcal{P}(\mathbb{H}) \subset \mathcal{M}^+(\mathbb{H})$ the set of all probability measures on $(\mathbb{H}, B(\mathbb{H}))$. For every $\eta \in \mathcal{P}(\mathbb{H})$ we have

$$\beta(\eta) = \sup_{\|\varphi\|_{BL} \leq 1} \left| \int_{\mathbb{H}} \eta(\mathrm{d}x)\varphi(x) \right| \leq \sup_{\|\varphi\|_{BL} \leq 1} \|\varphi\|_\infty \eta(\mathbb{H}) \leq 1.$$

The β norm metrizes weak convergence (Dudley, 2002, Theorem 11.3.3) in $\mathcal{M}(\mathbb{X})$: for every $\mu \in \mathcal{M}(\mathbb{H})$, and sequence $\{\mu_n\}_{n \geq 1}$ taking values in $\mathcal{M}(\mathbb{H})$, $\beta(\mu_n, \mu) \to 0$ is equivalent to $\mu_n(\varphi) \to \mu(\varphi)$ for all continuous bounded functions $\varphi \in C_b(\mathbb{H})$.

For any $\eta \in \mathcal{M}^+(\mathbb{H})$ and any positive function G integrable with respect to η we denote by $\Psi_G(\eta)(\mathrm{d}x)$ the Boltzmann-Gibbs transform

$$\Psi_G(\eta)(\mathrm{d}x) = \frac{1}{\eta(G)} G(x)\eta(\mathrm{d}x).$$

A Markov kernel M from $\mathbb{H}$ to $\mathbb{H}$ induces two operators. One acts upon measures in $\mathcal{M}(\mathbb{H})$ and takes values in $\mathcal{M}(\mathbb{H})$ and is defined by

$$\forall \eta \in \mathcal{M}(\mathbb{H}) \quad \eta M(\cdot) = \int_{\mathbb{H}} \eta(\mathrm{d}u)M(u, \cdot)$$

and the other acts upon functions in $\mathcal{B}_b(\mathbb{H})$ and takes values in $\mathcal{B}_b(\mathbb{H})$ and may be defined as

$$\forall u \in \mathbb{H} \quad \forall \varphi \in \mathcal{B}_b(\mathbb{H}) \quad M(\varphi)(u) = \int_{\mathbb{H}} M(u, \mathrm{d}v)\varphi(v).$$

For each $\omega \in \Omega$, we obtain a realization of the particle system with N particles at time n and a corresponding random measure denoted by $\eta_n^N : \omega \in \Omega \mapsto \eta_n^N(\omega) \in \mathcal{P}(\mathbb{H})$

$$\eta_n^N(\omega)(\cdot) = \frac{1}{N} \sum_{i=1}^N \delta_{(X_n^i(\omega), Y_n^i(\omega))}(\cdot),$$

where we suppress from the notation the dependence of $X_n^i(\omega)$ and $Y_n^i(\omega)$ upon N, as we shall throughout in the interest of readability.

B   Existence of the Fixed Point

Let us formally define the EMS map as a map from the set of unsigned measures of nonzero mass to the set of probability measures, $\mathrm{F}_{\mathrm{EMS}} : \mathcal{M}^+(\mathbb{X}) \to \mathcal{P}(\mathbb{X})$, such that

$$\mathrm{F}_{\mathrm{EMS}} : \eta \mapsto \mathrm{F}_{\mathrm{EMS}}\, \eta := \int_{\mathbb{X}} \eta(\mathrm{d}x') K(x', \cdot) \int_{\mathbb{Y}} \frac{g(y \mid x')h(y)}{\int_{\mathbb{X}} \eta(\mathrm{d}z)g(y \mid z)}\ \mathrm{d}y$$

and the EM map, $\mathrm{F}_{\mathrm{EM}} : \mathcal{M}^+(\mathbb{X}) \to \mathcal{P}(\mathbb{X})$ as in (7), slightly more formally as:

$$\mathrm{F}_{\mathrm{EM}}(\eta)(\mathrm{d}x) = \frac{1}{\eta(\bar{G}_\eta)} \bar{G}_\eta(x)\eta(\mathrm{d}x), \tag{26}$$

where the normalizing constant $\eta(\bar{G}_\eta) \equiv 1$ is introduced to highlight the connection with the particle methods introduced in Section 3. We introduce the smoothing operator, $\mathrm{K} : \mathcal{P}(\mathbb{X}) \to \mathcal{P}(\mathbb{X})$, corresponding to the smoothing kernel in (A2)

$$\mathrm{K} : \eta \mapsto \eta K := \int_{\mathbb{X}} \eta(\mathrm{d}v) K(v, \cdot) \tag{27}$$

and observe that $\mathrm{F}_{\mathrm{EMS}}\, \eta = \mathrm{K}\left(\mathrm{F}_{\mathrm{EM}}(\eta)\right) = \left(\mathrm{F}_{\mathrm{EM}}\, \eta\right) K$.

In order to prove that the EMS map admits a fixed point, a number of properties of the EM map, of the smoothing operator K and of the EMS map itself must be established. We show that $\mathrm{F}_{\mathrm{EMS}}$ is a compact operator on $\mathcal{M}^+(\mathbb{X})$ (Corollary 2). To do so, we show that F_{EM} is continuous and bounded (Proposition 6) then we prove that K is compact (Proposition 7). Compactness is needed to prove existence of a fixed point.

B.1   Properties of the Continuous EMS Map

Proposition 6. Under (A0) and (A1), the EM map F_{EM} in (26) is a continuous and bounded operator on $\mathcal{M}^+(\mathbb{X})$ endowed with the weak topology.

Proof. Let $\eta \in \mathcal{M}^+(\mathbb{X})$ and $\{\eta_n\}_{n \geq 1}$ be a sequence of measures in $\mathcal{M}^+(\mathbb{X})$ converging to η in the weak topology as $n \to \infty$. For any $\varphi \in C_b(\mathbb{X})$ consider

$$\left| \int_{\mathbb{X}} \mathrm{F}_{\mathrm{EM}}(\eta_n)(\mathrm{d}x)\varphi(x) - \int_{\mathbb{X}} \mathrm{F}_{\mathrm{EM}}(\eta)(\mathrm{d}x)\varphi(x) \right|$$

$$= \left| \int_{\mathbb{X}} \eta_n(\mathrm{d}x)\varphi(x) \int_{\mathbb{Y}} \frac{g(y \mid x)h(\mathrm{d}y)}{\eta_n\left(g(y \mid \cdot)\right)} - \int_{\mathbb{X}} \eta(\mathrm{d}x)\varphi(x) \int_{\mathbb{Y}} \frac{g(y \mid x)h(\mathrm{d}y)}{\eta\left(g(y \mid \cdot)\right)} \right|$$

$$= \left| \int_{\mathbb{X}} \int_{\mathbb{Y}} \varphi(x)g(y \mid x)h(\mathrm{d}y) \left[\frac{\eta_n(\mathrm{d}x)}{\eta_n\left(g(y \mid \cdot)\right)} - \frac{\eta(\mathrm{d}x)}{\eta\left(g(y \mid \cdot)\right)} \right] \right|$$

$$\leq \left| \int_{\mathbb{X}} \int_{\mathbb{Y}} \frac{\eta_n(\mathrm{d}x)\varphi(x)g(y \mid x)h(\mathrm{d}y)}{\eta_n\left(g(y \mid \cdot)\right)\eta\left(g(y \mid \cdot)\right)} \left[\eta\left(g(y \mid \cdot)\right) - \eta_n\left(g(y \mid \cdot)\right)\right] \right|$$

$$+ \left| \int_{\mathbb{X}} \int_{\mathbb{Y}} \left(\eta_n(\mathrm{d}x) - \eta(\mathrm{d}x)\right) \frac{\varphi(x)g(y \mid x)h(\mathrm{d}y)}{\eta\left(g(y \mid \cdot)\right)} \right|,$$

where the second equality follows from Fubini's Theorem since g, φ are bounded functions.

The first term can be bounded by

$$\left| \int_{\mathbb{X}} \int_{\mathbb{Y}} \frac{\eta_n(\mathrm{d}x)\varphi(x)g(y \mid x)h(\mathrm{d}y)}{\eta_n\left(g(y \mid \cdot)\right)\eta\left(g(y \mid \cdot)\right)} \left[\eta\left(g(y \mid \cdot)\right) - \eta_n\left(g(y \mid \cdot)\right)\right] \right|$$

$$\leq \|\varphi\|_\infty \int_{\mathbb{Y}} \frac{h(\mathrm{d}y)\int_{\mathbb{X}}\eta_n(\mathrm{d}x)g(y \mid x)}{\eta_n\left(g(y \mid \cdot)\right)\eta\left(g(y \mid \cdot)\right)} \left|\eta\left(g(y \mid \cdot)\right) - \eta_n\left(g(y \mid \cdot)\right)\right|$$

$$\leq \|\varphi\|_\infty \int_{\mathbb{Y}} \frac{h(\mathrm{d}y)}{\eta\left(g(y \mid \cdot)\right)} \left|\eta\left(g(y \mid \cdot)\right) - \eta_n\left(g(y \mid \cdot)\right)\right|.$$

Under (A1), g is bounded below by $1/m_g$ and we have

$$\eta\left(g(y \mid \cdot)\right) = \int_{\mathbb{X}} \eta(\mathrm{d}x)g(y \mid x) \geq \frac{1}{m_g}\int_{\mathbb{X}}\eta(\mathrm{d}x) = \frac{1}{m_g}\eta(\mathbb{X}) > 0$$

since $\eta \in \mathcal{M}^+(\mathbb{X})$ is an unsigned measure with nonzero mass. Therefore we obtain

$$\|\varphi\|_\infty \int_{\mathbb{Y}} \frac{h(\mathrm{d}y)}{\eta\left(g(y \mid \cdot)\right)} \left|\eta\left(g(y \mid \cdot)\right) - \eta_n\left(g(y \mid \cdot)\right)\right| \leq \|\varphi\|_\infty \frac{m_g}{\eta(\mathbb{X})} \int_{\mathbb{Y}} h(\mathrm{d}y)\left|\eta\left(g(y \mid \cdot)\right) - \eta_n\left(g(y \mid \cdot)\right)\right|$$

For fixed y, $g(y \mid \cdot) \in C_b(\mathbb{X})$, we have that

$$\left|\eta\left(g(y \mid \cdot)\right) - \eta_n\left(g(y \mid \cdot)\right)\right| \to 0$$

as $n \to \infty$ since η_n converges to η in the weak topology. Since g is uniformly bounded by m_g, the Dominated Convergence Theorem then gives

$$\int_{\mathbb{Y}} h(\mathrm{d}y)\left|\eta\left(g(y \mid \cdot)\right) - \eta_n\left(g(y \mid \cdot)\right)\right| \to 0$$

as $n \to \infty$, from which we obtain

$$\left| \int_{\mathbb{X}} \int_{\mathbb{Y}} \frac{\eta_n(\mathrm{d}x)\varphi(x)g(y \mid x)h(\mathrm{d}y)}{\eta_n\left(g(y \mid \cdot)\right)\eta\left(g(y \mid \cdot)\right)} \left[\eta\left(g(y \mid \cdot)\right) - \eta_n\left(g(y \mid \cdot)\right)\right] \right| \to 0 \qquad (28)$$

as $n \to \infty$.

For the second term, consider the function

$$x \mapsto \int_{\mathbb{Y}} \frac{\varphi(x)g(y \mid x)h(\mathrm{d}y)}{\eta\left(g(y \mid \cdot)\right)}. \qquad (29)$$

This function is bounded by $m_g^2\|\varphi\|_\infty/\eta(\mathbb{X})$; to see that it is also continuous, recall that φ, g are continuous functions while the continuity of $y \mapsto \eta\left(g(y \mid \cdot)\right)$ follows from the continuity of g and the Dominated Convergence Theorem. The Dominated Convergence theorem and the fact that g is continuous, bounded above and below give continuity of (29).

Using Fubini's Theorem, whose applicability is granted by the boundedness of g, φ, we obtain

$$\left| \int_{\mathbb{X}} \int_{\mathbb{Y}} (\eta_n(\mathrm{d}x) - \eta(\mathrm{d}x)) \frac{\varphi(x)g(y \mid x)h(\mathrm{d}y)}{\eta\left(g(y \mid \cdot)\right)} \right|$$

$$= \left| \int_{\mathbb{X}} (\eta_n(\mathrm{d}x) - \eta(\mathrm{d}x)) \int_{\mathbb{Y}} \frac{\varphi(x)g(y \mid x)h(\mathrm{d}y)}{\eta\left(g(y \mid \cdot)\right)} \right| \to 0 \qquad (30)$$

as $n \to \infty$.

