Ensemble Kalman inversion for general likelihoods

Samuel Duffield *, Sumeetpal S. Singh

Department of Engineering, University of Cambridge, CB2 1PZ, UK

ARTICLE INFO

ABSTRACT

In this letter we generalise Ensemble Kalman inversion techniques to general Bayesian models where previously they were restricted to additive Gaussian likelihoods — all in the difficult setting where the likelihood can be sampled from, but its density not necessarily evaluated.

1. Introduction

Approximate Bayesian computation (ABC) refers to a collection of Monte Carlo methods for inference in Bayesian models

$$x \sim p(x), \qquad y \sim p(y \mid x), \tag{1}$$

where x is an unknown parameter to be inferred, y is some data that is known, $p(x)$ is a prior distribution and $p(y \mid x)$ is a likelihood function describing the data generating process. With ABC, we have the further restriction that the likelihood function cannot be evaluated pointwise and therefore the likelihood is said to be *intractable* — this rules out classical techniques such as Markov chain Monte Carlo (MCMC) or importance sampling. Instead, ABC techniques rely on the ability to simulate synthetic data from the likelihood $y^{(i)} \sim p(y \mid x^{(i)})$. Here the superscript indicates that $(x^{(i)}, y^{(i)})$ are the ith pair from a collection of N parameter-simulated data pairs $\{(x^{(i)}, y^{(i)})\}_{i=1}^{N}$.

The majority of ABC techniques have focused on adapting classical techniques to target an extended distribution that can be sampled from without evaluating the likelihood. Simulated data $\{y^{(i)}\}_{i=1}^{N}$ can then be compared to the true observed data y and subsequently only those parameter values $\{x^{(i)}\}_{i=1}^{N}$ with generated data sufficiently *close* to the true data are retained (where the measure of closeness is typically defined by a combination of summary statistics, distance function and tolerance level). Popular ABC variants are based on rejection sampling (Pritchard et al., 1999), MCMC (Marjoram et al., 2003) or sequential Monte Carlo (SMC) (Jasra et al., 2011), whilst there have also been approaches to incorporate machine learning techniques (Lueckmann et al., 2019) and alternative posteriors (Matsubara et al., 2021). For a thorough review of ABC techniques, see Beaumont (2019). ABC methods are asymptotically biased, however this bias is typically quantifiable and controllable (at the expense of further computational cost).

This setting of intractable likelihoods is challenging. As yet, ABC methods have been restricted to relatively low-dimensional Bayesian models and are renowned for requiring many likelihood simulations and therefore long run-times. The development of efficient and scalable ABC methods remains a prominent research goal.

---

* Correspondence to: Department of Engineering, University of Cambridge, CB2 1PZ, UK
   E-mail address: s@mduffield.com (S. Duffield).

In contrast, ensemble Kalman techniques (Evensen, 1994; Iglesias et al., 2013) are a class of Monte Carlo algorithms that have become very popular for inference in high-dimensional models with the restriction of additive Gaussian likelihoods

$$x \sim p(x), \qquad y \sim \mathbf{N}(y \mid \mathcal{H}(x), R),\tag{2}$$

where $\mathcal{H}$ a deterministic forward operator and R a noise covariance matrix which are both assumed known. As noted in Nott et al. (2012), ensemble Kalman methods make no use of likelihood evaluations and therefore they are themselves ABC techniques in the specific setting of additive Gaussian noise (2).

In this letter, we remove this restriction and **apply ensemble Kalman techniques to general likelihoods** (1). Thus developing an extremely general algorithm for inference in Bayesian models with intractable likelihoods that can perform well even with few likelihood simulations and is scalable to higher dimensions compared to conventional ABC techniques. The rest of the letter is structured as follows. Section 2 provides a brief account of existing ensemble Kalman techniques. We then introduce our generalised ensemble Kalman method in Section 3 before examining its performance numerically for both optimisation and uncertainty quantification in Section 4. Section 5 concludes and discusses future work.

2. Ensemble Kalman approach

Ensemble Kalman techniques were first introduced (Evensen, 1994) for generating online Monte Carlo approximations for state–space models.

For simplicity, let us consider the offline setting of (2) where the likelihood takes the form $p(y \mid x) = \mathbf{N}(y \mid \mathcal{H}(x), R)$ with $x \in \mathbb{R}^{d_x}$ and $y \in \mathbb{R}^{d_y}$. Then a single update step of the ensemble Kalman filter as described in Houtekamer and Mitchell (2001) takes the form

$$
\begin{aligned}
x_1^{(i)} &= x_0^{(i)} + C_0^{xH} \left(C_0^{HH} + R \right)^{-1} \left(y - \mathcal{H}(x_0^{(i)}) - \eta^{(i)} \right), \\
\eta^{(i)} &\sim \mathbf{N}(\eta \mid 0, R),
\end{aligned}
\tag{3}
$$

with prior samples $x_0^{(i)} \sim p(x)$, empirical covariance matrices

$$
\begin{aligned}
C_0^{xH} &= \frac{1}{N-1} \sum_{i=1}^{N} (x_0^{(i)} - \bar{x}_0)(\mathcal{H}(x_0^{(i)}) - \bar{\mathcal{H}}_0)^{\mathsf{T}}, \\
C_0^{HH} &= \frac{1}{N-1} \sum_{i=1}^{N} (\mathcal{H}(x_0^{(i)}) - \bar{\mathcal{H}}_0)(\mathcal{H}(x_0^{(i)}) - \bar{\mathcal{H}}_0)^{\mathsf{T}},
\end{aligned}
\tag{4}
$$

and empirical means $\bar{x}_0 = \frac{1}{N} \sum_{i=1}^{N} x_0^{(i)}$ and $\bar{\mathcal{H}}_0 = \frac{1}{N} \sum_{i=1}^{N} \mathcal{H}(x_0^{(i)})$.

