Multilevel estimation of normalization constants using ensemble Kalman–Bucy filters

Hamza Ruzayqat · Neil K. Chada · Ajay Jasra

Received: 2 September 2021 / Accepted: 1 April 2022 / Published online: 4 May 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
In this article we consider the application of multilevel Monte Carlo, for the estimation of normalizing constants. In particular we will make use of the filtering algorithm, the ensemble Kalman–Bucy filter (EnKBF), which is an $N$-particle representation of the Kalman–Bucy filter (KBF). The EnKBF is of interest as it coincides with the optimal filter in the continuous-linear setting, i.e. the KBF. This motivates our particular setup in the linear setting. The resulting methodology we will use is the multilevel ensemble Kalman–Bucy filter (MLEnKBF). We provide an analysis based on deriving $L^q$-bounds for the normalizing constants using both the single-level, and the multilevel algorithms, which is largely based on previous work deriving the MLEnKBF Chada et al. (2022). Our results will be highlighted through numerical results, where we firstly demonstrate the error-to-cost rates of the MLEnKBFs comparing it to the EnKBF on a linear Gaussian model. Our analysis will be specific to one variant of the MLEnKBF, whereas the numerics will be tested on different variants. We also exploit this methodology for parameter estimation, where we test this on the models arising in atmospheric sciences, such as the stochastic Lorenz 63 and 96 model.

Keywords Multilevel Monte Carlo · Filtering · Kalman–Bucy filter · Normalizing constant · Parameter estimation

AMS subject classifications 60G35 · 62F15 · 65C05 · 62M20

1 Introduction
Filtering (Bain and Crisan 2009; Crisan and Rozovskii 2011; Del Moral 2004) is the mathematical discipline concerned with the conditional probability of an unobserved latent process, given sequentially observed data. It can be found in a wide array of applications, most notably; numerical weather prediction, mathematical finance, geophysical sciences and more recently machine learning (Bhar 2010; Majda and Wang 2006; Oliver et al. 2008). Mathematically, given a $d_x$-dimensional unobserved signal process $\{X_t\}_{t \geq 0}$, and a $d_y$-dimensional $\{Y_t\}_{t \geq 0}$ observed process, defined as

$$dY_t = h(X_t) dt + dV_t,$$
$$dX_t = f(X_t) dt + \sigma(X_t) dW_t,$$  \hspace{1cm} (1.1)  \hspace{1cm} (1.2)

the aim of the filtering problem is to compute the following expectation $\mathbb{E}[\varphi(X_t)|\mathcal{F}_t]$, where $\varphi : \mathbb{R}^{d_x} \to \mathbb{R}$ is an appropriately integrable function and $\{\mathcal{F}_t\}_{t \geq 0}$ is the filtration generated by the observed process (1.1). From (1.1)–(1.2), $V_t$ and $W_t$ are independent $d_y$ and $d_x$-dimensional Brownian motions respectively, with $h : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y}$ and $f : \mathbb{R}^{d_x} \to \mathbb{R}^{d_x}$ denoting potentially nonlinear functions, and $\sigma : \mathbb{R}^{d_x} \to \mathbb{R}^{d_x \times d_x}$ acting as a diffusion coefficient. Aside from computing the filtering distribution, filtering can also be exploited to compute normalizing constants associated with the filtering distribution (Brosse et al. 2018; Cerou et al. 2011; Gelman and Meng 1998; Kostov and Whiteley 2017; Rischard et al. 2018), i.e. the marginal likelihood, which is an important and useful computation in Bayesian statistics. It is useful as it can be used for model compari-
son which is commonly done through Bayes’ factor. These quantities are particularly useful in mixture models such as time-series state space models, and hierarchical models. Our motivation from this work is the estimation of the normalizing constants, associated with the filtering distribution in the continuous linear-Gaussian setting. It is well-known in this setting, that the optimal filter is the Kalman–Bucy filter (KBF) (Jazwinski 1970). However, the use of such a filter can be challenging to work with, such as firstly actually simulating such processes, or secondly the associated computational cost. As a result an alternative to this filter is the ensemble Kalman–Bucy filter (EnKBF), which is an $N$-particle representation of the KBF, or also as approximations of conditional McKean–Vlasov-type diffusion processes. This filter has been analyzed extensively of recent through various pieces of work which include, but not limited to, understanding stability, deriving uniform propagation of chaos bounds (Bishop and Del Moral 2020, 2017; Del Moral and Tugaut 2018) and acquiring a multilevel estimator (Chada et al. 2022; Giles 2008, 2015). However in the context of estimating normalizing constants, recent work has been done by Crisan et al. (2022) where the authors provide a way to do so using the EnKBF. However despite this, there is a computational burden associated with such costs, for example with the EnKBF or other Monte-Carlo algorithms, to attain a mean squared error (MSE) of a particular order. Therefore it would be of interest to apply techniques, to reduce the cost associated of attaining a particular MSE. This is the incentive behind the group of methods based on multilevel Monte Carlo (MLMC).

MLMC is a numerical technique for stochastic computation, concerned with reducing the computational complexity of Monte Carlo. Specifically it reduces the cost of attaining an MSE of an order $O(\epsilon^2)$, for $\epsilon > 0$. The ideas of MLMC go back to the original work of Giles (2008, 2015), which uses mesh refinements, and a telescoping sum property, to reduce the cost. Since then it has been applied to numerous applications, and disciplines, with filtering being one of them. Specifically for Kalman-based filtering this has been applied to the discrete-EnKF, with particular variants also analyzed, and applications arising in reservoir modeling (Chernov et al. 2021; Fossum et al. 2020; Hoel et al. 2016, 2021). However more recently there has been the extension to the continuous-time setting, for the EnKBF (Chada et al. 2022), entitled MLEnKBF.

In this article we are interested in developing multilevel estimators related to Kalman filtering, for the computation of normalizing constant. In particular we will propose the use of various MLEnKBFs for the application of normalizing constant estimation. Our motivation for this, as mentioned, is that firstly applying ML techniques can reduce the cost, associated to attaining an MSE of a particular order, compared to the normalizing constant estimator in Crisan et al. (2022). Due to this we will make use of the methodology and proof arguments discussed in Chada et al. (2022). Secondly as it coincides with KBF, it provides an incentive to work in the linear setting, which can be more cost-effective in terms of the cost of the algorithm, to other filtering techniques, which make use of the MLMC, such as the multilevel particle filter and sequential Monte Carlo sampler (Del Moral et al. 2017; Jasra et al. 2017, 2018). We emphasize with this paper, that we are not focused on such a comparison.