Combining (28) and (30) we obtain convergence of $\mathrm{F}_{\mathrm{EM}}\,\eta_n(\varphi)$ to $\mathrm{F}_{\mathrm{EM}}\,\eta(\varphi)$ for every $\varphi \in C_b(\mathbb{X})$, and thus convergence in the weak topology of $\mathrm{F}_{\mathrm{EM}}\,\eta_n$ to $\mathrm{F}_{\mathrm{EM}}\,\eta$ (Dudley, 2002, Theorem 11.3.3) whenever η_n converges weakly to η, proving that the EM map is continuous in $\mathcal{M}^+(\mathbb{X})$. Finally, consider boundedness. A non-linear operator is bounded if and only if it maps bounded sets into bounded sets (e.g. Zeidler (1985, page 757)). The EM map maps the space of positive finite measures $\mathcal{M}^+(\mathbb{X})$ into the space of probability measures $\mathcal{P}(\mathbb{X})$, whose elements have β norm uniformly bounded by 1; in particular F_{EM} maps any bounded subset of $\mathcal{M}^+(\mathbb{X})$ into a uniformly bounded subset of $\mathcal{P}(\mathbb{X})$, showing that F_{EM} is a bounded operator. $\qquad\square$

Proposition 7. Under (A0) and (A2), the smoothing operator K defined in (27) is compact on $\mathcal{P}(\mathbb{X})$ endowed with the weak topology.

Proof. To prove that K is compact we need to prove that it maps bounded subsets into relatively compact subsets (Kress, 2014, Definition 2.17). It is sufficient to observe that $\mathbb{X}$ is a complete subset of $\mathbb{R}^{d_\mathbb{X}}$ (as it is a compact subset of a metric space) from which it follows that $\mathcal{P}(\mathbb{X})$ is complete by Prokhorov's Theorem (e.g. Dudley (2002, Corollary 11.5.5)) and therefore $\mathcal{P}(\mathbb{X})$ is relatively compact (Dudley, 2002, Theorem 11.5.4). $\qquad\square$

Corollary 2 (Compactness of F_{EMS}). Under (A0), (A1) and (A2), the EMS map F_{EMS} is compact on $\mathcal{M}^+(\mathbb{X})$ endowed with the weak topology.

Proof. The EMS map is the composition of the continuous and bounded operator F_{EM} (by Proposition 6) which maps bounded sets into bounded sets with the compact smoothing operator K (by Proposition 7) which maps bounded sets into relatively compact sets. It follows that F_{EMS} is continuous and maps bounded sets into relatively compact sets, hence F_{EMS} is compact (e.g. Zeidler (1985, page 54)). $\qquad\square$

B.2   Proof of Proposition 1

The proposition may be established straightforwardly using the technical results obtained in the previous section.

Proof. Since $\mathbb{X}$ is a compact metric space (and therefore complete), the set of probability measures $\mathcal{P}(\mathbb{X}) \subset \mathcal{M}(\mathbb{X})$ is complete by Prokhorov's Theorem (e.g. Dudley (2002, Corollary 11.5.5)) and therefore $\mathcal{P}(\mathbb{X})$ is closed. Moreover, $\mathcal{P}(\mathbb{X})$ is non-empty, bounded (since all of its elements have β norm bounded by 1) and convex: take $\mu, \nu \in \mathcal{P}(\mathbb{X})$ and $t \in [0,1]$, then for every $A \in B(\mathbb{X})$

$$t\mu(A) + (1-t)\nu(A) \geq 0 \qquad\qquad t\mu(\mathbb{X}) + (1-t)\nu(\mathbb{X}) = 1,$$

showing that $t\mu + (1-t)\nu \in \mathcal{P}(\mathbb{X})$ for all $t \in [0,1]$ and all $\mu, \nu \in \mathcal{P}(\mathbb{X})$.

These properties and the compactness of the EMS map (Corollary 2) give the existence of a fixed point by Schauder's fixed point theorem see, e.g., Zeidler (1985, Theorem 2.A). $\qquad\square$

C   Convergence of the SMC Approximation

The theoretical characterization of the particle method approximating the EMS recursion is carried out by decomposing Algorithm 1 into three steps: mutation, reweighting and resampling. This decomposition is standard in the study of SMC algorithms (Crisan and Doucet, 2002; Chopin, 2004; Míguez et al., 2013) and allows us to examine the novelty of the particle approximation introduced in Section 3 by directly considering the contribution to the overall approximation error of the use of approximate weights G_n^N.

First, consider the following decomposition of the dynamics in (8) with potentials (11) and Markov kernels (10). In the selection step, the current state is weighted according to the potential function G_n

$$\hat{\eta}_n(x_{1:n}, y_{1:n}) \equiv \Psi_{G_n}(\eta_n)(x_{1:n}, y_{1:n}) = \frac{1}{\eta_n(G_n)} G_n(x_n, y_n)\eta_n(x_{1:n}, y_{1:n});$$

then, in the mutation step, a new state is proposed according to M_{n+1}

$$\eta_{n+1}(x_{1:n+1}, y_{1:n+1}) \propto \hat{\eta}_n(x_{1:n}, y_{1:n}) M_{n+1}(x_{n+1} \mid x_n).$$

Each step of the evolution above is then compared to its particle approximation: the weighted distribution $\Psi_{G_n}(\eta_n^N)$ is compared with $\Psi_{G_n}(\eta_n)$, the resampled distribution $\hat{\eta}_n^N$ is compared with $\hat{\eta}_n$ and finally η_n^N is compared with η_n.

The proof of the $\mathbb{L}_p$-inequality in Proposition 3 follows the inductive approach of Crisan and Doucet (2002); Míguez et al. (2013) and consists of 4 Lemmata. Lemmata 2, 4 and 5 are due to Crisan and Doucet (2002); Míguez et al. (2013) and establish $\mathbb{L}_p$-error estimates for the reweighting step performed with the exact potential G_n (exact reweighting), the multinomial resampling step and the mutation step. Lemma 3 compares the exact reweighting with the reweighting obtained by using the approximated potentials G_n^N and is the main element of novelty in the proof.

In the following we commit the usual abuse of notation and we denote by η_n both a measure and its density with respect to the Lebesgue measure.

C.1 Proof of Proposition 3

Before proceeding to the proof of Proposition 3 we introduce the following auxiliary Lemma giving some properties of the approximated potentials G_n^N:

Lemma 1. Under (A0) and (A1), the approximated and exact potentials are positive functions, bounded and bounded away from 0

$$\|G_n\|_\infty \leq m_g^2 < \infty \qquad \text{and} \qquad \inf_{(x,y)} |G_n(x,y)| \geq \frac{1}{m_g^2} > 0$$

$$\|G_n^N\|_\infty \leq m_g^2 < \infty \qquad \text{and} \qquad \inf_{(x,y)} |G_n^N(x,y)| \geq \frac{1}{m_g^2} > 0.$$

We have the following decomposition

$$G_n^N(x,y) - G_n(x,y) = G_n(x,y)\frac{\eta_n|_{\mathbb{X}}(g(y\mid\cdot)) - \eta_n^N|_{\mathbb{X}}(g(y\mid\cdot))}{\eta_n^N|_{\mathbb{X}}(g(y\mid\cdot))}$$

$$= G_n^N(x,y)\frac{\eta_n|_{\mathbb{X}}(g(y\mid\cdot)) - \eta_n^N|_{\mathbb{X}}(g(y\mid\cdot))}{\eta_n|_{\mathbb{X}}(g(y\mid\cdot))}$$

for fixed $(x,y) \in \mathbb{H}$.

Proof. The boundedness of G_n and G_n^N follows from definitions (11) and (15) and the boundedness of g. The second assertion is proved by considering the relative errors between the exact and the approximated potential:

$$\frac{G_n^N(x,y) - G_n(x,y)}{G_n(x,y)} = \frac{h_n(y)}{g(y\mid x)}\left[\frac{g(y\mid x)}{h_n^N(y)} - \frac{g(y\mid x)}{h_n(y)}\right]$$

$$= h_n(y)\left[\frac{1}{h_n^N(y)} - \frac{1}{h_n(y)}\right]$$

$$= \frac{h_n(y) - h_n^N(y)}{h_n^N(y)}$$

$$= \frac{\eta_n|_{\mathbb{X}}(g(y\mid\cdot)) - \eta_n^N|_{\mathbb{X}}(g(y\mid\cdot))}{\eta_n^N|_{\mathbb{X}}(g(y\mid\cdot))}$$

and

$$\frac{G_n^N(x,y) - G_n(x,y)}{G_n^N(x,y)} = \frac{\eta_n|_{\mathbb{X}}(g(y\mid\cdot)) - \eta_n^N|_{\mathbb{X}}(g(y\mid\cdot))}{\eta_n|_{\mathbb{X}}(g(y\mid\cdot))}$$

respectively.

$\square$

Lemma 2 (Exact reweighting)**.** Assume that for any $\varphi \in \mathcal{B}_b(\mathbb{H})$ and for some $p \geq 1$

$$\mathbb{E}\left[|\eta_n^N(\varphi) - \eta_n(\varphi)|^p\right]^{1/p} \leq \widetilde{C}_{p,n}\frac{\|\varphi\|_\infty}{N^{1/2}}$$

holds for some finite constant $\widetilde{C}_{p,n}$, then

$$\mathbb{E}\left[|\Psi_{G_n}(\eta_n^N)(\varphi) - \Psi_{G_n}(\eta_n)(\varphi)|^p\right]^{1/p} \leq \bar{C}_{p,n}\frac{\|\varphi\|_\infty}{N^{1/2}}$$

for any $\varphi \in \mathcal{B}_b(\mathbb{H})$ for some finite constant $\bar{C}_{p,n}$.

Proof. The proof follows that of Crisan and Doucet (2002, Lemma 4) by exploiting the boundedness of G_n.

$\square$

Lemma 3 (Approximate reweighting). Assume that for any $\varphi \in \mathcal{B}_b(\mathbb{H})$ and for some $p \geq 1$

$$\mathbb{E}\left[|\eta_n^N(\varphi) - \eta_n(\varphi)|^p\right]^{1/p} \leq \widetilde{C}_{p,n} \frac{\|\varphi\|_\infty}{N^{1/2}}$$

holds for some finite constant $\widetilde{C}_{p,n}$, then

$$\mathbb{E}\left[|\Psi_{G_n^N}(\eta_n^N)(\varphi) - \Psi_{G_n}(\eta_n^N)(\varphi)|^p\right]^{1/p} \leq \ddot{C}_{p,n} \frac{\|\varphi\|_\infty}{N^{1/2}}$$

for any $\varphi \in \mathcal{B}_b(\mathbb{H})$ and for some finite constant $\ddot{C}_{p,n}$.