This ensemble Kalman update is asymptotically unbiased in the special case of Gaussian prior $p(x) = \mathbf{N}(x \mid m, Q)$ and linear Gaussian likelihood $p(y \mid x) = \mathbf{N}(y \mid Hx, R)$ which gives

$$p(x, y) = \mathbf{N}\left(\begin{pmatrix} x \\ y \end{pmatrix} \mid \begin{pmatrix} m \\ Hm \end{pmatrix}, \begin{pmatrix} Q & QH^{\mathsf{T}} \\ HQ & HQH^{\mathsf{T}} + R \end{pmatrix} \right).\tag{5}$$

This model has an analytically tractable posterior distribution $p(x \mid y) = \mathbf{N}(x \mid m^y, Q^y)$ where

$$
\begin{aligned}
m^y &= m + QH^{\mathsf{T}} \left(HQH^{\mathsf{T}} + R \right)^{-1} (y - Hm), \\
Q^y &= Q - QH^{\mathsf{T}} \left(HQH^{\mathsf{T}} + R \right)^{-1} HQ.
\end{aligned}
$$

Outside of this special linear Gaussian case, the ensemble Kalman update (3) is known to be asymptotically biased. There is however, a vast amount of empirical evidence demonstrating stability in a variety of complex non-linear state–space models with a very small number of particles, e.g. Houtekamer et al. (2005), Roth et al. (2017). In our view, this summarises the ensemble Kalman paradigm:

Asymptotically unbiased in the linear Gaussian case, **otherwise biased but empirically stable**.

Where the term *empirically stable* represents the ability of the ensemble $\{x^{(i)}\}_{i=1}^{N}$ to cover the true underlying value of the state without degenerating to a single particle. Naturally, the acceptance of a bias induced by this paradigm could be extreme in some problems, however as mentioned, there is still very much a practical desire for these numerically efficient but biased methods in this difficult setting of intractable likelihoods.

For non-Gaussian priors and non-linear Gaussian likelihoods (1), the distribution of particles from a single step ensemble Kalman update may be quite different from the true posterior. As noted in Iglesias et al. (2013), this can be mitigated by iterating the ensemble Kalman update in (3). The idea was refined in Iglesias (2015) to be more in line with the tempered likelihood approach of SMC samplers (Del Moral et al., 2006; Jasra et al., 2011). This iterative ensemble

Kalman approach is termed *ensemble Kalman inversion* (EKI) and a single iterate takes the form

$$
\begin{aligned}
x_\ell^{(i)} &= x_{\ell-1}^{(i)} + C_{\ell-1}^{xH} \left(C_{\ell-1}^{HH} + h_\ell^{-1} R \right)^{-1} (y - \mathcal{H}(x_{\ell-1}^{(i)}) - \eta_\ell^{(i)}), \\
\eta_\ell^{(i)} &\sim \mathbf{N}(\eta_\ell \mid 0, h_\ell^{-1} R),
\end{aligned}
\tag{6}
$$

where $C_{\ell-1}^{xH}$ and $C_{\ell-1}^{HH}$ are as in (4) applied to the particles $\{x_{\ell-1}^{(i)}\}_{i=1}^N$, the subscript indicates the iteration of the algorithm, i.e. $\ell = 0, \dots, L$ where L is the total number of iterations. The parameter h_ℓ^{-1} can be considered a stepsize parameter or equally h_ℓ as an incremental inverse temperature parameter. In the fully linear Gaussian case (5), the particles $\{x_\ell^{(i)}\}_{i=1}^N$ at each step are asymptotically unbiased for a tempered version of the posterior

$$
\begin{aligned}
p_\ell(x) &\propto p(x)p(y \mid x)^{\lambda_\ell} \propto \mathbf{N}(x \mid m_\ell, Q_\ell), \\
m_\ell &= m + QH^T \left(HQH^T + \lambda_\ell^{-1} R \right)^{-1} (y - Hm), \\
Q_\ell &= Q - QH^T \left(HQH^T + \lambda_\ell^{-1} R \right)^{-1} HQ,
\end{aligned}
\tag{7}
$$

where $\lambda_\ell = \sum_{r=1}^\ell h_r$ is the inverse temperature. The tempered posterior can also be defined iteratively

$$
\begin{aligned}
m_\ell &= m_{\ell-1} + Q_{\ell-1} H^T \left(HQ_{\ell-1}H^T + h_\ell^{-1} R \right)^{-1} (y - Hm_{\ell-1}), \\
Q_\ell &= Q_{\ell-1} - Q_{\ell-1} H^T \left(HQ_{\ell-1}H^T + h_\ell^{-1} R \right)^{-1} HQ_{\ell-1}.
\end{aligned}
\tag{8}
$$

Thus, iterating for L steps with stepsizes such that $\lambda_L = \sum_{r=1}^L h_r = 1$ obtains particles $\{x_L^{(i)}\}_{i=1}^N$ that are asymptotically unbiased for the true posterior in the linear Gaussian case.