1.1 Contributions

Our contributions of this manuscript are highlighted through the following points:

- We introduce and approximate a multilevel estimator for the computation of normalizing constants associated with EnKBFs. This formulation is based on the multilevel EnKBF introduced by Chada et al. (2022), which looks to extend the single-level estimator which was proposed by Crisan et al. (2022).
- Through our formulation we provide firstly a propagation of chaos result for the normalizing constants associated with the EnKBF, which requires appropriate $L_q$-bounds, for $q \in [1, \infty)$. To achieve a MSE of order $O(\epsilon^2)$, we require a cost of $O(\epsilon^{-3})$, for $\epsilon > 0$. This is then extended to the multilevel setting, where we prove a similar result for an ‘ideal’, or i.i.d., system of the MLEnKBF. In particular to achieve a MSE of order $O(\epsilon^{-2})$, we require a cost of $O(\epsilon^{-2} \log(\epsilon))$, for $\epsilon > 0$. The analysis is specific to the vanilla variant, which will form our main result of this work and naturally follows from the analysis conducted in Chada et al. (2022).
- We verify the analysis derived in the paper for various numerical experiments. We firstly demonstrate the rates attained on an Ornstein–Uhlenbeck process, and provide parameter estimation on the stochastic Lorenz 63 and Lorenz 96 models. Furthermore we test other variants of the EnKBF, where the analysis does not directly apply, such as the deterministic and deterministic-transport variants, to see how they perform.

1.2 Outline

The outline of this paper is presented as follows. In Section 2 we review and discuss the KBF which will motivate the introduction of the EnKBF and the three variants we consider. This will lead onto Sect. 3 where we discuss normalizing constants associated to the filtering distribution, and how one can use the EnKBF and the MLEnKBF. Furthermore in this section we provide our main result and demonstrate it on a toy linear example. In Sect. 4 we provide an algorithm for parameter estimation using the MLEnKBF tested on various models arising in atmospheric sciences. We conclude with
some remarks and suggest future directions of work in the Sect. 5. Finally the proofs of our results are presented in the Appendix.

2 Model and background

In this section we provide an overview of the necessary background material required for the rest of the article. We begin by providing the continuous-time filtering problem through the Kalman–Bucy filters (KBF). This will lead onto a discussion of ensemble Kalman–Bucy filters (EnKBFs), which are \( N \)-particle representation of the KBF. Finally we will discuss the concept of multilevel Monte Carlo (MLMC), and review its extension to the EnKBF, referred to as the MLEnKBF.

2.1 Kalman–Bucy filters

Consider a linear-Gaussian filtering model of the following form

\[
dY_t = CX_t dt + R^{1/2}dV_t, \quad (2.1)
dX_t = AX_t dt + Q^{1/2}dW_t, \quad (2.2)
\]

where \((Y_t, X_t) \in \mathbb{R}^{d_y} \times \mathbb{R}^{d_x}, (V_t, W_t)\) is a \((d_y + d_x)\)-dimensional standard Brownian motion, \(A\) is a square \(d_x \times d_x\) matrix, \(C\) is a \(d_y \times d_x\) matrix, \(X_0 = 0, X_0 \sim N_{d_x}(\mu_0, \Sigma_0)\) \((d_x \times \text{dimensional Gaussian distribution, mean } \mu_0, \text{covariance matrix } \Sigma_0)\) and \(R^{1/2}, Q^{1/2}\) are square (of the appropriate dimension) and symmetric and invertible matrices. It is well-known that, letting \([\mathcal{F}_t]_{t \geq 0}\) be the filtration generated by the observations, the conditional probability of \(X_t\) given \(\mathcal{F}_t\) is a Gaussian distribution with mean and covariance matrix

\[
\mathcal{M}_t := \mathbb{E}[X_t|\mathcal{F}_t],
\]

\[
\mathcal{P}_t := \mathbb{E}[(X_t - \mathbb{E}(X_t|\mathcal{F}_t))(X_t - \mathbb{E}(X_t|\mathcal{F}_t))^\top],
\]

given by the Kalman–Bucy and Ricatti equations (Del Moral and Tugaut 2018)

\[
d\mathcal{M}_t = A\mathcal{M}_t dt + \mathcal{P}_t C^\top R^{-1} \left(dY_t - C\mathcal{M}_t dt\right), \quad (2.3)
d\mathcal{P}_t = \text{Ricc}(\mathcal{P}_t), \quad (2.4)
\]

where the Riccati drift term is defined as

\[
\text{Ricc}(G) = AG + GA^\top - GSG + Q,
\]

with \(S := C^\top R^{-1} C\).

A derivation of (2.3)-(2.4) can be found in Jazwinski (1970). KBF is viewed as the \(L_2\)-optimal state estimator for an Ornstein–Uhlenbeck process, given the state is partially observed with linear and Gaussian assumptions. An alternative approach is to consider a conditional McKean-Vlasov type diffusion process. For this article we work with three different processes of the form

\[
d\mathcal{X}_t = A \mathcal{X}_t dt + Q^{1/2} d\mathcal{W}_t + \mathcal{P}_\eta C^\top R^{-1} \left[dY_t - \left(C\mathcal{X}_t dt + R^{1/2} d\mathcal{V}_t\right)\right], \quad (2.5)
d\mathcal{X}_t = A \mathcal{X}_t dt + Q^{1/2} d\mathcal{W}_t + \mathcal{P}_\eta C^\top R^{-1} \left[dY_t - \left(\frac{1}{2} C [\mathcal{X}_t + \eta_t(e)]\right) dt\right], \quad (2.6)
d\mathcal{X}_t = A \mathcal{X}_t dt + Q\mathcal{P}_\eta^{-1} (\mathcal{X}_t - \eta_t(e)) dt + \mathcal{P}_\eta C^\top R^{-1} \left[dY_t - \left(\frac{1}{2} C [\mathcal{X}_t + \eta_t(e)]\right) dt\right], \quad (2.7)
\]

where \((\mathcal{V}_t, \mathcal{W}_t, \mathcal{X}_0)\) are copies of the process of \((V_t, W_t, X_0)\) and covariance

\[
\mathcal{P}_\eta = \eta_t \left( [e - \eta_t(e)][e - \eta_t(e)]^\top \right), \quad \eta_t := \text{Law}(\mathcal{X}_t|\mathcal{F}_t),
\]

such that \(\eta_t\) is the conditional law of \(\mathcal{X}_t\) given \(\mathcal{F}_t\) and \(e(x) = x\). We will explain the difference of each diffusion process, in succeeding subsections. It is important to note that the nonlinearity in (2.5)-(2.7) does not depend on the distribution of the state \(\text{Law}(\mathcal{X}_t)\) but on the conditional distribution \(\eta_t\), and \(\mathcal{P}_\eta\) alone does not depend on \(\mathcal{F}_t\). These processes are commonly referred to as Kalman–Bucy (nonlinear) diffusion processes. It is known that the conditional expectations of the random states \(\mathcal{X}_t\) and their conditional covariance matrices \(\mathcal{P}_\eta\), with respect to \(\mathcal{F}_t\), satisfy the Kalman–Bucy and the Riccati equations. In addition, for any \(t \in \mathbb{R}^+\)

\[
\eta_t := \text{Law}(\mathcal{X}_t|\mathcal{F}_t) = \text{Law}(X_t|\mathcal{F}_t).
\]

As a result, an alternative to recursively computing (2.3) - (2.4), is to generate \(N\) i.i.d. samples from any of (2.5)-(2.7) processes and apply a Monte Carlo approximation, as mentioned which we now discuss.