Proof. Apply the definition of Ψ_{G_n} and $\Psi_{G_n^N}$ and consider the following decomposition

$$
\begin{aligned}
|\Psi_{G_n^N}(\eta_n^N)(\varphi) - \Psi_{G_n}(\eta_n^N)(\varphi)| &= \left| \frac{\eta_n^N(G_n^N\varphi)}{\eta_n^N(G_n^N)} - \frac{\eta_n^N(G_n\varphi)}{\eta_n^N(G_n)} \right| \\
&\leq \left| \frac{\eta_n^N(G_n^N\varphi)}{\eta_n^N(G_n^N)} - \frac{\eta_n^N(G_n^N\varphi)}{\eta_n^N(G_n)} \right| \\
&\quad + \left| \frac{\eta_n^N(G_n^N\varphi)}{\eta_n^N(G_n)} - \frac{\eta_n^N(G_n\varphi)}{\eta_n^N(G_n)} \right|.
\end{aligned}
$$

Then, for the first term

$$
\begin{aligned}
\left| \frac{\eta_n^N(G_n^N\varphi)}{\eta_n^N(G_n^N)} - \frac{\eta_n^N(G_n^N\varphi)}{\eta_n^N(G_n)} \right| &= \left| \frac{\eta_n^N(G_n^N\varphi)}{\eta_n^N(G_n^N)} \right| \left| \frac{\eta_n^N(G_n) - \eta_n^N(G_n^N)}{\eta_n^N(G_n)} \right| \\
&\leq \frac{\|\varphi\|_\infty}{|\eta_n^N(G_n)|} \eta_n^N(|G_n - G_n^N|).
\end{aligned}
$$

For the second term

$$
\begin{aligned}
\left| \frac{\eta_n^N(G_n^N\varphi)}{\eta_n^N(G_n)} - \frac{\eta_n^N(G_n\varphi)}{\eta_n^N(G_n)} \right| &= \frac{1}{|\eta_n^N(G_n)|} |\eta_n^N(G_n^N\varphi) - \eta_n^N(G_n\varphi)| \\
&\leq \frac{\|\varphi\|_\infty}{|\eta_n^N(G_n)|} \eta_n^N(|G_n^N - G_n|).
\end{aligned}
$$

Hence,

$$|\Psi_{G_n^N}(\eta_n^N)(\varphi) - \Psi_{G_n}(\eta_n^N)(\varphi)| \leq 2 \frac{\|\varphi\|_\infty}{|\eta_n^N(G_n)|} \eta_n^N(|G_n^N - G_n|) \leq 2 m_g^2 \|\varphi\|_\infty \eta_n^N(|G_n^N - G_n|).$$

By applying Minkowski's inequality and the decomposition of the potentials in Lemma 1

$$
\begin{aligned}
&\mathbb{E}\left[|\eta_n^N(|G_n^N - G_n|)|^p\right]^{1/p} \\
&= \mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N |G_n^N(X_n^i, Y_n^i) - G_n(X_n^i, Y_n^i)|\right|^p\right]^{1/p} \\
&\leq \frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[|G_n^N(X_n^i, Y_n^i) - G_n(X_n^i, Y_n^i)|^p\right]^{1/p} \\
&\leq \frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[\left|\frac{G_n^N(X_n^i, Y_n^i)}{\eta_n|_{\mathbb{X}}(g(Y_n^i \mid \cdot))}\right|^p |\eta_n|_{\mathbb{X}}(g(Y_n^i \mid \cdot)) - \eta_n^N|_{\mathbb{X}}(g(Y_n^i \mid \cdot))|^p\right]^{1/p} \\
&\leq \frac{1}{N}\sum_{i=1}^N m_g^3 \mathbb{E}\left[|\eta_n|_{\mathbb{X}}(g(Y_n^i \mid \cdot)) - \eta_n^N|_{\mathbb{X}}(g(Y_n^i \mid \cdot))|^p\right]^{1/p}.
\end{aligned}
$$

Then, consider $\mathcal{S}_n^N := \sigma\left(Y_n^i : i \in \{1,\ldots,N\}\right)$, the σ-field generated by all the Y_n^i at time n. By construction, the evolution of X_n^i for $i = 1,\ldots,N$ is independent of $\mathcal{S}_n^N$ (this is due to the definition of the

mutation kernel (10)). Conditionally on $\mathcal{S}_n^N$, the Y_n^i are fixed for $i = 1, \ldots, N$ and we can use the fact that the integrals of functions from $\mathbb{X}$ to $\mathbb{R}$ with respect to η_n and $\eta_n|_{\mathbb{X}}$ coincide as do their integrals with respect to η_n^N and $\eta_n^N|_{\mathbb{X}}$, thus for fixed y:

$$\mathbb{E}\left[|\eta_n|_{\mathbb{X}}\left(g(y \mid \cdot)\right) - \eta_n^N|_{\mathbb{X}}\left(g(y \mid \cdot)\right)|^p\right]^{1/p} = \mathbb{E}\left[|\eta_n\left(g(y \mid \cdot)\right) - \eta_n^N\left(g(y \mid \cdot)\right)|^p\right]^{1/p}$$
$$\leq \frac{m_g \widetilde{C}_{p,n}}{N^{1/2}}$$

where the last inequality follows from the hypothesis of the Lemma because $g(y \mid \cdot)$ is a bounded and measurable function for all fixed $y \in \mathbb{Y}$.

Hence, since Y_n^i is $\mathcal{S}_n^N$-measurable and independent of $\eta_n^N|_{\mathbb{X}}$, we have

$$\mathbb{E}\left[|\eta_n^N(|G_n^N - G_n|)|^p\right]^{1/p} \leq m_g^3 \frac{1}{N} \sum_{i=1}^N \mathbb{E}\left[|\eta_n|_{\mathbb{X}}\left(g(Y_n^i \mid \cdot)\right) - \eta_n^N|_{\mathbb{X}}\left(g(Y_n^i \mid \cdot)\right)|^p\right]^{1/p}$$
$$\leq m_g^3 \frac{1}{N} \sum_{i=1}^N \mathbb{E}\left[\mathbb{E}\left[|\eta_n|_{\mathbb{X}}\left(g(Y_n^i \mid \cdot)\right) - \eta_n^N|_{\mathbb{X}}\left(g(Y_n^i \mid \cdot)\right)|^p|\mathcal{S}_n^N\right]\right]^{1/p}$$
$$\leq \frac{m_g^4 \widetilde{C}_{p,n}}{N^{1/2}}.$$

Therefore,

$$\mathbb{E}\left[|\Psi_{G_n^N}(\eta_n^N)(\varphi) - \Psi_{G_n}(\eta_n^N)(\varphi)|^p\right]^{1/p} \leq 2\widetilde{C}_{p,n} m_g^6 \frac{\|\varphi\|_\infty}{N^{1/2}},$$

with the constant $\ddot{C}_{p,n} = 2\widetilde{C}_{p,n} m_g^6$. $\qquad\square$

Lemma 4 (Multinomial resampling). Assume that for any $\varphi \in \mathcal{B}_b(\mathbb{H})$ and for some $p \geq 1$

$$\mathbb{E}\left[|\Psi_{G_n^N}(\eta_n^N)(\varphi) - \hat{\eta}_n(\varphi)|^p\right]^{1/p} = \mathbb{E}\left[|\Psi_{G_n^N}(\eta_n^N)(\varphi) - \Psi_{G_n}(\eta_n)(\varphi)|^p\right]^{1/p} \leq \widehat{C}_{p,n} \frac{\|\varphi\|_\infty}{N^{1/2}}$$

holds for some finite constant $\widehat{C}_{p,n}$, then after the resampling step performed through multinomial re-sampling

$$\mathbb{E}\left[|\hat{\eta}_n^N(\varphi) - \hat{\eta}_n(\varphi)|^p\right]^{1/p} \leq C_{p,n} \frac{\|\varphi\|_\infty}{N^{1/2}}$$

for any $\varphi \in \mathcal{B}_b(\mathbb{H})$ for some finite constant $C_{p,n}$.

Proof. The proof follows that of Crisan and Doucet (2002, Lemma 5) using the Marcinkiewicz-Zygmund type inequality in Del Moral (2004, Lemma 7.3.3) and the hypothesis.

$\qquad\square$

Lemma 5 (Mutation). Assume that for any $\varphi \in \mathcal{B}_b(\mathbb{H})$ and for some $p \geq 1$

$$\mathbb{E}\left[|\hat{\eta}_n^N(\varphi) - \hat{\eta}_n(\varphi)|^p\right]^{1/p} \leq C_{p,n} \frac{\|\varphi\|_\infty}{N^{1/2}}$$

holds for some finite constant $C_{p,n}$, then, after the mutation step

$$\mathbb{E}\left[|\eta_{n+1}^N(\varphi) - \eta_{n+1}(\varphi)|^p\right]^{1/p} \leq \widetilde{C}_{p,n+1} \frac{\|\varphi\|_\infty}{N^{1/2}}$$

for any $\varphi \in \mathcal{B}_b(\mathbb{H})$ for some finite constant $\widetilde{C}_{p,n+1}$.

Proof. The proof follows that of Crisan and Doucet (2002, Lemma 3), where after applying Minkowski's inequality

$$\mathbb{E}\left[|\eta_{n+1}^N(\varphi) - \eta_{n+1}(\varphi)|^p\right]^{1/p} = \mathbb{E}\left[|\eta_{n+1}^N(\varphi) - \hat{\eta}_n M_{n+1}(\varphi)|^p\right]^{1/p}$$
$$\leq \mathbb{E}\left[|\eta_{n+1}^N(\varphi) - \hat{\eta}_n^N M_{n+1}(\varphi)|^p\right]^{1/p}$$
$$+ \mathbb{E}\left[|\hat{\eta}_n^N M_{n+1}(\varphi) - \hat{\eta}_n M_{n+1}(\varphi)|^p\right]^{1/p},$$

we can bound the first term with the Marcinkiewicz-Zygmund type inequality in Del Moral (2004, Lemma 7.3.3) and the second term with the hypothesis.

$\qquad\square$

The proof of the $\mathbb{L}_p$-inequality in Proposition 3 is based on an inductive argument which uses Lemmata 2-5:

Proof of Proposition 3. At time $n = 1$, the particles $(X_1^i, Y_1^i)_{i=1}^N$ are sampled i.i.d. from $\eta_1 \equiv \hat{\eta}_1$ thus $\mathbb{E}\left[\varphi(X_1^i, Y_1^i)\right] = \eta_1(\varphi)$ for $i = 1, \ldots, N$. We can define the sequence of functions $\Delta_1^i : \mathbb{X} \times \mathbb{Y} \mapsto \mathbb{R}$ for $i = 1, \ldots, N$

$$\Delta_1^i(x, y) := \varphi(x, y) - \mathbb{E}\left[\varphi(X_1^i, Y_1^i)\right]$$

so that,

$$\eta_1^N(\varphi) - \eta_1(\varphi) = \frac{1}{N}\sum_{i=1}^N \Delta_1^i(X_1^i, Y_1^i),$$

and apply Lemma 7.3.3 in Del Moral (2004) to get

$$\mathbb{E}\left[|\eta_1^N(\varphi) - \eta_1(\varphi)|^p\right]^{1/p} \le 2b(p)^{1/p}\frac{\|\varphi\|_\infty}{N^{1/2}},$$

with $b(p) < \infty$, for every $p \ge 1$.

Then, assume that the result holds at time n: for every $\varphi \in \mathcal{B}_b(\mathbb{H})$, every $p \ge 1$ and some finite constant $\widetilde{C}_{p,n}$

$$\mathbb{E}\left[|\eta_n^N(\varphi) - \eta_n(\varphi)|^p\right]^{1/p} \le \widetilde{C}_{p,n}\frac{\|\varphi\|_\infty}{N^{1/2}}.$$

The $\mathbb{L}_p$-inequality in (19) is obtained by combining the results of Lemma 2 and Lemma 3 using Minkowski's inequality

$$\mathbb{E}\left[|\Psi_{G_n^N}(\eta_n^N)(\varphi) - \Psi_{G_n}(\eta_n)(\varphi)|^p\right]^{1/p} \le (\bar{C}_{p,n} + \ddot{C}_{p,n})\frac{\|\varphi\|_\infty}{N^{1/2}}$$

for every $\varphi \in \mathcal{B}_b(\mathbb{H})$ and some finite constants $\bar{C}_{p,n}, \ddot{C}_{p,n}$. Thus, $\widehat{C}_{p,n} = \bar{C}_{p,n} + \ddot{C}_{p,n}$.

Lemma 4 gives

$$\mathbb{E}\left[|\hat{\eta}_n^N(\varphi) - \hat{\eta}_n(\varphi)|^p\right]^{1/p} \le C_{p,n}\frac{\|\varphi\|_\infty}{N^{1/2}}$$

for every $\varphi \in \mathcal{B}_b(\mathbb{H})$ and some finite constants $C_{p,n}$, and Lemma 5 gives

$$\mathbb{E}\left[|\eta_{n+1}^N(\varphi) - \eta_{n+1}(\varphi)|^p\right]^{1/p} \le \widetilde{C}_{p,n+1}\frac{\|\varphi\|_\infty}{N^{1/2}}$$

for every $\varphi \in \mathcal{B}_b(\mathbb{H})$ and some finite constant $\widetilde{C}_{p,n+1}$.

The result follows for all $n \in \mathbb{N}$ by induction. $\square$

C.2 Proof of Proposition 4

Using standard techniques following Dudley (2002, Chapter 11, Theorem 11.4.1) and Berti et al. (2006) and given in detail for the context of interest by Schmon et al. (2021, Theorem 4), the result of Corollary 1 can be strengthened to the convergence of the measures in the weak topology.

Proof of Proposition 4. Consider $\mathrm{BL}(\mathbb{H}) \subset \mathcal{B}_b(\mathbb{H})$, the Banach space of bounded Lipschitz functions. As shown in Dudley (2002, Theorem 11.4.1), see also Schmon et al. (2021, Proposition 5) for a more accessible presentation, $\mathrm{BL}(\mathbb{H})$ admits a countable dense subclass $C \subset \mathrm{BL}(\mathbb{H})$.