Iterating until $\lambda_L = 1$ is only one possible method of termination, another approach is to adopt an optimisation style stopping criterion, for example Iglesias (2015) suggest terminating when the average of the forward operations $\bar{\mathcal{H}}_L$ is suitably close the true data, e.g. when $(y - \bar{\mathcal{H}}_L)^T R^{-1} (y - \bar{\mathcal{H}}_L) < \tau$ for some stopping parameter τ.

3. Generalised ensemble Kalman inversion

In this section, we generalise ensemble Kalman inversion to the case where data is generated according to any likelihood $p(y \mid x)$ rather than the setting described in Section 2 which is restricted to likelihoods of the form $\mathbf{N}(y \mid \mathcal{H}(x), R)$. The resulting algorithm only requires samples from the prior $p(x)$ and the ability to simulate from the likelihood $p(y \mid x)$ for a given parameter x, as in approximate Bayesian computation. The final particle approximation is asymptotically biased for the true posterior with the exception of the fully linear Gaussian case (5) and therefore remains faithful to the ensemble Kalman paradigm.

3.1. General likelihoods

In this more general case, we can no longer form the empirical covariance matrices in (4) as we do not have access to the deterministic $\mathcal{H}$. Instead we can form

$$
C_\ell^{xx} = \frac{1}{N-1} \sum_{i=1}^N (x_\ell^{(i)} - \bar{x}_\ell)(x_\ell^{(i)} - \bar{x}_\ell)^T,
\tag{9a}
$$

$$
C_\ell^{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_\ell^{(i)} - \bar{x}_\ell)(y_\ell^{(i)} - \bar{y}_\ell)^T,
\tag{9b}
$$

$$
C_\ell^{yy} = \frac{1}{N-1} \sum_{i=1}^N (y_\ell^{(i)} - \bar{y}_\ell)(y_\ell^{(i)} - \bar{y}_\ell)^T,
\tag{9c}
$$

and $C_\ell^{yx} = C_\ell^{xy\,T}$. Here we have simulated data $y_\ell^{(i)} \sim p(\cdot \mid x_\ell^{(i)})$ and means $\bar{x}_\ell = \frac{1}{N} \sum_{i=1}^N x_\ell^{(i)}, \bar{y}_\ell = \frac{1}{N} \sum_{i=1}^N y_\ell^{(i)}$.

In the fully linear Gaussian case (5) or rather the tempered version (7), (8) we have $C_\ell^{xy} \to HQ_\ell$ and $C_\ell^{yy} \to HQ_\ell H^T + R$. We now note that we can also calculate

$$
C_\ell^{y|x} = C_\ell^{yy} - C_\ell^{yx} C_\ell^{xx\,-1} C_\ell^{xy},
\tag{10}
$$

and observe that in the linear Gaussian case $C_\ell^{y|x} \to (HQ_{\ell-1}H^T + R) - (HQ_{\ell-1})Q_{\ell-1}^{-1}(Q_{\ell-1}H^T) = R$. Thus we can use the particles to approximate the covariance of the likelihood empirically. We therefore adjust the tempered ensemble Kalman iteration (6) to the following

$$
\begin{aligned}
x_\ell^{(i)} &= x_{\ell-1}^{(i)} + C_{\ell-1}^{xy} \left(C_{\ell-1}^{yy} + (h_\ell^{-1} - 1) C_{\ell-1}^{y|x} \right)^{-1} \left(y - y_{\ell-1}^{(i)} - \eta_\ell^{(i)} \right), \\
y_{\ell-1}^{(i)} &\sim p(y_{\ell-1} \mid x_{\ell-1}^{(i)}),
\end{aligned}
\tag{11}
$$

$$\eta_\ell^{(i)} \sim \mathbf{N}(\eta \mid 0, (h_\ell^{-1} - 1)C_{\ell-1}^{y|x}).$$

In the linear Gaussian case and large sample limit, the above ensemble Kalman move becomes

$$x_\ell^{(i)} = x_{\ell-1}^{(i)} + Q_{\ell-1}H^T \left((HQ_{\ell-1}H^T + R) + (h_\ell^{-1} - 1)R\right)^{-1} (y - y_{\ell-1}^{(i)} - \eta_\ell^{(i)}),$$

$$y_{\ell-1}^{(i)} \sim \mathbf{N}(y_{\ell-1} \mid Hx_{\ell-1}^{(i)}, R),$$

$$\eta_\ell^{(i)} \sim \mathbf{N}(\eta_\ell \mid 0, (h_\ell^{-1} - 1)R).$$

Where the $(h_\ell^{-1} - 1)$ scaling has been chosen to ensure that the statistics $\mathbb{E}[x_\ell^{(i)}]$ and $\mathrm{Cov}[x_\ell^{(i)}]$ match those in (8). Therefore the ensemble Kalman update in (11) is asymptotically unbiased in the linear Gaussian case, and consistent with the ensemble Kalman paradigm. The full procedure is listed in Algorithm 1.

Algorithm 1 Ensemble Kalman Inversion for General Likelihoods

---

1: Given a sequence of inverse temperatures $\{\lambda_\ell\}_{\ell=0}^L$ and stopping criterion − which may both be adaptive, see Sections 3.2 and 3.3.