2.2 Ensemble Kalman–Bucy filters

Exact simulation from (2.5)-(2.7) is typically not possible, or feasible, as one cannot compute \(\mathcal{P}_\eta\) exactly. The ensemble Kalman–Bucy filter (EnKBF) can be used to deal with this issue. The EnKBF coincides with the mean-field particle interpretation of the Kalman-Bucy diffusion processes. EnKBFs is an \(N\)-particle system simulated for the \(i^{th}\)-particle, \(i \in \{1, \ldots, N\}\). The first variant of the EnKBF we consider is given as
which is known as the vanilla EnKBF (VEnKBF). This is the standard EnKBF used in theory and practice which contains perturbed observations, through the brownian motion $\bar{W}_i$. It is a continuous-time derivation of the EnKF (Evensen 2009). The second EnKBF given, defined as,
\[
d\xi^i_t = A \xi^i_t \, dt + Q^{1/2} \, d\bar{W}_t + P_i^N C^T R^{-1} \left[ dY_t - \left( C \xi^i_t + R^{1/2} \, d\bar{V}_t \right) \right],
\]
(2.8)

which is referred to as the deterministic EnKBF (DEnKBF), where unlike (2.8), it contains no perturbed observations. This is the continuous-limiting object of the deterministic EnKBF defined in Sakov and Oke (2008). It is well-known in ensemble data assimilation, that deterministic filters can perform better, partially due to the containing exact observations with no noise. Our final variant of the EnKBF is the deterministic transport EnKBF (DEnKBF),
\[
d\xi^i_t = A \xi^i_t \, dt + Q P_i^{-1} \left( \xi^i_t - M_t \right) \, dt + P_i^N C^T R^{-1} \left[ dY_t - \left( \frac{1}{2} C \left[ \xi^i_t + M_t \right] dt \right) \right],
\]
(2.9)

where the modification is that it does not contain $\bar{V}_t$ and $\bar{W}_t$, implying it is completely deterministic. This filter is motivated from the use of optimal transport methodologies within data assimilation (Reich 2019). Such processes are discussed in further detail in Bishop and Del Moral (2020). For all the variants of the EnKBF the sample mean and covariances are defined as
\[
\begin{align*}
P_i^N &= \frac{1}{N-1} \sum_{i=1}^{N} (\xi^i_t - m_i^N) (\xi^i_t - m_i^N)^\top, \\
m_i^N &= \frac{1}{N} \sum_{i=1}^{N} \xi^i_t,
\end{align*}
\]

where $\xi^i_0 \overset{i.i.d.}{\sim} \mathcal{N}_{d_i}(\mathcal{M}_0, \mathcal{P}_0)$. Note that when $C = 0$, (2.8) and (2.9) reduce to $N$–independent copies of an Ornstein–Uhlenbeck process.

In practice, one will not have access to an entire trajectory of observations. Thus numerically, one often works with a time discretization, such as an Euler-type discretization. Let $\Delta_t = 2^{-l}$ denote our level of discretization, then we will generate the system for $(i, k) \in \{1, \ldots, N\} \times \mathbb{N}_0 = \mathbb{N} \cup \{0\}$ as

Remarks

Remark 2.1 As mentioned in the introduction, the function $f$ in (1.2) can be nonlinear. If we assume in addition that $h$ is the identity function, then one can modify equations (2.8)–(2.10) by replacing the term $A X_t$ with $f(X_t)$ and equations (2.11)–(2.13) by replacing the term $A \xi^i_{k\Delta_t}$ with $f(\xi^i_{k\Delta_t})$.

Remark 2.2 When mentioning perturbed observations, we are referring to the additional noise in the innovation term. However, we emphasize it is not the observation being perturbed, but rather the projected state. hence once can think of this as a stochastic filter.

\[
\begin{align*}
\xi^i_{(k+1)\Delta_t} &= \xi^i_{k\Delta_t} + A \xi^i_{k\Delta_t} \Delta_t + Q^{1/2} \left[ \tilde{W}^i_{(k+1)\Delta_t} - \tilde{W}^i_{k\Delta_t} \right] \\
&+ P_{k\Delta_t}^N C^T R^{-1} \left( \{Y_{(k+1)\Delta_t} - Y_{k\Delta_t}\} - \left[ C \xi^i_{k\Delta_t} \Delta_t + R^{1/2} \left[ \tilde{V}_{(k+1)\Delta_t} - \tilde{V}^i_{k\Delta_t} \right] \right] \right),
\end{align*}
\]
(2.11)

\[
\begin{align*}
\xi^i_{(k+1)\Delta_t} &= \xi^i_{k\Delta_t} + A \xi^i_{k\Delta_t} \Delta_t + Q^{1/2} \left[ \tilde{W}^i_{(k+1)\Delta_t} - \tilde{W}^i_{k\Delta_t} \right] \\
&+ P_{k\Delta_t}^N C^T R^{-1} \left( \{Y_{(k+1)\Delta_t} - Y_{k\Delta_t}\} - C \left( \xi^i_{k\Delta_t} + m_{k\Delta_t}^N \right) \right) \Delta_t,
\end{align*}
\]
(2.12)

\[
\begin{align*}
\xi^i_{(k+1)\Delta_t} &= \xi^i_{k\Delta_t} + A \xi^i_{k\Delta_t} \Delta_t + Q \left( P_{k\Delta_t} \right)^{-1} \left( \xi^i_{k\Delta_t} - m_{k\Delta_t}^N \right) \Delta_t \\
&+ P_{k\Delta_t}^N C^T R^{-1} \left( \{Y_{(k+1)\Delta_t} - Y_{k\Delta_t}\} - C \left( \xi^i_{k\Delta_t} + m_{k\Delta_t}^N \right) \right) \Delta_t,
\end{align*}
\]
(2.13)

such that
\[
\begin{align*}
P_{k\Delta_t}^N &= \frac{1}{N-1} \sum_{i=1}^{N} (\xi^i_t - m_i^N) (\xi^i_t - m_i^N)^\top, \\
m_{k\Delta_t}^N &= \frac{1}{N} \sum_{i=1}^{N} \xi^i_t,
\end{align*}
\]

and $\xi^i_0 \overset{i.i.d.}{\sim} \mathcal{N}_{d_i}(\mathcal{M}_0, \mathcal{P}_0)$. For $l \in \mathbb{N}_0$ given, denote by $\eta^N_{l\Delta_t}$ as the $N$–empirical probability measure of the particles $(\xi^i_0, \ldots, \xi^i_N)$, where $t \in \{0, \Delta_t, 2\Delta_t, \ldots\}$. For $\varphi : \mathbb{R}^{d_i} \to \mathbb{R}^{d_i}$ we will use the notation $\eta^N_{l\Delta_t}(\varphi) := \frac{1}{N} \sum_{i=1}^{N} \varphi(\xi^i_t)$.