For every $\varphi \in C$ define $A_\varphi := \{\omega \in \Omega : \eta_n^N(\omega)(\varphi) \to \eta_n(\varphi) \ N \to \infty\}$. Then $\mathbb{P}(A_\varphi) = 1 \ \forall \varphi \in C$ by Corollary 1 and

$$\mathbb{P}\left(\{\omega \in \Omega : \eta_n^N(\omega)(\varphi) \to \eta_n(\varphi) \ N \to \infty \ \forall \varphi \in C\}\right) = \mathbb{P}\left(\bigcap_{\varphi \in C} A_\varphi\right) = 1.$$

The result follows from the fact that C is dense in $\mathrm{BL}(\mathbb{H})$ and the Portmanteau Theorem (e.g. Dudley (2002, Theorem 11.1.1)). $\square$

D Convergence of Density Estimates

D.1 Auxiliary Results

Using a version of the Dominated Convergence Theorem for weakly converging measures (Serfozo, 1982; Feinberg et al., 2020), standard results on kernel density estimation (e.g. Parzen (1962); Cacoullos (1966)) and an argument based on compactness as in Newey (1991) we can establish the following result

Proposition 8. Under (A0), (A1) and (A2), if $s_N \to 0$ as $N \to \infty$, the estimator $f_{n+1}^N(x)$ in (17) converges uniformly to $f_{n+1}(x)$ with probability 1 for all $n \geq 1$.

Proof. Let us define for $N \in \mathbb{N}$

$$\varphi^N(t,x) := \int_{\mathbb{X}} K(x',x) s_N^{-d_{\mathbb{X}}} |\Sigma|^{-1/2} S\left((s_N^2 \Sigma)^{-1/2}(t-x')\right) dx',$$

and note that the estimator (17) is given by $f_{n+1}^N(x) = \Psi_{G_n^N}(\eta_n^N)\left(\varphi^N(\cdot,x)\right)$ for any fixed $x \in \mathbb{X}$. Standard results in the literature on kernel density estimation show that $\varphi^N(\cdot,x)$ converges to $K(\cdot,x)$ pointwise for all $x \in \mathbb{X}$ (e.g. Cacoullos (1966, Theorem 2.1)). Because $\mathbb{X}$ is compact, Assumption (A2) ensures that K is *uniformly* continuous on $\mathbb{X}$ (e.g. Rudin (1964, Theorem 4.19)), then, as argued in Parzen (1962, Theorem 3.A), the sequence $\varphi^N(\cdot,x)$ converges uniformly to $K(\cdot,x)$ in $\mathbb{X}$ (see also Cacoullos (1966, Theorem 3.3)). As a consequence, the sequence $\{\varphi^N(\cdot,x)\}_{N \in \mathbb{N}}$ is uniformly equicontinuous and uniformly bounded (e.g. Rudin (1964, Theorem 7.25)). It follows that $\{\varphi^N(\cdot,x)\}_{N \in \mathbb{N}}$ is (asymptotically) uniformly integrable in the sense of Feinberg et al. (2020, Definition 2.6).

Using an argument analogous to that in Proposition 4 we can establish that $\Psi_{G_n^N}(\eta_n^N)$ converges to $\Psi_{G_n}(\eta_n)$ almost surely in the weak topology, then using the fact that the sequence $\{\varphi^N(\cdot,x)\}_{N \in \mathbb{N}}$ is asymptotically uniformly integrable and equicontinuous with continuous limit $K(\cdot,x)$, the Dominated Convergence theorem for weakly converging measures (Feinberg et al. (2020, Corollary 5.2); see also Serfozo (1982, Theorem 3.3)) implies that

$$f_{n+1}^N(x) = \Psi_{G_n^N}(\eta_n^N)\left(\varphi^N(\cdot,x)\right) \to \Psi_{G_n}(\eta_n)\left(K(\cdot,x)\right) = f_{n+1}(x) \tag{31}$$

almost surely as $N \to \infty$ for any fixed $x \in \mathbb{X}$.

To turn the result above into almost sure uniform convergence, i.e.

$$\mathbb{P}\left(\limsup_{N \to \infty}\left\{\sup_{x \in \mathbb{X}} |f_{n+1}^N(x) - f_{n+1}(x)| > \varepsilon\right\}\right) = 0$$

for every $\varepsilon > 0$, we exploit assumption (A0) and the resulting continuity properties of K.

Under (A0)–(A2), K is uniformly continuous and we have that for any $\varepsilon > 0$, there exists some $\delta_\varepsilon > 0$ such that

$$|f_{n+1}(x) - f_{n+1}(x')| = |\Psi_{G_n}(\eta_n)\left(K(\cdot,x) - K(\cdot,x')\right)|$$
$$\leq \sup_{z \in \mathbb{X}} |K(z,x) - K(z,x')| \leq \frac{\varepsilon}{3}$$

whenever $\|x - x'\|_2 < \delta_\varepsilon$. Using the definition of φ^N and exploiting again the uniform continuity of K we also have that for every $\varepsilon > 0$

$$|\varphi^N(t,x) - \varphi^N(t,x')| \leq \int_{\mathbb{X}} |K(u,x) - K(u,x')| s_N^{-d_{\mathbb{X}}} |\Sigma|^{-1/2} S\left((s_N^2 \Sigma)^{-1/2}(t-u)\right) du$$
$$\leq \frac{\varepsilon}{3} \int_{\mathbb{X}} s_N^{-d_{\mathbb{X}}} |\Sigma|^{-1/2} S\left((s_N^2 \Sigma)^{-1/2}(t-u)\right) du \leq \frac{\varepsilon}{3}$$

if $\|x - x'\|_2 < \delta_\varepsilon$. It follows that f_{n+1}^N is uniformly continuous: for any $\varepsilon > 0$

$$|f_{n+1}^N(x) - f_{n+1}^N(x')| = |\Psi_{G_n^N}(\eta_n^N)\left(\varphi^N(\cdot,x) - \varphi^N(\cdot,x')\right)|$$
$$\leq \sup_{z \in \mathbb{X}} |\varphi^N(z,x) - \varphi^N(z,x')| \leq \frac{\varepsilon}{3}$$

27

whenever $\|x - x'\|_2 < \delta_\varepsilon$.

Let $B(x, \delta_\varepsilon) := \{x' \in \mathbb{X} : \|x - x'\|_2 < \delta_\varepsilon\}$ denote the ball in $\mathbb{X}$ centred around x of radius δ_ε. Under (A0), $\mathbb{X}$ is compact and therefore there exists a finite subcover $\{B(x^j)\}_{j=1}^J$ of $\{B(x, \delta_\varepsilon)\}_{x \in \mathbb{X}}$. Using the uniform continuity above and the following decomposition, we obtain for all $x \in B(x^j)$, $j = 1, \ldots, J$ and for all N,

$$
\begin{aligned}
|f_{n+1}^N(x) - f_{n+1}(x)| &\leq |f_{n+1}^N(x) - f_{n+1}^N(x^j)| + |f_{n+1}^N(x^j) - f_{n+1}(x^j)| + |f_{n+1}(x^j) - f_{n+1}(x)| \\
&\leq \frac{\varepsilon}{3} + |f_{n+1}^N(x^j) - f_{n+1}(x^j)| + \frac{\varepsilon}{3} \\
&\leq \frac{2}{3}\varepsilon + \max_{j=1,\ldots,J} |f_{n+1}^N(x^j) - f_{n+1}(x^j)|,
\end{aligned}
$$

from which follows

$$
\sup_{x \in \mathbb{X}} |f_{n+1}^N(x) - f_{n+1}(x)| \leq \frac{2}{3}\varepsilon + \max_{j=1,\ldots,J} |f_{n+1}^N(x^j) - f_{n+1}(x^j)|.
$$

Therefore, to obtain almost sure uniform convergence, it is sufficient to show that

$$
\mathbb{P}\left(\left\{\omega \in \Omega : \max_{j=1,\ldots,J} |f_{n+1}^N(\omega)(x^j) - f_{n+1}(x^j)| \to 0 \ N \to \infty\right\}\right) = 1.
$$

Let us define $A_j := \{\omega \in \Omega : f_{n+1}^N(\omega)(x^j) \to f_{n+1}(x^j) \ N \to \infty\}$. As a consequence of (31) we have $\mathbb{P}(A_j) = 1$ for all $j = 1, \ldots, J$ and

$$
\mathbb{P}\left(\left\{\omega \in \Omega : \max_{j=1,\ldots,J} |f_{n+1}^N(\omega)(x^j) - f_{n+1}(x^j)| \to 0 \ N \to \infty\right\}\right) = \mathbb{P}\left(\bigcap_{j=1,\ldots,J} A_j\right) = 1,
$$

which gives the result. $\qquad\square$

D.2 Proof of Proposition 5

Proof. A direct consequence of Proposition 8 is the almost sure pointwise convergence of f_{n+1}^N to f_{n+1}. As both $f_{n+1}^N(x)$ and $f_{n+1}(x)$ are probability densities on $\mathbb{X}$, we can extend them to $\mathbb{R}^{d_{\mathbb{X}}}$ by taking

$$
\psi_{n+1}^N(x) := \begin{cases} f_{n+1}^N(x) & x \in \mathbb{X} \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad \psi_{n+1}(x) := \begin{cases} f_{n+1}(x) & x \in \mathbb{X} \\ 0 & \text{otherwise} \end{cases}
$$

respectively. Both $\psi_{n+1}^N(x)$ and $\psi_{n+1}(x)$ are probability densities on $\mathbb{R}^{d_{\mathbb{X}}}$ and are measurable functions. Moreover, $\psi_{n+1}^N(x)$ converges almost surely to $\psi_{n+1}(x)$ for all $x \in \mathbb{R}^{d_{\mathbb{X}}}$. Hence, we can apply Glick's extension to Scheffé's Lemma (e.g. Devroye and Wagner (1979)) to obtain

$$
\int_{\mathbb{R}^d} |\psi_{n+1}^N(x) - \psi_{n+1}(x)| \, \mathrm{d}x \overset{\text{a.s.}}{\to} 0
$$

from which we can conclude

$$
\int_{\mathbb{X}} |f_{n+1}^N(x) - f_{n+1}(x)| \, \mathrm{d}x = \int_{\mathbb{X}} |\psi_{n+1}^N(x) - \psi_{n+1}(x)| \, \mathrm{d}x \leq \int_{\mathbb{R}^{d_{\mathbb{X}}}} |\psi_{n+1}^N(x) - \psi_{n+1}(x)| \, \mathrm{d}x \to 0
$$

almost surely as $N \to \infty$.

Convergence of the MISE is a consequence of Proposition 8, (A0) and the Dominated Convergence Theorem

$$
\mathbb{E}\left[\int_{\mathbb{X}} |f_{n+1}^N(x) - f_{n+1}(x)|^2 \mathrm{d}x\right] \leq \lambda(\mathbb{X}) \, \mathbb{E}\left[\|f_{n+1}^N - f_{n+1}\|_\infty^2\right] \mathrm{d}x \to 0
$$

as $N \to \infty$, where $\lambda(\mathbb{X}) < \infty$ denotes the Lebesgue measure of $\mathbb{X}$.
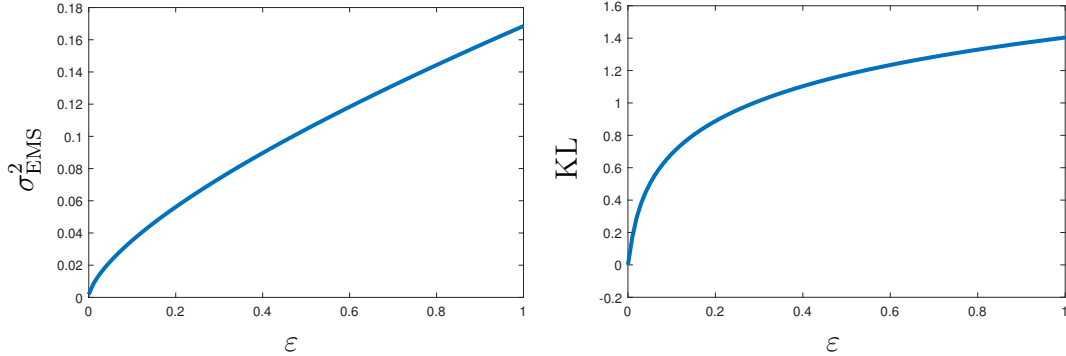
$\qquad\square$

Figure 3: Functional dependence of the variance of the resulting approximation σ^2_{EMS} (left) and the Kullback–Leibler divergence (33) (right) on the smoothing parameter ε.