2: Sample from prior

$$x_0^{(i)} \sim p(x) \qquad\qquad\qquad i = 1, \ldots, N$$

3: **for** $\ell = 1, \ldots, L$ **do**

4:     Simulate observations

$$y_{\ell-1}^{(i)} \sim p(y \mid x_{\ell-1}^{(i)}) \qquad\qquad\qquad i = 1, \ldots, N$$

5:     Form sample covariances $C_{\ell-1}^{xx}, C_{\ell-1}^{xy}, C_{\ell-1}^{yy}$ (9) and $C_{\ell-1}^{y|x}$ (10).

6:     Set stepsize $h_\ell = \lambda_\ell - \lambda_{\ell-1}$

7:     Generate observation perturbations

$$\eta_\ell^{(i)} \sim \mathbf{N}\left(\eta \mid 0, \left(h_\ell^{-1} - 1\right) C_{\ell-1}^{y|x}\right) \qquad\qquad\qquad i = 1, \ldots, N$$

8:     Move particles

$$x_\ell^{(i)} = x_{\ell-1}^{(i)} + C_{\ell-1}^{xy} \left(C_{\ell-1}^{yy} + (h_\ell^{-1} - 1)C_{\ell-1}^{y|x}\right)^{-1} \left(y - y_{\ell-1}^{(i)} - \eta_\ell^{(i)}\right) \qquad\qquad\qquad i = 1, \ldots, N$$

9: **return** $\{x_L^{(i)}\}_{i=1}^N$

3.2. Stepsize selection

The motivation for iterative ensemble Kalman inversion is that for difficult non-Gaussian problems, moving directly from prior to posterior in one step is an extremely difficult task. By taking many smaller steps the particles can explore the state–space and have a better chance of settling in regions of high posterior probability. There is, therefore, a trade-off to be made − more steps means greater exploration but also more likelihood simulations and a longer runtime.

The stepsizes $\{h_\ell\}_{\ell=1}^L$ equivalently define an inverse temperature schedule $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_L$ where $h_\ell = \lambda_\ell - \lambda_{\ell-1} > 0$ for $\ell = 1, \ldots, L$.

For SMC samplers (Del Moral et al., 2006), it is common for the next inverse temperature to be selected adaptively (Jasra et al., 2011) such that the effective sample size (of the sequential importance weights $\{w_\ell^{(i)}\}_{i=1}^N$) decreases by a fixed amount at each iteration. That is, select λ_ℓ such that $\mathrm{ESS}(\{w_\ell^{(i)}\}_{i=1}^N) = (\sum_{i=1}^N w_\ell^{(i)2})^{-1} \approx \rho N$, where the normalised weights are a function of λ_ℓ and $\rho \in (0, 1)$ is a tuning parameter that controls the size of the steps. The root for λ_ℓ can be found using a numerical bisection algorithm in $(\lambda_{\ell-1}, \lambda_L)$ and requires no additional likelihood evaluations. This way the algorithm smoothly transitions from prior to posterior whereas simply using a constant stepsize might induce a degenerate series of intermediate distributions unless an excessively small stepsize is used. This idea was ported to ensemble Kalman inversion in Iglesias et al. (2018) where the following pseudo-weights are used $\hat{w}_\ell^{(i)} \propto \exp\left(-\frac{1}{2}(\lambda_\ell - \lambda_{\ell-1})\left(y - \mathcal{H}(x_{\ell-1}^{(i)})\right)^T R^{-1}\left(y - \mathcal{H}(x_{\ell-1}^{(i)})\right)\right)$, as ensemble Kalman inversion does not compute importance weights inherently. Here we directly adapt these pseudo-weights to the general likelihood case

$$\hat{w}_\ell^{(i)} \propto \exp\left(-\frac{1}{2}(\lambda_\ell - \lambda_{\ell-1})\left(y - y_{\ell-1}^{(i)}\right)^T C_{\ell-1}^{y|x-1}\left(y - y_{\ell-1}^{(i)}\right)\right). \tag{12}$$

As noted in Iglesias et al. (2018), the lack of resampling means the user can be more aggressive with the choice of ρ, in our experiments we set $\rho = \frac{1}{2}$.

3.3. Stopping criteria

We consider two stopping criteria.

- **Sampling** — terminate when $\lambda_L = 1$. This stopping criterion mimics that of tempered likelihood approaches for tractable likelihoods and is applied to ensemble Kalman inversion in Iglesias et al. (2018). This approach is asymptotically unbiased for the true posterior in the linear Gaussian case and aims to quantify uncertainty around parameters.
- **Optimisation** — terminate when the marginal variance (of the particles) in each dimension becomes sufficiently small, i.e. when $[C_\ell^{xx}]_{kk} < \upsilon [C_0^{xx}]_{kk}, \forall k \in \{1, \ldots d_x\}$ for a predefined $\upsilon \in [0, 1]$ (which we set to 10^{-2}). Here the notation $[A]_{ij}$ indexes the i, j coordinate in the matrix A. Under this approach the algorithm is iterated until the particles form a consensus approaching a single point estimate, which in the linear Gaussian case will (asymptotically) be the maximum likelihood estimator.

4. Numerical experiments

We now examine the performance of both ensemble Kalman inversion for sampling and optimisation versus approximate Bayesian computation methods for two static Bayesian inference problems with intractable likelihoods.