Remark 2.3 For ensemble based data assimilation, a common recurring assumption is that the particles are i.i.d. While this is true, through ensemble filtering methodologies the assumption is practically false. This is based on the fact that ensemble covariances are computed from all ensemble members, which introduces a dependence. Furthermore when the EnKF is applied to nonlinear functions, this can result in the ensemble not being normally distributed.
2.3 Multilevel EnKBFs

Let us define \( \pi \) to be a probability on a measurable space \((X, \mathcal{X})\) and for \( \pi \)-measurable \( \varphi : X \to \mathbb{R} \) consider the problem of estimating \( \pi(\varphi) = \mathbb{E}_\pi[\varphi(X)] \). Now let us assume that we only have access to a sequence of approximations of \( \pi \), \( \{\pi_l\}_{l \in \mathbb{N}_0} \), also each defined on \( \mathbb{R} \), and we are now interested in estimating \( \pi_l(\varphi) \), such that \( \lim_{l \to \infty} [\pi_l - \pi](\varphi) = 0 \). Therefore one can use the telescoping sum

\[
\pi_L(\varphi) = \pi_0(\varphi) + \sum_{l=1}^{L} [\pi_l - \pi_{l-1}](\varphi), \tag{2.14}
\]

as we know that the approximation error between \( \pi \) and \( \pi_l \) gets smaller as \( l \to \infty \). The idea of multilevel Monte Carlo (MLMC) (Giles 2008, 2015) is to construct a coupled system, related to the telescoping sum, such that the mean squared error can be reduced, relative to i.i.d. sampling from \( \pi_L \). Therefore our MLMC approximation of\( \mathbb{E}_{\pi_L}[\varphi(X)] \) is

\[
\pi^{\text{ML}}_L(\varphi) := \frac{1}{N_0} \sum_{i=1}^{N_0} \varphi(X_i,0) + \sum_{l=1}^{L} \frac{1}{N_l} \sum_{i=1}^{N_l} (\varphi(X^{l,i}, \varphi(\tilde{X}^{l,i-1})),
\]

where \( N_0 \in \mathbb{N} \) i.i.d. samples from \( \pi_0 \) as \( (X^{1,0}, \ldots, X^{N_0,0}) \) and for \( l \in \{1, \ldots, L\}, N_l \in \mathbb{N} \) samples from a coupling of \( (\pi_l, \pi_{l-1}) \) as \( ((X^{1,l}, \tilde{X}^{1,l-1}), \ldots, (X^{N_l,l}, \tilde{X}^{N_l,l-1})) \). Then the MSE is then

\[
\mathbb{E}[\pi^{\text{ML}}_L(\varphi) - \pi(\varphi)^2] = \text{var}[\pi^{\text{ML}}_L(\varphi)] + [\pi_L(\varphi) - \pi(\varphi)]^2.
\]

It is important to note that the telescoping sum (2.14), despite its simplicity and usefulness, is not an optimal choice to construct the multilevel estimator. In particular Giles (2015) makes this assumption, but states that a potentially more robust weighted sum would be more optimal. The work of Chada et al. (2022) proposed the application of MLMC for the VEnKBF, whom we will now briefly review this. If we consider the discretized VEnKBF (2.11), then, for \( l \in \mathbb{N} \) and \((i, k) \in \{1, \ldots, N\} \times \mathbb{N}_0 \) the ML adaption is given as

\[
\xi^{i,l}_{(k+1),\Delta_l} = \xi^{i,l}_{k,\Delta_l} + A^{i,l}_{k,\Delta_l} \Delta_l + Q^{1/2} [W^{(k+1),\Delta_l} - W^{k,\Delta_l}] \\
+ p^{N,l}_{k,\Delta_l} C^T R^{-1} \left( Y^{(k+1),\Delta_l} - Y^{k,\Delta_l} \right) \\
- C \left( \xi^{i,l}_{k,\Delta_l} + m^{N,l}_{k,\Delta_l} \right) \Delta_l, \tag{F1}
\]

\[
\xi^{i,l-1}_{(k+1),\Delta_{l-1}} = \xi^{i,l-1}_{k,\Delta_{l-1}} + A^{i,l-1}_{k,\Delta_{l-1}} \Delta_{l-1} \\
+ Q^{1/2} [W^{(k+1),\Delta_{l-1}} - W^{k,\Delta_{l-1}}] \\
+ p^{N,l-1}_{k,\Delta_{l-1}} C^T R^{-1} \left( Y^{(k+1),\Delta_{l-1}} - Y^{k,\Delta_{l-1}} \right) \\
- C \left( \xi^{i,l-1}_{k,\Delta_{l-1}} + m^{N,l-1}_{k,\Delta_{l-1}} \right) \Delta_{l-1}, \tag{F2}
\]

Similarly for the deterministic variant (2.12) we have

\[
\xi^{i,l}_{(k+1),\Delta_l} = \xi^{i,l}_{k,\Delta_l} + A^{i,l}_{k,\Delta_l} \Delta_l + Q^{1/2} [W^{(k+1),\Delta_l} - W^{k,\Delta_l}] \\
+ p^{N,l}_{k,\Delta_l} C^T R^{-1} \left( Y^{(k+1),\Delta_l} - Y^{k,\Delta_l} \right) \\
- C \left( \xi^{i,l}_{k,\Delta_l} + m^{N,l}_{k,\Delta_l} \right) \Delta_l, \tag{F3}
\]

and finally for the deterministic-transport variant (2.13)

\[
\xi^{i,l-1}_{(k+1),\Delta_{l-1}} = \xi^{i,l-1}_{k,\Delta_{l-1}} + A^{i,l-1}_{k,\Delta_{l-1}} \Delta_{l-1} \\
+ Q^{1/2} [W^{(k+1),\Delta_{l-1}} - W^{k,\Delta_{l-1}}] \\
+ p^{N,l-1}_{k,\Delta_{l-1}} C^T R^{-1} \left( Y^{(k+1),\Delta_{l-1}} - Y^{k,\Delta_{l-1}} \right) \\
- C \left( \xi^{i,l-1}_{k,\Delta_{l-1}} + m^{N,l-1}_{k,\Delta_{l-1}} \right) \Delta_{l-1},
\]

where our sample covariances and means are defined accordingly as

\[
P^{N,l}_{k,\Delta_l} = \frac{1}{N-1} \sum_{i=1}^{N} (\xi^{i,l}_{k,\Delta_l} - m^{N,l}_{k,\Delta_l})(\xi^{i,l}_{k,\Delta_l} - m^{N,l}_{k,\Delta_l})^T,
\]

\[
m^{N,l}_{k,\Delta_l} = \frac{1}{N} \sum_{i=1}^{N} \xi^{i,l}_{k,\Delta_l},
\]

\[
P^{N,l-1}_{k,\Delta_{l-1}} = \frac{1}{N-1} \sum_{i=1}^{N} (\xi^{i,l-1}_{k,\Delta_{l-1}} - m^{N,l-1}_{k,\Delta_{l-1}})(\xi^{i,l-1}_{k,\Delta_{l-1}} - m^{N,l-1}_{k,\Delta_{l-1}})^T,
\]