E Additional Examples

E.1 Analytically tractable example

Here we consider a toy example involving Gaussian densities for which both the EM recursion (2) and the EMS recursion (6) can be solved at least implicitly. The Fredholm integral equation we consider is

$$\mathcal{N}(y; \mu, \sigma_f^2 + \sigma_g^2) = \int_{\mathbb{X}} \mathcal{N}(x; \mu, \sigma_f^2) \mathcal{N}(y; x, \sigma_g^2) \, \mathrm{d}x, \qquad y \in \mathbb{Y}$$

where $\mathbb{X} = \mathbb{Y} = \mathbb{R}$. The initial distribution $f_1(x)$ is $\mathcal{N}(x; \mu, \sigma^2_{\text{EMS},1})$ for some $\sigma^2_{\text{EMS},1} > 0$.

The fixed point f_{EMS} of the EMS recursion (6) with Gaussian smoothing kernel $K(x', x) = \mathcal{N}(x; x', \varepsilon^2)$ is a Gaussian with mean μ and variance σ^2_{EMS} solving

$$\sigma^6_{\text{EMS}} + \sigma^4_{\text{EMS}}(\sigma_g^2 - \sigma_h^2) - 2\sigma^2_{\text{EMS}}\varepsilon^2\sigma_g^2 - 2\varepsilon^2\sigma_g^2 = 0. \tag{32}$$

We can compute the Kullback–Leibler divergence achieved by f_{EMS}:

$$\text{KL}\left(h, \int_{\mathbb{X}} f_{EMS}(x)g(y \mid \cdot) \, \mathrm{d}x\right) = \frac{1}{2}\log\frac{\sigma^2_{\text{EMS}} + \sigma_g^2}{\sigma_h^2} + \frac{\sigma_h^2}{2(\sigma^2_{\text{EMS}} + \sigma_g^2)} - \frac{1}{2}, \tag{33}$$

as $\int_{\mathbb{X}} f_{EMS}(x)g(y \mid \cdot) \, \mathrm{d}x$ is the Gaussian density $\mathcal{N}(y; \mu, \sigma^2_{\text{EMS}} + \sigma_g^2)$. The fixed point for the EM recursion (2) is obtained setting $\varepsilon = 0$. The corresponding value of the Kullback–Leibler divergence is 0. Figure 3 shows the dependence of σ^2_{EMS} and of the KL divergence on ε.

The conjugacy properties of this model allow us to obtain an exact form for the potential (11)

$$G_n(x_n, y_n) = \frac{g(y_n \mid x_n)}{h_n(y_n)} = \frac{\mathcal{N}(y_n; x_n, \sigma_g^2)}{\mathcal{N}(y_n; \mu, \sigma_g^2 + \sigma^2_{\text{EMS},n})} \tag{34}$$

where $\sigma^2_{\text{EMS},n}$ is the variance of $f_n(x)$.

We use this example to show that the maximum likelihood estimator obtained with the EM iteration (4) does not enjoy good properties, and to motivate the addition of a smoothing step in the iterative process (Figure 4).

Taking $\sigma_f^2 = 0.043^2$ and $\sigma_g^2 = 0.045^2$ we have $|1 - \int_0^1 f(x)\mathrm{d}x| < 10^{-30}$, thus we can restrict our attention to $[0, 1]$ and implement the discretized EM and EMS by taking $B = D = 100$ equally spaced intervals in this interval. The number of iterations $n = 100$ is fixed for EM, EMS and SMC. The number of particles for SMC is $N = 10^4$ and $\varepsilon = 10^{-2}$. The smoothing matrix for EMS is obtained by discretization of the smoothing kernel $K(x', x) = \mathcal{N}(x; x', \varepsilon^2)$.

Figure 4 clearly shows that the EM estimate, despite identifying the correct support of the solution, cannot recover the correct shape and is not smooth. On the contrary, both EMS and SMC give good reconstruction of f while preserving smoothness.
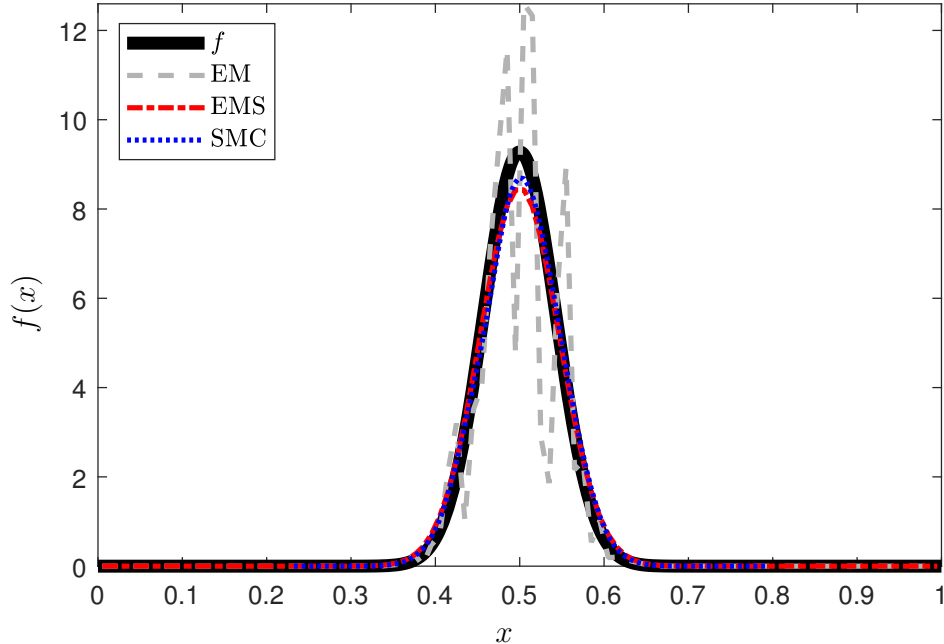
Figure 4: Comparison of EM, EMS and SMC with exact potential G_n for the analytically tractable example.

Then we compare the deterministic discretization (5) of the EMS recursion (6) with the stochastic one given by SMC with the exact potential (34). To do so, we consider the variance of the obtained reconstructions, their integrated square error (23), the mean integrated square error for between h and

$$\hat{h}_{n+1}^N(y) = \int_{\mathbb{X}} f_{n+1}^N(x)g(y \mid x)\mathrm{d}x$$

and the Kullback–Leibler divergence $\mathrm{KL}(h, \hat{h}_{n+1}^N)$ (restricting to the $[0, 1]$ interval and computing by numerical integration) as the value of the smoothing parameter ε increases (Figure 5). We consider one run of discretized EMS and compare it with 1,000 repetitions of SMC for each value of ε (this choice follows from the fact that discretized EMS is a deterministic algorithm). The number of particles for SMC is $N = 10^3$ and for each run we draw a sample $\mathbf{Y}$ of size 10^4 from h and resample from it $M = \min(N, 10^4)$ particles in line 2 of Algorithm 1. Both algorithms correctly identify the mean for every value of ε while the estimated variances increase from that obtained with the EM algorithm ($\varepsilon = 0$) to the variance of a Uniform distribution over $[0, 1]$ (Figure 5 top left). Unsurprisingly, the ISE for both f_{n+1}^N and h_{n+1}^N increases with ε (Figure 5 top right and bottom left), showing that an excessive amount of smoothing leads to poor reconstructions. In particular for values of $\varepsilon \geq 0.5$ the reconstructions of f become flatter and tend to coincide with a Uniform distribution in the case of EMS and with a normal distribution centered at μ and with high variance (≥ 0.08) in the case of SMC. This difference reflects in the behavior of the Kullback–Leibler divergence, which stabilizes around 133 for EMS while keeps increasing for SMC (Figure 5 bottom right).

We now consider the effect of the use of the approximated potentials G_n^N in place of the exact ones G_n in the SMC scheme. We compare the ISE for f_{n+1}^N given by the SMC scheme with exact and approximated potentials for values of the number M of samples Y_n^{ij} drawn from h at each time step between 1 and 10^3 with 1,000 repetitions for each M. Through this comparison we also address the computational complexity $O(MN)$ of the algorithm, with focus on the choice of the value of M. Figure 6 shows the results for $N = 10^3$ and $\varepsilon = 10^{-2}$. The behavior for different values of N and ε is similar. The plot of $\mathrm{ISE}(f_{n+1}^N)$ shows a significant improvement when $M > 1$ but little further improvement for $M > 10$.

To further investigate the choice of M we compare the reconstructions obtained using the exact and the approximated potentials for $M = 10$, $M = 10^2$ and $M = N = 10^3$. Figure 7 shows pointwise means and pointwise MSE (24) for 1,000 reconstructions. The means of the reconstructions with the exact potentials (blue) coincide for the three values of M, the means of the reconstructions with the approximated potentials (red) also coincide but have heavier tails than those obtained with the exact
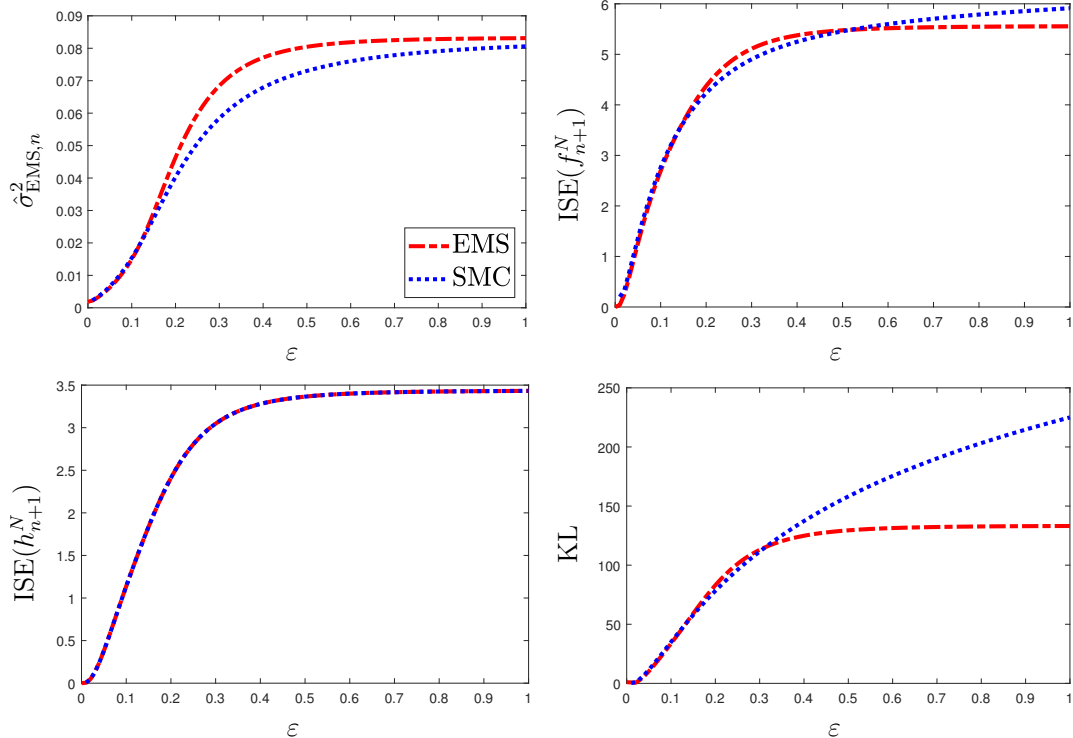
Figure 5: Estimated variance (top left), $\mathrm{ISE}(f_{n+1}^N)$ (top right), $\mathrm{ISE}(h_{n+1}^N)$ (bottom left) and Kullback–Leibler divergence (bottom right) as functions of the smoothing parameter ε for the analytically tractable example. The deterministic discretization (5) (red) and the stochastic discretization via SMC with the exact potentials (34) (blue) are compared.

potentials. The MSE is similar for reconstructions with exact and approximated potentials with the same value of M. In particular, the little improvement of the MSE from $M = 10^2$ to $M = 10^3$ suggests that $M = 10^2$ could be used instead of $M = N = 10^3$ if the computational resources are limited. Using $M = 10^2$ instead of $M = 10^3$ reduces the average runtime by $\approx 80\%$ for both the algorithm using the exact potentials and that using the approximated potentials.