We consider two ABC techniques, the first of which being that of ABC-SMC (Del Moral et al., 2012). In ABC-SMC, a series of importance weighted Monte Carlo approximations are generated iteratively to target

$$v_{\kappa_\ell}(x, y') \propto p(x)p(y' \mid x)\mathbb{I}[\|y - y'\| < \kappa_\ell], \tag{13}$$

for a series of decreasing parameters $\kappa_0 > \kappa_1 > \cdots > \kappa_L$. Here y is the true data, y' is simulated data, whilst $\mathbb{I}[\cdot]$ is the indicator function and $\|\cdot\|$ is the Euclidean norm. As in Del Moral et al. (2006), particles are rejuvenated using a random-walk Metropolis–Hastings kernel, with covariance adapted to $2.38^2 d_x^{-1} C_{\ell-1}^{xx}$ a scaled version of the empirical covariance of the previous particles and threshold parameters κ_ℓ adaptively chosen such that ESS $\approx 0.9N$. Particles are resampled when the ESS drops below $0.5N$ and the algorithm terminates when the acceptance rate of the Metropolis–Hastings kernel first falls below 1.5%.

For our second ABC technique, we run the random-walk Metropolis–Hastings kernel directly as an ABC-MCMC algorithm. As described in Vihola and Franks (2020) we use a Robbins–Monro schedule to adaptively tune the stepsize, preconditioner combination to $2.38^2 d_x^{-1}\hat{\Sigma}$ where $\hat{\Sigma}$ is the empirical covariance of the historical chain and adapt the threshold parameter κ from (13) such that 10% of samples are accepted.

As mentioned, for EKI we determine the stepsize h_ℓ (equivalently the next inverse temperature λ_ℓ) adaptively such that the ESS of the pseudo-weights (12) is approximately $0.5N$. We examine both the sampling and optimisation stopping criteria discussed in Section 3.3.

4.1. g-and-k distribution

Popular as a benchmark for ABC algorithms, the g-and-k distribution family is defined by the quantile function

$$F^{-1}(u) = A + B\left(1 + c\frac{1 - \exp(-gz(u))}{1 + \exp(-gz(u))}\right)(1 + z(u)^2)^k z(u), \tag{14}$$

where $z(\cdot)$ is the quantile function of a standard normal distribution. The constant c is typically set to 0.8 and considered known. We set the remaining parameters $x = (A, B, g, k)$ with *true* values $(3, 1, 2, 1/2)$ but consider them unknown (to be inferred) each with prior $\mathcal{U}(\cdot \mid 0, 10)$. The g-and-k likelihood is defined implicitly by the quantile function (14). This likelihood function cannot be obtained analytically and thus we cannot easily utilise traditional posterior inference methods such as MCMC or importance sampling — although their ABC counterparts are still possible as the likelihood can be easily simulated for given parameters by simply evaluating the quantile function (14) at a sampled standard uniform random variable.

We assume we have data of 1000 i.i.d. samples from (14) which we summarise into 100 evenly spaced order statistics for both EKI and ABC. It is common practice for Monte Carlo algorithms to apply a transformation converting constrained variables to ones in $\mathbb{R}$, this alleviates the possibility of parameters being moved outside of the bounded domain, here we unconstrain all parameters using the transformation $z(\cdot/10)$ for both ABC and EKI (other transformations are possible although the choice is not typically considered a significant factor in performance).

In Fig. 1, we compare the marginal distributions produced by EKI for sampling and those from ABC-SMC. We observe that EKI has centred on the vicinity of the truth for all 4 parameters whereas ABC-SMC has failed to concentrate in the g variable in particular. We note that the two posteriors differ quite significantly – an indication of the high levels of non-linearity in the g-and-k likelihood – yet the EKI posterior remains informative.

We then examine the second EKI stopping criterion, that of optimisation. The EKI optimisation procedure, Section 3.3, iterates beyond $\lambda_\ell = 1$ until a consensus is reached on a single point approximation for the maximum likelihood estimator. Fig. 2 displays 7 equally spaced iterations as the inverse temperature (which is adaptively selected at each iteration based
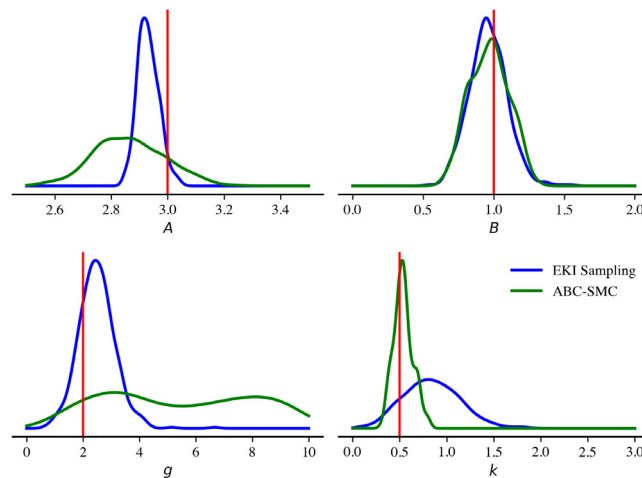
Fig. 1. *g*-and-*k* marginal posteriors for EKI sampling and ABC-SMC, $N = 500$, truth in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
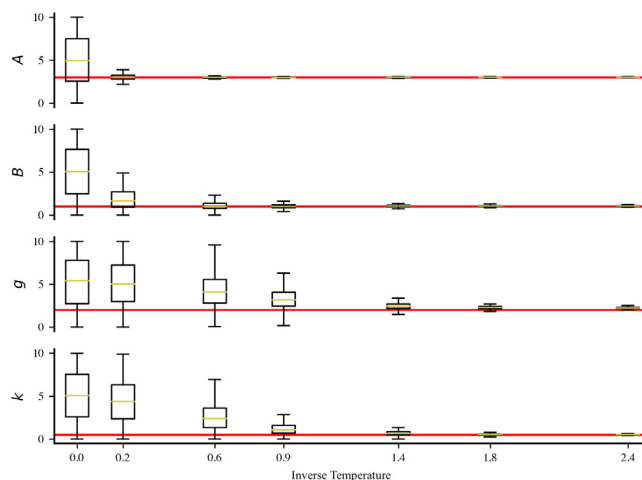