\[
m^{N,l-1}_{k,\Delta_{l-1}} = \frac{1}{N} \sum_{i=1}^{N} \xi^{i,l-1}_{k,\Delta_{l-1}},
\]

and \( \xi_{0}^{i,l \Delta_0} \sim \mathcal{N}_{d_i}(M_0, P_0) \), \( \xi_{0}^{i,l-1} = \xi_{0}^{i,l} \). Then, one has the approximation of \( [\eta_l - \eta_{l-1}]^{-1}(\varphi) \), \( l \in \mathbb{N}_0 \), \( \varphi : \mathbb{R}^{d_x} \to \mathbb{R} \), given as
with the filtering distribution. This is defined as

\[ \eta^N_{t,i} - \eta^N_{t,i-1} (\varphi) = \frac{1}{N} \sum_{i=1}^{N} [\varphi(\xi^N_{t,i}) - \varphi(\xi^N_{t,i-1})]. \]

Therefore for the multilevel estimation, one has an approximation for \( t \in \mathbb{N}_0 \)

\[ \eta^{ML}_{t} (\varphi) := \eta^{N,0}_{t} (\varphi) + \sum_{l=1}^{L} [\eta^{N,l}_{t} - \eta^{N,l-1}_{t}] (\varphi). \] (2.15)

Similarly with Remark 2.3, the i.i.d. assumption regarding each particle does not also hold in the multilevel setting. Largely due to the correlation of the particle in the multilevel setting.

**Remark 2.4** For the above forms of the MLEnKBF, only (F1) - (F2) were introduced and numerically tested in Chada et al. (2022). Therefore this article aims to further test the deterministic-transport variant of (F3).

### 3 Normalizing constant estimation

In this section we introduce the notion of normalizing constants (NC), which is what we are concerned with estimating. We begin by recalling the NC estimator using the EnKBF, aiming to show its cost associated to attain an MSE of a particular order. We then present our NC estimator through the MLEnKBF. This will lead onto our main result of the paper, which is the error-to-cost ratio of our normalizing constant estimator, i.e. the MLEnKBF, which is presented through Theorem 3.1.

#### 3.1 Normalizing constants

We begin by defining the normalizing constant associated with the filtering distribution. This is defined as \( Z_t := \mathcal{L}_{X_t,Y_t}^{X_0,Y_0} \) to be the density of \( \mathcal{L}_{X_t,Y_t} \), the law of the process \((X,Y)\) and that of \( \mathcal{L}_{X_0,Y_0} \), the law of the process \((X,W)\). That is,

\[ \mathbb{E}[f(X_{0,t})g(Y_{0,t})] = \mathbb{E}[f(X_{0,t})g(W_{0,t})Z_t(X,Y)]. \]

One can show that (see Exercise 3.14 pp 56 in Bain and Crisan (2009))

\[ Z_t(X,Y) = \exp \left[ \int_0^t \left( \langle CX_s, R^{-1}dY_s \rangle - \frac{1}{2} \langle X_s, SX_s \rangle \right) ds \right]. \] (3.1)

recalling that \( S := C^T R^{-1} C \). Now we let \( Z_t(Y) \) denote the likelihood function defined by

\[ Z_t(Y) := \mathbb{E}_Y (Z_t(X,Y)), \]

where \( \mathbb{E}_Y (\cdot) \) stands for the expectation w.r.t. the signal process when the observation is fixed and independent of the signal. From the work of Crisan et al. (2022), the authors show that the normalizing constant is given by

\[ Z_t(Y) = \exp \left[ \int_0^t \left( \langle CM_s, R^{-1}dY_s \rangle - \frac{1}{2} \langle M_s, SM_s \rangle ds \right) \right], \] (3.2)

and a sensible estimator of it is given as

\[ Z^N_t(Y) = \exp \left[ \int_0^t \left( \langle CM^N_s, R^{-1}dY_s \rangle - \frac{1}{2} \langle M^N_s, SM^N_s \rangle ds \right) \right], \] (3.3)

which follows from replacing the conditional mean of the signal process with the sample mean associated with the EnKBF. To briefly describe how (3.2) was attained in Crisan et al. (2022), the authors used a change of measure rule, through Girsanov’s Theorem, then an application of Bayes’ Theorem which can be expressed through as the Kallianpur-Striebel formula defined on path space. The derivation is a standard approach, which is discussed in Chapter 3 in Bain and Crisan (2009).

In practice, one must time-discretize the EnKBF, and the normalizing constant estimator (3.3) to yield for \( t \in \mathbb{N} \)

\[ Z^N_t(Y) = \exp \left\{ \sum_{k=0}^{t\Delta t - 1} \left[ \langle CM_k^N, R^{-1}(Y_{(k+1)\Delta t} - Y_{k\Delta t}) \rangle - \frac{\Delta t}{2} \langle m^N_{k\Delta t}, Sm^N_{k\Delta t} \rangle \right] \right\}. \] (3.4)

Let \( \overline{Z}^N_t(Y) = \log(Z^N_t(Y)) \), where we now consider the estimation of log-normalization constants. To enhance the efficiency we consider a coupled ensemble Kalman–Bucy filter, as described in Sect. 2. Let \( l \in \mathbb{N} \) then we run the coupled system of the different MLEnKBFs.

**Remark 3.1** In order to use the analysis derived in Chada et al. (2022), for the normalizing constant estimation, our results will be specific to the vanilla MLEnKBF variant (F1). This is important as we will use the results presented in Chada et al. (2022), however, in terms of the numerics, we will use all ML variants described for various parameter estimation experiments.

#### 3.2 Single-level EnKBF

To present a multilevel NC estimator using the EnKBF, we first require to understand the single-level EnKBF. To aid our analysis we will consider the i.i.d. particle system based
upon the Euler discretization of EnKBF for (F1), such that for \((i, k) \in \{1, \ldots, N\} \times \mathbb{N}_0\)

\[
\zeta_{i(k+1)\Delta t} = (I + A\Delta t)\zeta_{i(k)\Delta t} + Q^{1/2}[\overline{W}_{(k+1)\Delta t} - \overline{W}_{k\Delta t}]
+ P_{k\Delta t}C^TR^{-1}\left([Y_{(k+1)\Delta t} - Y_{k\Delta t}]
- C\zeta_{i(k+1)\Delta t} + R^{1/2}[^V_{(k+1)\Delta t} - \overline{V}_{k\Delta t}]\right),
\]

(3.5)
such that \(\zeta_{i(k+1)\Delta t} | \mathcal{F}_{(k+1)\Delta t} \stackrel{i.i.d.}{\sim} \mathcal{N}_{d_x}(m_{(k+1)\Delta t}, \sigma_{(k+1)\Delta t})\), where the moments are defined as

\[
m_{(k+1)\Delta t} = m_{k\Delta t} + Am_{k\Delta t}\Delta t + P_{k\Delta t}C^TR^{-1}\left([Y_{(k+1)\Delta t} - Y_{k\Delta t}] - Cm_{k\Delta t}\Delta t\right),
\]

(3.6)

\[
P_{(k+1)\Delta t} = P_{k\Delta t} + \text{Ricc}(P_{k\Delta t})\Delta t + (A - P_{k\Delta t}S)P_{k\Delta t}(AT - SP_{k\Delta t})\Delta t^2,
\]

(3.7)

which are satisfied by the Kalman–Bucy diffusion (2.5). We now consider the NC estimator \(\widehat{\mathcal{O}}_{t+k\Delta t}\), instead using the recursion of (3.5).