Silverman et al. (1990, Section 5.4) conjectured that under suitable assumptions the EMS map (6) has a unique fixed point. This conjecture is empirically confirmed by the results in Figure 8. We run EM, EMS and SMC with approximated potentials for $n = 100$ iterations starting from three initial distributions $f_1(x)$: a Uniform on $[0, 1]$, a Dirac δ centered at 0.5 and the solution $\mathcal{N}(x; \mu, \sigma_f^2)$. The number of particles is set to $N = 10^3$ and the smoothing parameter $\varepsilon = 10^{-1}$. Both EMS and SMC converge to the same value of the Kullback–Leibler divergence regardless of the starting distribution. The speed of convergence of the three algorithms is similar, in each case little further change is observed once 4 iterations have occurred.
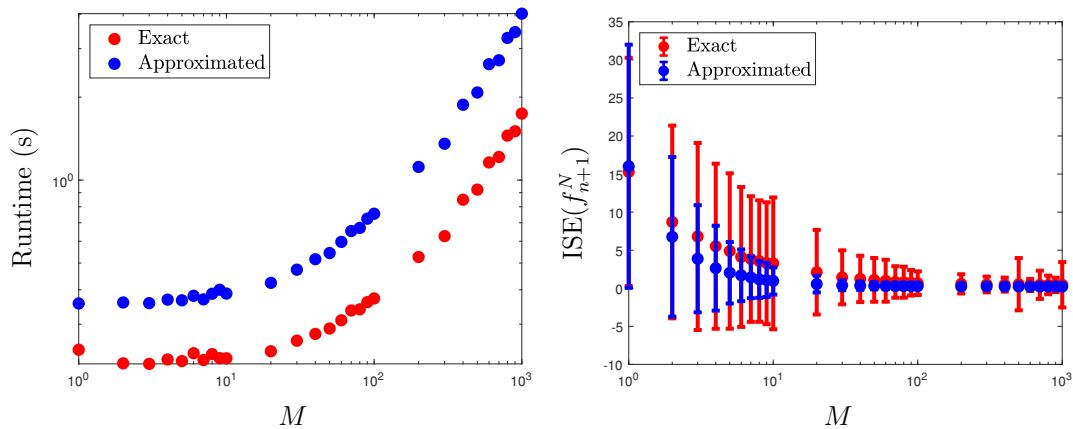
31

Figure 6: Dependence of runtime and ISE(f_{n+1}^N) on the value of M, the number of samples drawn from h at each iteration, for the SMC scheme run with the exact potential (blue) and the approximated potential (red). The error bars represent twice the standard deviation of ISE(f_{n+1}^N).
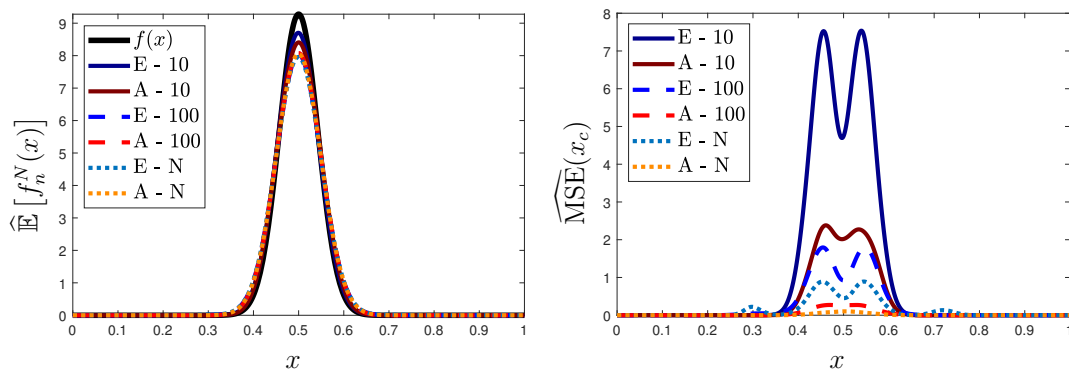

Figure 7: Reconstruction of $f(x) = \mathcal{N}(x; 0.5, 0.043^2)$ from data distribution $h(y) = \mathcal{N}(y; 0.5, 0.043^2 + 0.045^2)$. The number of particles N is 10^3 and the smoothing parameter $\varepsilon = 10^{-2}$. $M = 10$, $M = 10^2$ and $M = N$ are compared through the pointwise means of the reconstructions and the pointwise mean squared error (MSE).
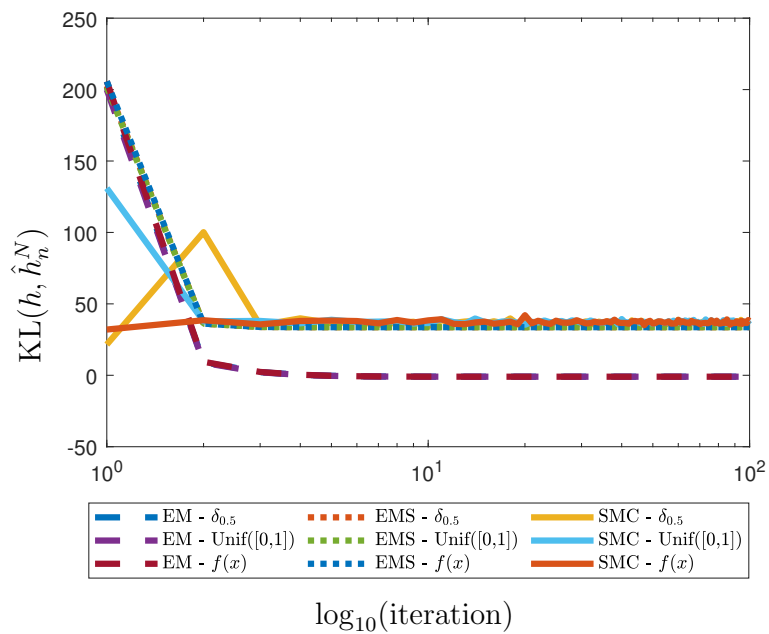
Figure 8: Kullback–Leibler divergence between h and $\hat{h}_n^N$ as function of the number of iterations. Three starting distributions are considered: Uniform$([0, 1])$, $\delta_{0.5}$, $\mathcal{N}(x; \mu, \sigma_f^2)$. The behavior of EM (dashed lines), EMS (dotted lines) and SMC (solid lines) is compared.

E.2 Motion deblurring

Consider a simple example of motion deblurring where the observed picture h is obtained while the object of interest is moving with constant speed b in the horizontal direction (Vardi and Lee, 1993; Lee and Vardi, 1994). The constant motion in the horizontal direction is modeled by multiplying the density of a uniform random variable on $[-b/2, b/2]$ describing the motion in the horizontal direction and a Gaussian, $\mathcal{N}(v; y, \sigma^2)$, with small variance, $\sigma^2 = 0.02^2$, describing the relative lack of motion in the vertical direction

$$g(u, v \mid x, y) = \mathcal{N}(v; y, \sigma^2) \mathrm{Uniform}_{[x-b/2, x+b/2]}(u).$$

We obtain the corrupted image in Figure 9b from the reference image in Figure 9a using the model above with constant speed $b = 128$ pixels and adding multiplicative noise as in Lee and Vardi (1994, Section 6.2). Figure 9b is a noisy discretization of the unknown $h(u, v)$ on a 300×600 grid. The addition of multiplicative noise makes the model (1) misspecified, but still suitable to describe the deconvolution problem when the amount of noise is low. For higher levels of noise, the noise itself should be taken into account when modeling the generation of the data corresponding to h.

We compare the reconstruction obtained using the SMC scheme with that given by the `deconvlucy` function in MATLAB$^{\copyright}$ (The MathWorks Inc., 1993), an efficient implementation of the Richardson–Lucy algorithm (i.e. EM for Poisson counts) for image processing which considers the data image as a discretization of the unknown density h into bins. The same image is used to draw the samples necessary for the SMC implementation.

The smoothing parameter is $\varepsilon = 10^{-3}$, and the number of particles is $N = 5,000$. These values are chosen to achieve a trade-off between smoothing and accuracy of the reconstruction and to keep the runtime under three minutes on a standard laptop.

The distance between the reconstructions and the original image is evaluated using both the ISE (23) and the match distance, i.e. the $\mathbb{L}_1$ norm of the cumulative histogram of the image, a special case of the Earth Mover's Distance for gray-scale images (Rubner et al., 2000). SMC gives visibly smoother images and is better at recovering the shape of the original image (ISE(f_{n+1}^N) is 1.4617 for SMC and 2.0863 for RL). In contrast, the RL algorithm performs better in terms of match distance (0.0054 for RL and 0.0346 for SMC).

(a) Original sharp image

(b) Blurred image with 0.5% multiplicative noise

(c) Reconstruction with RL
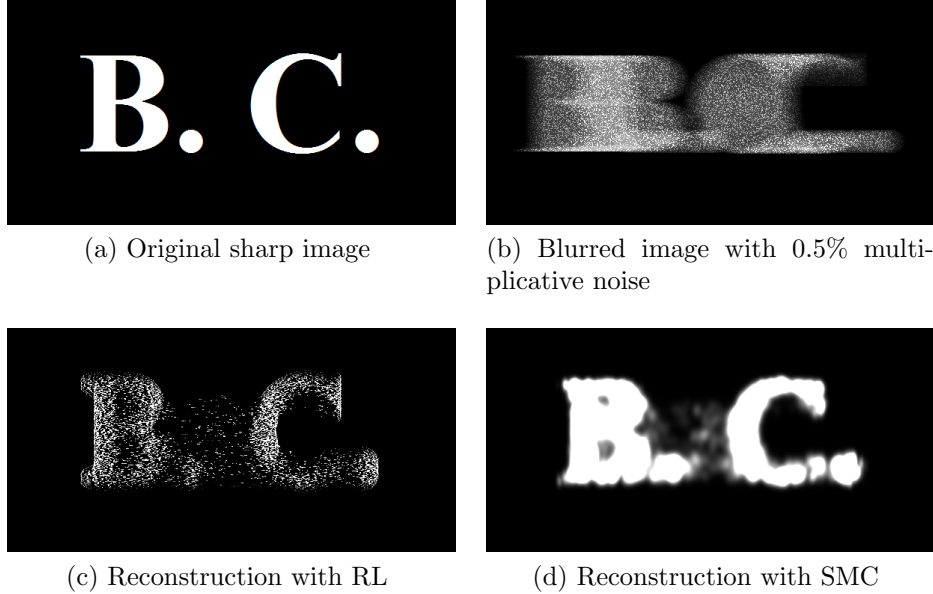

(d) Reconstruction with SMC

Figure 9: Reference image, blurred noisy data distribution and reconstructions for the motion deblurring example. Each scheme used 100 iterations; the SMC scheme used $N = 5,000$ particles.

F   Additional Results for PET Example

The reconstruction of cross-sectional images from projections given by PET scanners is modeled by the Radon transform (Radon, 1986)

$$h(\phi, \xi) = \int_{-\infty}^{+\infty} f(\xi \cos \phi - t \sin \phi, \xi \sin \phi + t \cos \phi) \mathrm{d}t, \tag{35}$$

for $(\phi, \xi) \in \mathbb{Y} = [0, 2\pi] \times [-R, R]$, where the right hand side is the line integral along the line with equation $x \cos \phi + y \sin \phi = \xi$. We rewrite (35) as a Fredholm integral equation (1) modelling the alignment between the projections onto (ϕ, ξ) and the corresponding location (x, y) in the reference image using a Gaussian distribution with small variance (in the experiments we use $\sigma^2 = 0.02^2$)

$$h(\phi, \xi) = \int_{\mathbb{X}} \mathcal{N}\left(x \cos(\phi) + y \sin(\phi) - \xi; 0, \sigma^2\right) f(x, y) \mathrm{d}x \mathrm{d}y,$$

where $\mathbb{X} = [-r, r]^2$. The kernel $g(\phi, \xi \mid x, y) = \mathcal{N}\left(x \cos(\phi) + y \sin(\phi) - \xi; 0, \sigma^2\right)$ is not a Markov kernel (in the sense that it does not integrate to 1 for fixed (x, y)), however, we can use the re-normalization described in Chae et al. (2018, Section 6) to obtain the Markov kernel

$$\tilde{g}(\phi, \xi \mid x, y) = \frac{g(\phi, \xi \mid x, y)}{C(x, y, \sigma^2)}$$

where $C(x, y, \sigma^2)$ is the normalizing constant for each fixed $(x, y) \in \mathbb{X}$

$$
\begin{aligned}
C(x, y, \sigma^2) &= \int_{\mathbb{Y}} g(\phi, \xi \mid x, y) \mathrm{d}\phi \, \mathrm{d}\xi \\
&= \int_0^{2\pi} \frac{1}{2} \left[\mathrm{erf}\left(\frac{R - x \cos \phi - y \sin \phi}{\sqrt{2}\sigma} \right) + \mathrm{erf}\left(\frac{R + x \cos \phi + y \sin \phi}{\sqrt{2}\sigma} \right) \right] \mathrm{d}\phi
\end{aligned}
$$

with erf the error function. Recalling that $R = 92$, $\phi \in [0, 2\pi]$ and $(x, y) \in [-64, 64]^2$ (i.e. we want to reconstruct a 128×128 pixels image) and selecting $\sigma = 0.02$ gives

$$\left| \frac{1}{2} \left[\mathrm{erf}\left(\frac{R - x \cos \phi - y \sin \phi}{\sqrt{2}\sigma} \right) + \mathrm{erf}\left(\frac{R + x \cos \phi + y \sin \phi}{\sqrt{2}\sigma} \right) \right] - 1 \right| < 10^{-17}$$

for all $\phi \in [0, 2\pi]$ and $(x, y) \in [-64, 64]^2$. The above shows that, for σ^2 sufficiently small (e.g. $\sigma^2 = 0.02^2$ as we use in our experiments), i.e. if the Gaussian distribution appropriately describes the alignment onto $x \cos \phi + y \sin \phi = \xi$,

$$\left| C(x, y, \sigma^2) - 2\pi \right| < 10^{-17}$$

for all $(x, y) \in \mathbb{X}$. Therefore we obtain, up to a negligible approximation, an integral equation satisfying (A0)–(A1) dividing h by 2π:

$$\frac{h(\phi, \xi)}{2\pi} = \int_{\mathbb{X}} \frac{\mathcal{N}\left(x \cos(\phi) + y \sin(\phi) - \xi; 0, \sigma^2\right)}{2\pi} f(x, y) \mathrm{d}x \mathrm{d}y.$$