Fig. 2. *g*-and-*k* convergence of EKI for optimisation, $N = 500$, truth in red. Boxes display range and quartiles of ensemble. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

on the distance of between simulated and true data (12)) increases above $\lambda_\ell = 1$ before terminating when all standard deviations are suitably small, note the increasing stepsizes as the ensemble converges to a point estimate. The particles converge quickly and very closely to the true underlying parameter values.

Finally, we vary the sample size N between 200 and 5000 then repeat the experiment 10 times. We plot the root mean squared error against the number of likelihood simulations utilised by each algorithm as N changes in Fig. 3. Note that the adaptive stopping of the EKI and ABC-SMC techniques means that the number of iterations and therefore likelihood simulations can vary across seeds. We observe that both EKI stopping criteria are consistently outperforming both ABC-MCMC and ABC-SMC for the same computational cost. Curiously, we observe that the performance of EKI for sampling deteriorates as N increases although this is not the case for the optimisation approach. We posit that this might be due to the smaller sample sizes under estimating the noise covariance (a common occurrence in ensemble Kalman techniques (Tong et al., 2016)) and therefore moving the particles further.

4.2. Stochastic Lorenz 96

We now consider the task of inferring the initial conditions of a noisy version of the Lorenz 96 model (Lorenz, 1995). The Lorenz 96 model is a simplified representation of oceanic flows that is commonly used as a testbed for high-dimensional data assimilation techniques.
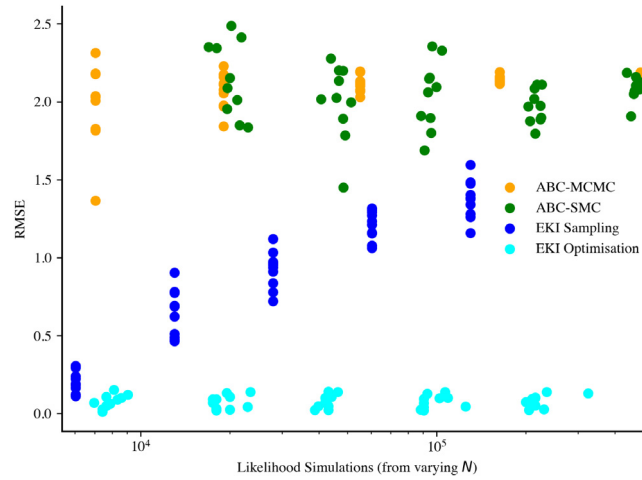
Fig. 3. Root mean squared error versus number of likelihood simulations induced by varying N (from 200 to 5000), on g-and-k distribution. Repeated over 10 randomly generated sets of observations.
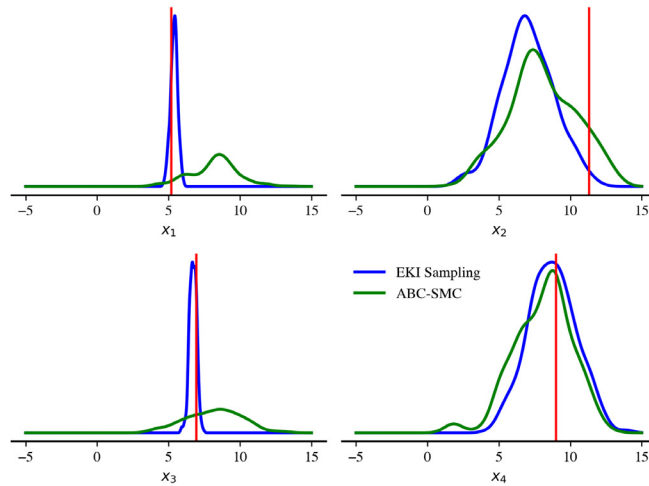
Fig. 4. Lorenz 96 marginal densities for EKI sampling and ABC-SMC, $N = 500$, truth in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The Lorenz 96 dynamics (with added stochasticity) are defined by the SDE

$$dx_t[m] = (-x_t[m-2]x_t[m-1] + x_t[m-1]x_t[m+1] - x_t[m] + F)dt + dW_t,$$

for $m = 1, \ldots, d_x$ with cyclic coordinates $x_t[0] = x_t[d_x]$, $x_t[-1] = x_t[d_x - 1]$ and $x_t[d_x + 1] = x_t[1]$. Here the notation $x[i]$ simply accesses the ith coordinate of the vector x. We adopt the common high-dimensional setup of $d_x = 40$ and $F = 8$, which is known to produce challenging, chaotic dynamics.

We define our inference goal as obtaining the initial conditions $x_0 \in \mathbb{R}^{d_x}$ given observations of every other dimension, perturbed by Gaussian noise with variance 0.1, at times $t = 1, 2, 3, 4, 5$ — resulting in $d_y = 100$ observations. We set the prior to $p(x_0) = \mathbf{N}(x_0 \mid F, 5\mathbb{I}_{40})$ and simulate from the likelihood with an Euler–Maruyama scheme (stepsize 0.001). The stochasticity in the Lorenz 96 dynamics provides a more realistic influence of uncertainty but the iterative addition of noise in the Euler–Maruyama scheme makes evaluating the likelihood density intractable.