**Remark 3.2** Our reason for introducing the new estimator, based on (3.5)–(3.7), is that the usual described estimator of \(\mathcal{U}^T_{ML}(Y)\) will not hold, in terms of the analysis, due to the recursion of the EnKBF, which is mentioned and documented in Chada et al. (2022). However, as we will see by using a number of arguments one can show that both the multilevel estimators are close in probability, as \(N \to \infty\). Therefore our analysis will be specific to ML NC estimator \(\widehat{\mathcal{O}}_{t+k\Delta t}\) i.e.

\[
\widehat{\mathcal{O}}^N_{i(k+1)\Delta t} = \widehat{\mathcal{O}}^N_{i(k)\Delta t} + \sum_{l=i}^{L} (\widehat{\mathcal{O}}^N_{i(l+1)\Delta t} - \widehat{\mathcal{O}}^N_{i(l)\Delta t}),
\]

while we continue to use \(\widehat{\mathcal{O}}^T_{ML}(Y)\) for the numerical examples.

We now present our first result of the paper which is a propagation of chaos result for the single-level vanilla EnKBF NC estimator. For a \(d_x\)-dimensional vector \(x\) we denote \(\|x\|_2 = (\sum_{j=1}^{d_x} x(j)^2)^{1/2}\), where \(x(j)\) is the \(j^{th}\)-element of \(x\). We use the notation \([\mathcal{O}^N_{i(t+k\Delta t)} - \mathcal{O}^N_{i(t+k\Delta t)}] = \mathcal{O}^N_{i(t+k\Delta t)}(Y) - \mathcal{O}^N_{i(t+k\Delta t)}(Y)\).

**Proposition A.1** For any \((l, t, k) \in \mathbb{N}_0 \times \mathbb{R}^+ \times \{0, 1, \ldots, \Delta_t^{-1}\}\) almost surely:

\[
\lim_{N \to \infty} [\mathcal{O}^N_{i(t+k\Delta t)} - \mathcal{O}^N_{i(t+k\Delta t)}](Y) = 0.
\]

The proof follows by using Proposition A.1 in the appendix and the Marcinkiewicz–Zygmund inequality for i.i.d. random variables, along with a standard first Borel–Cantelli lemma argument.

Then using the same arguments, as described in Chada et al. (2022), one can show, through the Markov inequality that

\[
P\left(\left[\mathcal{O}^N_{i(t+k\Delta t)} - \mathcal{O}^N_{i(t+k\Delta t)}\right] > \varepsilon\right) \leq \frac{C}{\varepsilon^2 N^{-2}.}
\]

(3.8)

for any \(\varepsilon > 0\) and \(q > 0\), where \(C\) is a constant that depends on \((l, q, t, k)\) but not \(N\). Related to this in Proposition C.1, in the appendix, we have shown that:

\[
P\left(\sqrt{\mathcal{O}^N_{i(t+k\Delta t)} - \mathcal{O}^N_{i(t+k\Delta t)}}(Y) \geq \varepsilon\right) \leq \frac{C}{\varepsilon^2 N^{-2}}.
\]

(3.9)

\[
\mathcal{O}^N_{i(t+k\Delta t)} - \mathcal{O}^N_{i(t+k\Delta t)}(Y) = \mathcal{O}^N_{i(t+k\Delta t)}(Y) - \mathcal{O}^N_{i(t+k\Delta t)}(Y).
\]

(3.10)

Now we set

\[
\widehat{\mathcal{O}}^N_{i(t)\Delta t} := \widehat{\mathcal{O}}^{N,0}_{i(t)}(Y) + \sum_{l=1}^{L} (\widehat{\mathcal{O}}^N_{i(l+1)\Delta t} - \widehat{\mathcal{O}}^N_{i(l)\Delta t})(Y).
\]

(3.11)

Therefore using the same argument as in Chada et al. (2022), one can establish, almost surely

\[
\lim_{N \to \infty} (\mathcal{O}^N_{i(t\Delta t)} - \mathcal{O}^N_{i(t\Delta t)})(Y) = 0
\]

\[
P\left(\left[\mathcal{O}^N_{i(t\Delta t)} - \mathcal{O}^N_{i(t\Delta t)}\right](Y) > \varepsilon\right) \leq \frac{C}{\varepsilon^2 N^{-2}}.
\]
where \( C \) is a constant that can depend on \( (L, q, t) \) but not \( N_{0:L} \).

We are now in a position to state our main result, concerned with using the vanilla MLEnKBF, for the i.i.d. system, as a NC estimator.

**Theorem 3.1** For any \( T \in \mathbb{N} \) fixed and \( t \in [0, T - 1] \) there exists a \( C < +\infty \) such that for any \( (L, N_{0:L}) \in \mathbb{N} \times \{2, 3, \ldots\}^{L+1} \):

\[
E \left[ \left\| \hat{U}_T^{ML} - U_T \right\|^2 \right] 
\leq C \left( \sum_{l=0}^{L} \frac{\Delta_l}{N_l} + \sum_{l=1}^{L} \sum_{q=1,q \neq l}^{L} \frac{\Delta_l \Delta_q}{N_l N_q} + \Delta_L^2 \right).
\]

The above theorem translates as, in order to achieve an MSE of order \( O(\varepsilon^2) \), for \( \varepsilon > 0 \), we have a cost of \( O(\varepsilon^{-2} \log(\varepsilon)^2) \). This implies a reduction in cost compared to the single-level NC estimator.

**Remark 3.3** It is important to note that the rates obtained in Proposition A.1 and Theorem 3.1 for the normalizing constants, are identical to those derived in Chada et al. (2022). However despite this it should be emphasized as we will see in the appendix, that the proofs are not trivial, and all of them do not directly follow exactly.

## 4 Numerical simulations

In this section we provide various numerical experiments, which include both verifying the ML rates obtained, and for parameter estimation on both linear and nonlinear models. The former will be tested on an linear example, which the later will include a toy linear Gaussian example, a stochastic Lorenz 63 model and Lorenz 96 model, which are common models that arise in atmospheric sciences. Our parameter estimation will be conducted through a combination of using recursive maximum likelihood estimation, and simultaneous perturbation stochastic approximation, which is a gradient-free methodology.