Figure 10 shows relative error and ISE for the reconstructions in Figure 2; the ISE between the original image and the reconstructions at iteration 50 to 100 stabilizes below 0.08. The stopping criterion (18) is a trade-off between Monte Carlo error and convergence to a fixed point. In particular, when $\zeta(f_k^N) = \int_{\mathbb{X}} |f_k^N(x)|^2 \mathrm{d}x$, larger values of N will make the r.h.s. of (18) smaller which corresponds to a smaller tolerance to assess the convergence to the fixed point. On the other hand, small values of N will give poorer reconstructions and might require more iterations n to satisfy the stopping criterion (18). For instance, for $N = 1,000$ the stopping criterion is not satisfied in 100 iterations despite the r.h.s. of (18) being of order 10^{-3} against the 10^{-5} order when $N = 20,000$.
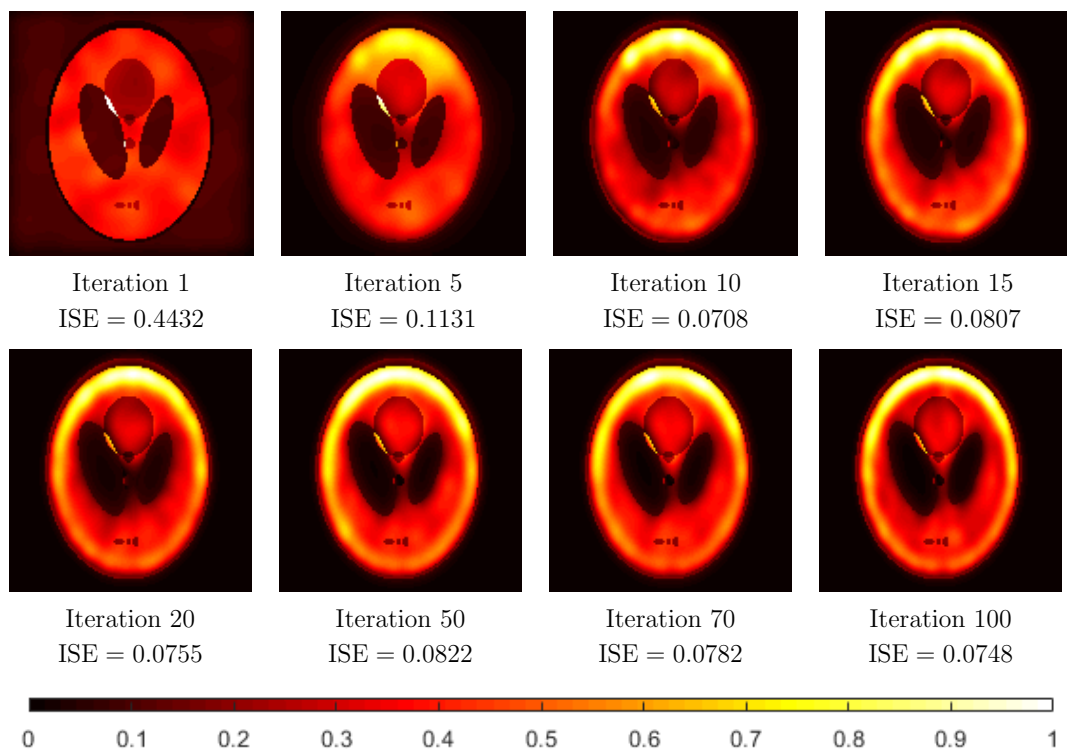
Figure 10: Relative error for the reconstructions in Figure 2. The ISE at each iteration is given in the captions and stabilises below 0.08.

G   Effect of Lower Bound on Gaussian Mixture Example

Consider the example in Section 4.1 and instead of defining the integrals on $\mathbb{X} = \mathbb{Y} = \mathbb{R}$ take $\mathbb{X} = \mathbb{Y} = [0.4 - a, 0.4 + a]$ with $a \to \infty$ so that we obtain the integral equation

$$\tilde{h}(y) = \int_{\mathbb{X}} \tilde{f}(x)\tilde{g}(y \mid x)\mathrm{d}x$$

with

$$\tilde{h}(y) = \frac{h(y)}{\frac{1}{3}C(a, 0.3, 0.045^2 + 0.015^2) + \frac{2}{3}C(a, 0.5, 0.045^2 + 0.043^2)}$$

$$\tilde{g}(y \mid x) = \frac{g(y \mid x)}{C(a, x, 0.045^2)}$$

$$\tilde{f}(x) = \frac{f(x)C(a, x, 0.045^2)}{\frac{1}{3}C(a, 0.3, 0.045^2 + 0.015^2) + \frac{2}{3}C(a, 0.5, 0.045^2 + 0.043^2)}$$

where

$$C(a, \mu, \sigma) := \int_{0.4-a}^{0.4+a} \mathcal{N}(x; \mu, \sigma^2)\mathrm{d}x = \frac{1}{2}\left(\mathrm{erf}\left(\frac{(a + 0.4 - \mu)}{\sqrt{2\sigma^2}}\right) - \mathrm{erf}\left(\frac{(0.4 - a - \mu)}{\sqrt{2\sigma^2}}\right)\right).$$

In any of the intervals $[0.4 - a, 0.4 + a]$ assumption (A1) is satisfied, in particular $\tilde{g}$ is bounded below. We study the behaviour of the reconstructions as $a \to \infty$ to check the influence of the lower bound on g on the accuracy of the reconstructions measured through the average ISE in (23) over 100 repetitions. The algorithmic set up is the same of Section 4.1. Figure 11 show that for $a \in [0.2, 1]$ (which corresponds to $\mathbb{X} = \mathbb{Y} = [0.2, 0.6]$ up to $\mathbb{X} = \mathbb{Y} = [-0.6, 1.4]$) the average reconstruction error is not influenced by the lower bound on g, the behaviour for larger values of a is equivalent since $|1 - \int_{-0.6}^{1.4} f(x)\mathrm{d}x| < 10^{-30}$.

H   Scaling with dimension

To explore the scaling with the dimension $d_{\mathbb{X}}$ of the domain of f of the discretized EMS (5) and the SMC implementation of EMS we revisit the Gaussian mixture model in Section 4.1 and extend it to higher dimension

$$f(x) = \frac{1}{3}\mathcal{N}(x; 0.3 \cdot \mathbf{1}_{d_{\mathbb{X}}}, 0.07^2 I_{d_{\mathbb{X}}}) + \frac{2}{3}\mathcal{N}(x; 0.7 \cdot \mathbf{1}_{d_{\mathbb{X}}}, 0.1^2 I_{d_{\mathbb{X}}}),$$
$$g(y \mid x) = \mathcal{N}(y; x, 0.15^2 I_{d_{\mathbb{X}}}),$$
$$h(y) = \frac{1}{3}\mathcal{N}(y; 0.3 \cdot \mathbf{1}_{d_{\mathbb{X}}}, (0.07^2 + 0.15^2)I_{d_{\mathbb{X}}}) + \frac{2}{3}\mathcal{N}(y; 0.7 \cdot \mathbf{1}_{d_{\mathbb{X}}}, (0.1^2 + 0.15^2)I_{d_{\mathbb{X}}}),$$

where $\mathbb{X} = \mathbb{Y} = \mathbb{R}^{d_{\mathbb{X}}}$ and $\mathbf{1}_{d_{\mathbb{X}}}$ and $I_{d_{\mathbb{X}}}$ denote the unit function in $\mathbb{R}^{d_{\mathbb{X}}}$ and the $d_{\mathbb{X}} \times d_{\mathbb{X}}$ identity matrix, respectively. In particular, note that for $d_{\mathbb{X}}$ up to 5 at least 97% of the mass of f is contained in $[0, 1]^{d_{\mathbb{X}}}$. We do not consider DKDE as these estimators approximate each marginal of f separately and then use the product of the marginals as approximation for f. In the particular mixture model we consider, this results in reconstructions with additional modes due to the underlying independence assumption (e.g. reconstructions of the 2-dimensional model in Figure 12 present two additional modes at $(0.7, 0.3)$ and $(0.3, 0.7)$).

First, we take $d_{\mathbb{X}} = 2$ and investigate the minimum number of bins/particles necessary to achieve reasonably good reconstructions. We consider three particle sizes $N = 10^2, 50^2, 100^2$ and set the total number of bins $B \approx N$ so that we obtain $\approx N^{1/2}$ equally spaced bins for each dimension. We stop iterating after 30 steps since we observed that convergence occurs within 30 iterations, the value of $\varepsilon = 10^{-3}$ is fixed and used for both the smoothing kernel and the smoothing matrix. The initial distribution is a uniform over $[0, 1]^2$ and we assume we have a sample $\mathbf{Y}$ of size 10^6 from h, so that $M = N$. This corresponds to the highest computational cost for Algorithm 1 but as observed in Appendix E.1 smaller values of M could be considered and would reduce the computational cost of running the SMC implementation of EMS.
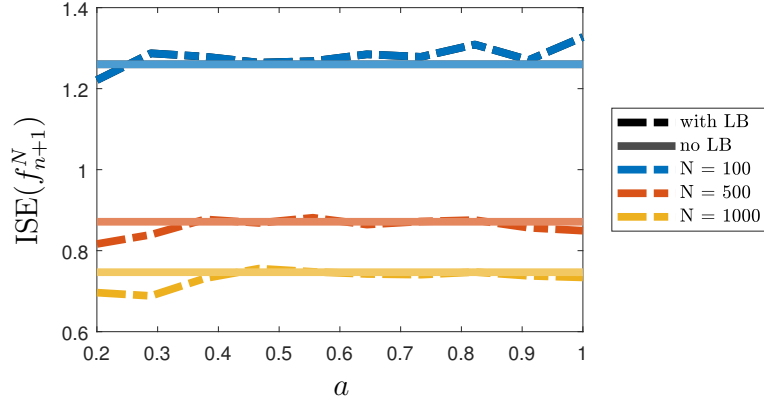
Figure 11: Influence of the lower bound (LB) on g on average reconstruction accuracy over 100 repetitions. The solid lines represent the average $\text{ISE}(f_{n+1}^N)$ for $a = \infty$ while the dashed lines the average $\text{ISE}(f_{n+1}^N)$ for finite a.

For small values of B the runtimes of EMS and SMC are similar, however the reconstructions obtained with EMS are poor and low resolution due to the very coarse discretization (Figure 12-left panel); on the contrary, the presence the kernel density estimator (17) guarantees smooth reconstructions even when the particle size is small (Figure 12-middle panel). In addition, as the number of particles N increases the accuracy of the reconstructions provided by SMC keeps increasing, while the EMS reconstructions do not improve as quickly, a phenomenon we already observed for the one-dimensional example in Section 4.1. For $N = B \geq 100$ the runtime of SMC is roughly 30% less than that of EMS with the accuracy of SMC being always larger than that of EMS.