In our experiments, underlying true values for the initial conditions are sampled from the prior and then observations generated using the same Euler–Maruyama scheme as above.

Recall that odd dimensions are directly observed whereas even dimensions are unobserved. We see in Fig. 4 that EKI for sampling converges very closely around the truth in observed dimensions and is understandably less certain about unobserved dimensions. In contrast ABC-SMC, performs similarly for all dimensions and is likely struggling with the dimensionality of the problem.
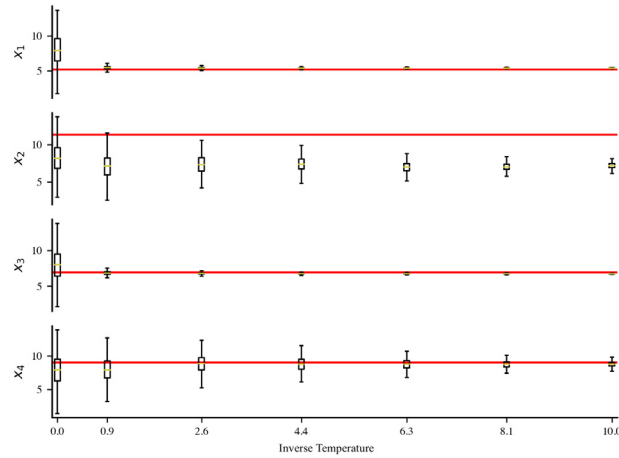
Fig. 5. Lorenz 96 convergence of EKI for optimisation, $N = 500$, truth in red. Boxes display range and quartiles of ensemble. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
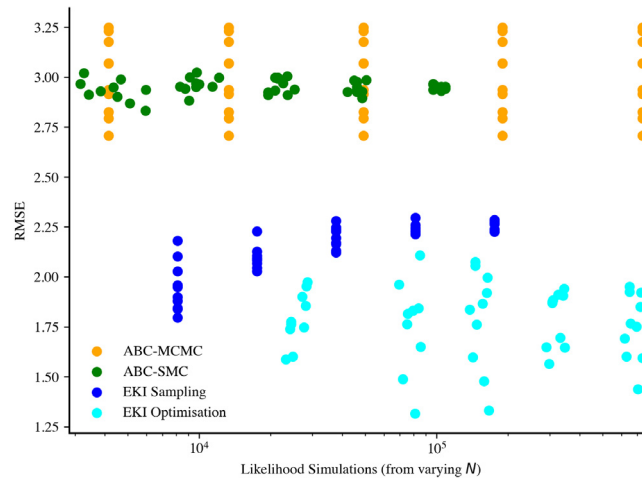
Fig. 6. Root mean squared error for number of likelihood simulation induced by varying N (from 200 to 5000), on Lorenz 96 example. Repeated over 10 randomly generated sets of observations.

When we push the EKI into high inverse temperatures in Fig. 5 we first see that the particles struggle to collapse in unobserved dimensions. This is an indication that the given observations are insufficient to provide a confident point estimate in those unobserved dimensions. We also notice that the particles converge in the second dimension away from the true underlying value of the parameter — this may be an indication that the true maximum likelihood is not necessarily guaranteed to be close to the truth (for all dimensions) under this model setup.

In Fig. 6, we also observe the phenomenon of decreasing performance in EKI for sampling as N increases for the L96 example — although less so. However, again we see that EKI consistently outperforms ABC, although this is perhaps not surprising in a high-dimensional example given the regular use of ensemble Kalman techniques in similar settings. In this case, the EKI for optimisation utilised significantly more likelihood simulations as it requires more iterations to converge — it also suffered high variance results whereas the performance of the EKI for sampling was more stable.

5. Conclusion

We have extended the work of ensemble Kalman inversion (Iglesias et al., 2013, 2018) to problems with general likelihoods as opposed to the common restriction of additive Gaussian noise. In doing so, we remain faithful to the ensemble Kalman paradigm, i.e. our generalisation remains asymptotically unbiased in the linear Gaussian case. We described how to apply the technique for both optimisation and for sampling or uncertainty quantification. We have demonstrated both speed and accuracy of the novel ensemble Kalman inversion algorithm in a difficult benchmark problem as well as a high dimensional spatial example. The computational cost of the ensemble Kalman inversion is

$O(Ld_x^3 + LNd_x^2)$ and only requires LN likelihood simulations which is the typical computational bottleneck for problems within approximate Bayesian computation.

We observe the curious phenomenon that increasing the number of particles fails to increase accuracy when applying ensemble Kalman inversion for sampling — at least in terms of mean square error. An outstanding question is how to correct for this. This phenomenon is not entirely new but is not well understood, it would be intriguing to investigate whether it could potentially be mitigated by covariance regularisation (Houtekamer and Mitchell, 2001) or moment-matching ideas (Lei and Bickel, 2011).

A natural extension of the stochastic ensemble Kalman inversion algorithm presented in this chapter would be the conversion to the square-root ensemble Kalman variants (Bishop et al., 2001; Anderson, 2001) that instead deterministically move particles in a way that remains asymptotically unbiased for fully linear Gaussian problems. By removing a layer of stochasticity we hope to increase the numerical stability of the inversion algorithm.