### 4.1 Verification of multilevel rates

We now seek to verify our theory from Proposition A.1 and Theorem 3.1 on an Ornstein–Uhlenbeck process, taking the form of (2.1)–(2.2). We will compare the error-to-cost rates of the MLEnKBF estimation versus the EnKBF estimation of log-normalizing constant. This will be numerically tested using all variants (F1) - (F3), and their respective multilevel counterparts. We will take \( d_x = d_y = 5 \), \( A = -0.8 \times I_d \), \( \lambda_0 = 0.1 \times I_d \), \( P_0 = 0.05 \times I_d \), \( R^{1/2} = 2 \times I_d \), \( C \) to be a random matrix and \( Q^{1/2} \) is a tri-diagonal matrix defined as

\[
Q^{1/2} = \begin{bmatrix}
2/3 & 1/3 & \ldots & 0 \\
1/3 & 2/3 & \ddots & \\
\vdots & \ddots & \ddots & 1/3 \\
0 & \ldots & 1/3 & 2/3
\end{bmatrix},
\]

where \( I_d \) is the identity matrix of appropriate dimension and \( \mathbf{1} \) is a vector of ones. We computed the MSE for the target levels \( L \in \{7, 8, 9, 10\} \) using 416 simulations with \( \varepsilon = 1 \). In order to better estimate the quantity in (3.4), we used \( l_a = 6 \). The number of samples on each level is given by

\[
N_l = \lfloor C_0 2^{2L-l} (L - l_a + 1) \rfloor.
\]

The reference value is the mean of 208 simulations computed at a discretization level \( l = 11 \). For the NC estimator, using the EnKBF, the cost of the estimator is \( N_l \Delta_{L,l}^{-1} \). When using the MLEnKBF-NC, the cost of the estimator is \( \sum_{l=0}^{L} N_l \Delta_{L,l}^{-1} \).

\( \text{ Springer} \)
Fig. 1 MSE vs Cost plots on a log-log scale for the linear Gaussian model (2.1)–(2.2), comparing EnKBF and MLEnKBF estimators of the normalizing constant

Our numerical results are presented in Fig. 1, where we have plotted the ML and single-level (SL) rates for each variant, within each subplot. Let us first consider (F1), which is the vanilla MLEnKBF. The results obtained match that from both Proposition A.1 and Theorem 3.1, which suggest error-to-cost rates of $O(\epsilon^{-3})$ and $O(\epsilon^{-2} \log(\epsilon)^2)$. This is indicated through the slopes, which show that to attain an order of MSE, the MLEnKBF-NC is computationally cheaper than that of the EnKBF-NC. Our next subplot concerns (F2). Despite no existing theory for this case, or its SL counterpart, we see that the results are similar to the vanilla variant. Specifically, rates given for both slopes are similar, where there is little distinction. Finally for the final variant of (F3), the results obtained are quite different. Firstly we notice the slopes are more different to the other variants, but also that the theory of the rate $O(\epsilon^{-3})$ does not hold as convincingly. This could be related to the lack of stochasticity within the methodology. However from all subplots one can see that by applying MLMC, one does attain a lower cost for a particular order of MSE.

4.2 Parameter estimation

Let us firstly assume that the model (1.1)–(1.2) contains unknown parameters $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$. To account for this, we rewrite the normalizing constant (3.2) with the additional subscript $\theta$ as $Z_{t,\theta}(Y)$. To estimate these parameters we focus on maximum likelihood inference and stochastic gradient methods that are performed in an online manner. Mainly, we follow a recursive maximum likelihood (RML) method, which has been proposed originally in Arapostathis and Marcus (1990), in Le Gland and Mevel (1997) for finite spaces, and in Beskos et al. (2021), Del Moral et al. (2010), Poyiadjis et al. (2011) in the context of sequential Monte Carlo (SMC) approximations. Let

$$U_{t,t+1,\theta}(Y) := \log \frac{Z_{t+1,\theta}(Y)}{Z_{t,\theta}(Y)}.$$ 

RML relies on the following update scheme at any time $t \in \mathbb{N}$:

$$\theta_{t+1} = \theta_t + a_t \left( \nabla_\theta \log Z_{t+1,\theta_t}(Y) - \nabla_\theta \log Z_{t,\theta_t}(Y) \right)$$

where $\{a_t\}_{t \in \mathbb{N}}$ is a sequence of positive real numbers such that we assume the usual Robbins–Munro conditions, i.e. $\sum_{t \in \mathbb{N}} a_t = \infty$ and $\sum_{t \in \mathbb{N}} a_t^2 < \infty$. Given an initial $\theta_0 \in \Theta$, this formula enables us to update $\theta$ online as we obtain a new observation path in each unit time interval. Computing
the gradients in the above formula can be expensive (see e.g. Beskos et al. 2021), therefore we use a gradient-free method that is based on some type of finite differences with simultaneous perturbation stochastic approximation (SPSA) (Spall 1992, 2003). In a standard finite difference approach, one perturbs $\theta_t$ in the positive and negative directions of a unit vector $\vec{e}_k$ (a vector of zeros in all directions except $k$ it is 1). This means evaluating $U_{t,t+1,\theta_t}(Y)$ 2$d\theta$-times.

Whereas in SPSA, we perturb $\theta_t$ with a magnitude of $b_t$ in the positive and negative directions of a $d\theta$-dimensional random vector $\Psi_t$. The numbers $(b_t)_{t \in \mathbb{N}}$ are a sequence of positive real numbers such that $b_t \to 0$, $\sum_{t \in \mathbb{N}} a_t^2 / b_t^2 < \infty$, and for $k \in \{1 \cdots, d\}$, $\Psi_t(k)$ is sampled from a Bernoulli distribution with success probability 1/2 and support $\{-1, 1\}$. Therefore, this method requires only 2 evaluations of $U_{t,t+1,\theta_t}(Y)$ to estimate the gradient. The MLEnKBF-NC estimator, presented in Algorithm 1, was used to estimate the ratio $U_{t,t+1,\theta_t}(Y)$. In Algorithm 2 we illustrate how to implement these approximations in order to estimate the model’s static parameters.

### Algorithm 2 Parameter Estimation: using MLEnKBF-NC and RML-SPSA

1. **Input**: Target level $L \in \mathbb{N}$, start level $l_s \in \mathbb{N}$ such that $l_s < L$, the number of particles on each level $\{N_l\}_{l=1}^L$, the number of iterations $M \in \mathbb{N}$, initial $\theta_u \in \Theta$, step size sequences of positive real numbers $(a_l)_{l \in \mathbb{N}}$, $(b_l)_{l \in \mathbb{N}}$ such that $a_l \to 0$, $\sum_{l \in \mathbb{N}} a_l^2 b_l^2 < \infty$, and initial ensembles $\{\xi_l^{N_l}\}_{l=1}^L \sim \mathcal{N}(\Theta_0, \Psi_0)$, where $N_{tot} = \sum_{l=1}^L N_l$.