Since the accuracy of kernel density estimators decreases when the dimension increases (Silverman, 1986) and is primarily used in this work for visualization and human interpretation (which becomes less informative in higher dimension, with the exception of low dimensional projections), to compare the performances of EMS and SMC in dimension $d_{\mathbb{X}} \geq 2$ we focus on approximating expectations w.r.t. η_{n+1} of appropriate test functions φ, in this case, in fact, Proposition 3 gives the rate of convergence in terms of the number of particles N. In particular, we consider mean, variance, the probability of the region $[0, 0.5]^{d_{\mathbb{X}}}$ and the probability of a hyper-sphere of radius 0.3 around the mode at $(0.3, \ldots, 0.3)$. We compare three particle sizes $N = 10^2, 10^3, 10^4$ and obtain the number of bins for each dimension as $\lceil N^{1/d_{\mathbb{X}}} \rceil$ so that the total number of bins, $B = \lceil N^{1/d_{\mathbb{X}}} \rceil^{d_{\mathbb{X}}}$, where $\lceil \cdot \rceil$ denotes the ceiling function, roughly matches N. This choice allows us to compare EMS and SMC reconstructions which require roughly the same runtime (Table 2). The SMC implementation in generally better at recovering the variance and the probability of the region $[0, 0.5]^{d_{\mathbb{X}}}$. For small values of N, B, both SMC and EMS have larger errors with discretized EMS achieving better crude estimates. However, as N, B increase SMC is consistently better at approximating the four quantities considered, in particular, in the case of mean and variance the estimates are at least one order of magnitude more accurate. This is achieved at a computational cost which is always smaller than that of EMS and that could be in principle reduced by considering smaller values of M.
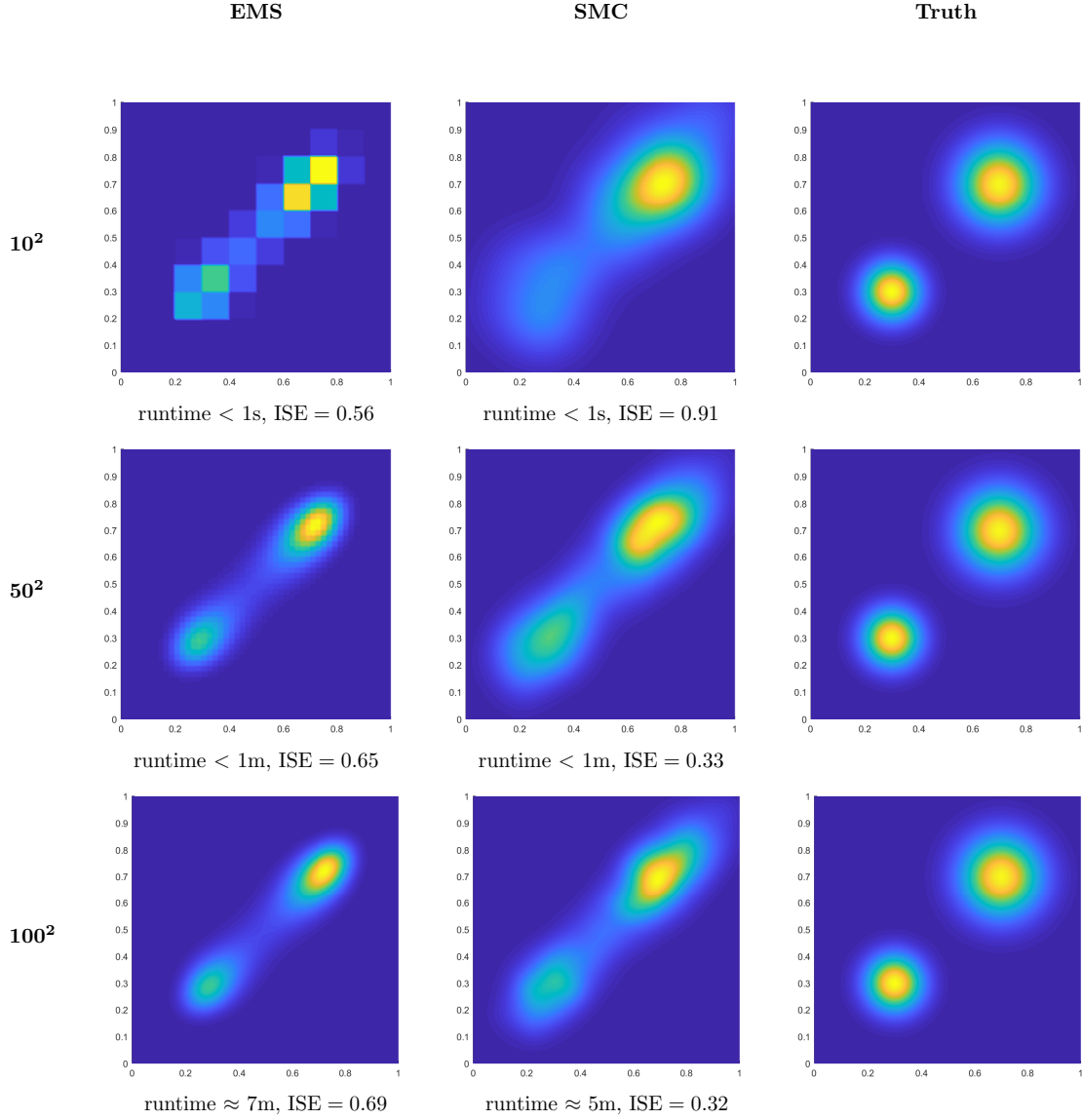
Figure 12: Reconstructions of a 2-dimensional mixture of Gaussian obtained with EMS and SMC. The number of bins/particles increases from 10^2 to 100^2. Runtime and accuracy are reported too.

Table 2: Mean squared error over 100 repetitions for mean, variance, probability of the lower quadrant and probability of a circle around the mode for the $d_{\mathbb{X}}$-dimensional Gaussian mixture model. Runtimes are reported too. Best values are in **bold**.

	mean	variance	$\mathbb{P}(\square)$	$\mathbb{P}(\bigcirc)$	$\log_{10}(\text{runtime / s})$
$d_{\mathbb{X}} = 2$					
EMS - $B = 10^2$	**1.38e-04**	4.96e-05	5.30e-02	**7.04e-03**	-1.71
SMC - $N = 10^2$	3.87e-04	**1.26e-05**	4.90e-02	1.02e-02	**-2.02**
EMS - $B = 32^2$	1.42e-04	5.31e-05	5.17e-02	**5.86e-03**	1.28
SMC - $N = 10^3$	**4.29e-05**	**5.81e-06**	4.69e-02	6.14e-03	**0.94**
EMS - $B = 100^2$	1.42e-04	5.38e-05	5.15e-02	6.11e-03	5.31
SMC - $N = 10^4$	**3.84e-06**	**4.51e-06**	4.67e-02	**5.68e-03**	**5.11**
$d_{\mathbb{X}} = 3$					
EMS - $B = 5^3$	**2.53e-04**	1.26e-04	1.46e-01	8.59e-03	-1.47
SMC - $N = 10^2$	3.76e-04	**3.23e-05**	**7.19e-02**	**3.92e-03**	**-2.06**
EMS - $B = 10^3$	2.00e-04	5.75e-05	9.00e-02	2.42e-03	1.40
SMC - $N = 10^3$	**4.62e-05**	**8.50e-06**	**8.24e-02**	**1.36e-03**	**1.08**

40

	mean	variance	$\mathbb{P}(\square)$	$\mathbb{P}(\bigcirc)$	$\log_{10}(\text{runtime / s})$
continued from previous page					
EMS - $B = 22^3$	2.04e-04	6.12e-05	8.83e-02	1.64e-03	5.66
SMC - $N = 10^4$	**3.53e-06**	**6.68e-06**	**8.23e-02**	**9.24e-04**	**5.30**
$d_{\mathbb{X}} = 4$					
EMS - $B = 4^4$	**1.98e-04**	**1.55e-05**	1.22e-01	**1.16e-03**	-0.65
SMC - $N = 10^2$	4.77e-04	9.77e-05	**7.35e-02**	4.41e-03	**-2.08**
EMS - $B = 6^4$	2.43e-04	4.02e-05	1.09e-01	7.80e-04	1.70
SMC - $N = 10^3$	**3.45e-05**	**1.80e-05**	**9.13e-02**	**5.47e-04**	**0.95**
EMS - $B = 10^4$	2.60e-04	6.59e-05	1.03e-01	5.54e-04	5.32
SMC - $N = 10^4$	**4.10e-06**	**8.58e-06**	**9.45e-02**	**2.38e-04**	**5.12**
$d_{\mathbb{X}} = 5$					
EMS - $B = 3^5$	**5.66e-05**	2.67e-04	2.12e-01	**1.27e-02**	-0.56
SMC - $N = 10^2$	6.59e-04	**1.34e-04**	**5.10e-02**	1.10e-02	**-1.96**
EMS - $B = 4^5$	2.42e-04	**2.08e-05**	1.29e-01	7.59e-04	1.51
SMC - $N = 10^3$	**5.57e-05**	4.54e-05	**7.84e-02**	**4.49e-04**	**1.14**
EMS - $B = 7^5$	2.82e-04	5.71e-05	1.36e-01	2.09e-04	6.63
SMC - $N = 10^4$	**3.39e-06**	**1.27e-05**	**9.57e-02**	**7.43e-05**	**5.36**

References for Supplementary Material

Berti, P., Pratelli, L. and Rigo, P. (2006), 'Almost sure weak convergence of random probability measures', *Stochastics* **78**(2), 91–97.

Cacoullos, T. (1966), 'Estimation of a multivariate density', *Ann Inst Statist Math* **18**(1), 179–189.

Chae, M., Martin, R. and Walker, S. G. (2018), 'On an algorithm for solving Fredholm integrals of the first kind', *Stat Comput* **29**, 645–654.

Chopin, N. (2004), 'Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference', *Ann Stat* **32**(6), 2385–2411.

Crisan, D. and Doucet, A. (2002), 'A survey of convergence results on particle filtering methods for practitioners', *IEEE T Signal Proces* **50**(3), 736–746.

Del Moral, P. (2004), *Feynman-Kac Formulae*, Springer.

Devroye, L. and Wagner, T. (1979), 'The L1 convergence of kernel density estimates', *Ann Stat* **7**(5), 1136–1139.

Dobrushin, R. L. (1956), 'Central limit theorem for nonstationary Markov chains. I', *Theory of Probability & Its Applications* **1**(1), 65–80.

Dudley, R. M. (2002), *Real Analysis and Probability*, 2 edn, Cambridge University Press.

Feinberg, E. A., Kasyanov, P. O. and Liang, Y. (2020), 'Fatou's lemma in its classical form and Lebesgue's convergence theorems for varying measures with applications to Markov decision processes', *Theor Probab Appl* **65**(2), 270–291.

Kress, R. (2014), *Linear Integral Equations*, Vol. 82, Springer.

Lee, D. and Vardi, Y. (1994), 'Experiments with maximum likelihood method for image motion deblurring', *J Appl Stat* **21**(1-2), 355–383.

Míguez, J., Crisan, D. and Djurić, P. M. (2013), 'On the convergence of two sequential Monte Carlo methods for maximum a posteriori sequence estimation and stochastic global optimization', *Stat Comput* **23**(1), 91–107.

Newey, W. K. (1991), 'Uniform convergence in probability and stochastic equicontinuity', *Econometrica* pp. 1161–1167.

Parzen, E. (1962), 'On estimation of a probability density function and mode', *Ann Math Stat* **33**(3), 1065–1076.

Radon, J. (1986), 'On the determination of functions from their integral values along certain manifolds', *IEEE T Med Imaging* **5**(4), 170–176.

Rubner, Y., Tomasi, C. and Guibas, L. J. (2000), 'The Earth Mover's Distance as a metric for image retrieval', *Int J Comput Vision* **40**(2), 99–121.

Rudin, W. (1964), *Principles of Mathematical Analysis*, Vol. 3, McGraw-Hill New York.

Schmon, S. M., Deligiannidis, G., Doucet, A. and Pitt, M. K. (2021), 'Large sample asymptotics of the pseudo-marginal method', *Biometrika* **108**(1), 37–51.

Serfozo, R. (1982), 'Convergence of lebesgue integrals with varying measures', *Sankhyā: The Indian Journal of Statistics, Series A* pp. 380–402.

Silverman, B. W., Jones, M. C., Wilson, J. D. and Nychka, D. W. (1990), 'A smoothed EM approach to indirect estimation problems, with particular, reference to stereology and emission tomography', *J R Stat Soc B* **52**(2), 271–324.

The MathWorks Inc. (1993), 'DECONVLUCY. Deblur image using Lucy-Richardson method.', `https://uk.mathworks.com/help/images/ref/deconvlucy.html`. [Online; accessed 25-March-2019].

Vardi, Y. and Lee, D. (1993), 'From image deblurring to optimal investments: Maximum likelihood solutions for positive linear inverse problems', *J R Stat Soc B* **55**(3), 569–612.

Zeidler, E. (1985), *Nonlinear Functional Analysis and its Applications: Part I: Fixed-Point Theorems*, Vol. 1, Springer.