Outside of linear Gaussian problems, ensemble Kalman inversion is asymptotically biased. It would interesting to investigate embedding an ensemble Kalman kernel within an ABC-SMC sampler. This way, the ensemble Kalman inversion would inherit the theory of approximate Bayesian computation and become asymptotically unbiased for ν_K. However, it is not clear how to choose the backward kernel to induce stable importance weights.

A further application of the presented ensemble Kalman inversion would be to investigate its use within state–space models. It may be that utilising an iterative tempered approach to the update step within an ensemble Kalman filter may improve sample quality. Additionally, we could adapt ensemble Kalman inversion to be used in the online setting of state–space models with intractable observation densities (Jasra et al., 2012).

Acknowledgements

References

Anderson, J.L., 2001. An ensemble adjustment Kalman filter for data assimilation. Mon. Weather Rev. 129 (12), 2884–2903. http://dx.doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

Beaumont, M.A., 2019. Approximate Bayesian computation. Annu. Rev. Stat. Appl. 6 (1), 379–403. http://dx.doi.org/10.1146/annurev-statistics-030718-105212.

Bishop, C.H., Etherton, B.J., Majumdar, S.J., 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. Mon. Weather Rev. 129 (3), 420–436. http://dx.doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Del Moral, P., Doucet, A., Jasra, A., 2006. Sequential Monte Carlo samplers. J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (3), 411–436.

Del Moral, P., Doucet, A., Jasra, A., 2012. An adaptive sequential Monte Carlo method for approximate Bayesian computation. Stat. Comput. 22 (5), 1009–1020. http://dx.doi.org/10.1007/s11222-011-9271-y.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. Oceans 99 (C5), 10143–10162. http://dx.doi.org/10.1029/94JC00572.

Houtekamer, P.L., Mitchell, H.L., 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Weather Rev. 129 (1), 123–137. http://dx.doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Houtekamer, P.L., Mitchell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., Hansen, B., 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. Mon. Weather Rev. 133 (3), 604–620. http://dx.doi.org/10.1175/MWR-2864.1.

Iglesias, M., 2015. Iterative regularization for ensemble data assimilation in reservoir models. Comput. Geosci. 19 (1), 177–212. http://dx.doi.org/10.1007/s10596-014-9456-5.

Iglesias, M., Law, K.J.H., Stuart, A.M., 2013. Ensemble Kalman methods for inverse problems. Inverse Problems 29 (4), 045001. http://dx.doi.org/10.1088/0266-5611/29/4/045001.

Iglesias, M., Park, M., Tretyakov, M.V., 2018. Bayesian inversion in resin transfer molding. Inverse Problems 34 (10), 105002. http://dx.doi.org/10.1088/1361-6420/aad1cc.

Jasra, A., Singh, S.S., Martin, J.S., McCoy, E., 2012. Filtering via approximate Bayesian computation. Stat. Comput. 22 (6), 1223–1237. http://dx.doi.org/10.1007/s11222-010-9185-0.

Jasra, A., Stephens, D.A., Doucet, A., Tsagaris, T., 2011. Inference for Lévy-driven stochastic volatility models via adaptive sequential Monte Carlo. Scand. J. Stat. 38 (1), 1–22. http://dx.doi.org/10.1111/j.1467-9469.2010.00723.x.

Lei, J., Bickel, P., 2011. A moment matching ensemble filter for nonlinear non-Gaussian data assimilation. Mon. Weather Rev. 139, 3964–3973.

Lorenz, E.N., 1995. Predictability: A problem partly solved. In: Seminar on Predictability, 4-8 September 1995, Vol. 1. ECMWF, Shinfield Park, Reading, pp. 1–18.

Lueckmann, J.-M., Bassetto, G., Karaletsos, T., Macke, J.H., 2019. Likelihood-free inference with emulator networks. In: Symposium on Advances in Approximate Bayesian Inference. PMLR, pp. 32–53.

Marjoram, P., Molitor, J., Plagnol, V., Tavaré, S., 2003. Markov Chain Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. 100 (26), 15324–15328. http://dx.doi.org/10.1073/pnas.0306899100.

Matsubara, T., Knoblauch, J., Briol, F.-X., Oates, C., 2021. Robust generalised Bayesian inference for intractable likelihoods.

Nott, D.J., Marshall, L., Ngoc, T.M., 2012. The ensemble Kalman filter is an ABC algorithm. Stat. Comput. 22 (6), 1273–1276. http://dx.doi.org/10.1007/s11222-011-9300-x.

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A., Feldman, M.W., 1999. Population growth of human Y chromosomes: A study of Y chromosome microsatellites. Mol. Biol. Evol. 16 (12), 1791−1798. http://dx.doi.org/10.1093/oxfordjournals.molbev.a026091.

Roth, M., Hendeby, G., Fritsche, C., Gustafsson, F., 2017. The ensemble Kalman filter: A signal processing perspective. EURASIP J. Adv. Signal Process. 2017 (1), 56. http://dx.doi.org/10.1186/s13634-017-0492-x.

Tong, X.T., Majda, A.J., Kelly, D., 2016. Nonlinear stability and ergodicity of ensemble based Kalman filters. Nonlinearity 29 (2), 657–691. http://dx.doi.org/10.1088/0951-7715/29/2/657.

Vihola, M., Franks, J., 2020. On the use of approximate Bayesian computation Markov chain Monte Carlo with inflated tolerance and post-correction. Biometrika 107 (2), 381–395. http://dx.doi.org/10.1093/biomet/asz078.