2. **Iterate**: For $t \in \{0, \cdots, M - 1\}$:
   - Set $\{\xi_l^{N_l}\}_{l=1}^L = \{\xi_0^{N_0}\}_{l=1}^L$, $\{\xi_l^{N_l}\}_{l=1}^L = \{\xi_l^{N_l}\}_{l=1}^L$.
   - For $k \in \{1, \cdots, d\}$, sample $\Psi_t(k)$ from a Bernoulli distribution with success probability 1/2 and support $\{-1, 1\}$.
   - Set $\theta_t^+ = \theta_t + b_t \Psi_t$, $\theta_t^- = \theta_t - b_t \Psi_t$.
   - Run Algorithm 1 twice, with $T = 1$ and initial ensembles $\{\xi_l^{N_l}\}_{l=1}^L, \{\xi_l^{N_l}\}_{l=1}^L$ to generate the estimates $U_{t,t+1,\theta_t^+}(Y)$ and $U_{t,t+1,\theta_t^-}(Y)$.
   - Set for $k \in \{1, \cdots, d\}$, $\theta_{t+1}(k) = \eta_t (k) + \frac{a_{t+1}}{2b_{t+1}} \Psi_{t+1}(k) \left[ U_{t,t+1,\theta_t^+}(Y) - U_{t,t+1,\theta_t^-}(Y) \right]$.
   - Run the EnKBF up to time 1 under the new parameter $\theta_{t+1}$ with discretization level $L$ and initial ensembles $\{\xi_l^{N_l}\}_{l=1}^L$.

We will now introduce our different models we work with. This includes a linear and two nonlinear models, namely, the stochastic Lorenz 63 and Lorenz 96.

### 4.2.1 Linear Gaussian model

Our first numerical example is a linear Gaussian model with $d_x = d_y = 2$, based on (2.1)–(2.2). Specifically, we choose the parameter values $A = \theta_1 I_d$ and $Q^{1/2} = \theta_2 Q$ where $Q$ is a tri-diagonal matrix defined as

$$Q = \begin{bmatrix}
1 & 1/2 & 0 & \ldots & 0 \\
1/2 & 1 & 1/2 & \ddots & \\
0 & \ddots & \ddots & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & 1/2 \\
0 & \cdots & \cdots & 0 & 1/2 & 1
\end{bmatrix}$$

We also take $C$ to be a uniform random matrix and $R^{1/2} = 0.556 I_d$. The parameters of interest we aim to estimate are $(\theta_1, \theta_2) \in \mathbb{R}^2$. We choose the target level as $L = 9$ and the start level as $l_s = 7$, and specify an initial state of $X_0 \sim \mathcal{N}(4l, I_d)$, where $l$ is a vectors of 1’s. In all cases, (F1) - (F3), we take

$$N_l = \lfloor 0.04 \cdot 2^{L-l} (L - l_s + 1) \rfloor.$$  

The results are displayed in Fig. 2. In order to reduce the variance between the different simulations, we used the same Brownian increments (in the cases (F1) and (F2)) that were needed to generate $U_{t,t+1,\theta_t^+}(Y)$ and $U_{t,t+1,\theta_t^-}(Y)$ as well. We applied this trick only on the linear model, but it can be extended to the nonlinear models below. We observe from Fig. 2 that all the variants learn of the true parameters $(\theta_1^*, \theta_2^*) = (-2, 1)$, and from the bottom panel, (F3) has the smallest standard deviation.

**Remark 4.1** When testing (F3) in the linear example, there were difficulties in learning $\theta_2$. This is because there is no natural coupling, hence why we had to choose a specific step-size as described in Fig. 2. However for the nonlinear examples we did not experience this, where no specific modification of the step-size was required.

### 4.2.2 Stochastic Lorenz 63 model

Our next example is the Lorenz 63 model (Lorenz 1963) with $d_x = d_y = 3$, which is a model for atmospheric convection. The model is based on three ordinary differential equations, where now we have three parameters of interest to estimate, i.e. $(\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3$. The stochastic Lorenz 63 model is given as

$$dX_t = f(X_t)dt + Q^{1/2}dW_t,$$
$$dY_t = CX_t dt + R^{1/2}dV_t,$$
such that

\[ f_1(X_t) = \theta_1(X_t(2) - X_t(1)), \]
\[ f_2(X_t) = \theta_2 X_t(1) - X_t(2) - X_t(1)X_t(3), \]
\[ f_3(X_t) = X_t(1)X_t(2) - \theta_3 X_t(3), \]

where \( X_t(i) \) is the \( i^{th} \) component of \( X_t \). Furthermore we have that \( Q^{1/2} = Id \) and the variable \( C \) is specified as

\[ C = \begin{cases} 
\frac{1}{2}, & \text{if } i = j, \\
\frac{1}{2}, & \text{if } i = j-1, \ j \in \{1, 2, 3\}, \\
0, & \text{otherwise},
\end{cases} \]

and \( (R^{1/2})_{ij} = 2 q \left( \frac{2}{3} \min(|i-j|, r_2 - |i-j|) \right) \), \( i, j \in \{1, 2, 3\} \) such that

\[ q(x) = \begin{cases} 
1 - \frac{3}{2} x + \frac{1}{2} x^3, & \text{if } 0 \leq x \leq 1, \\
0, & \text{otherwise}.
\end{cases} \]

In Fig. 3, we show the results for the parameters estimation of \((\theta_1, \theta_2, \theta_3)\) using the cases (F1) - (F3). We set the target level to be \( L = 9 \), the start level \( l_s = 7 \), and specify the initial state \( X_0 \sim N(1, 0.5 \ Id) \). The number of samples on each level are the same as in (4.2). From Fig. 3 it is clear that all variants of the MLEnKBF perform similarly for this inference problem. Interestingly we notice that the learning of the parameter \( \theta_2^* = 8/3 \) seems the most accurate, while taking the least amount of time to reach the value of \( \theta_2^* \). This is unlike the learning of the other parameters, which require at least a time of \( T = 5000 \) to get close to \( \theta_1^* \) and \( T = 3000 \) to get close to \( \theta_2^* \). Part of the reason for this could be that \( \theta_3 \) acts as a coefficient only for \( X_t(3) \), unlike for \( X_t(1) \) which depends on both \( \theta_1 \) and \( \theta_2 \). In addition, we see that the learning of \( \theta_2^* \) is biased towards 27.6 in all variants.

### 4.2.3 Stochastic Lorenz 96 model

Our final test model will be the Lorenz 96 model (Lorenz 1996) with \( d_x = d_y = 40 \), which is a dynamical system designed to describe equatorial waves in atmospheric science. The stochastic Lorenz 96 model takes the form

\[ dX_t = f(X_t)dt + Q^{1/2}dW_t, \]
Fig. 3 Lorenz 63 results: The outcomes of running Algorithm 2 for the estimation of \((\theta_1, \theta_2, \theta_3)\) in the cases (F1) (top), (F2) (middle) and (F3) (bottom). The black curve is the average of 6 independent runs and the shaded area is the mean ± the standard deviation. The initial values of the parameters are \((6, 27, 6.5)\). The dashed lines represent the true parameters values \((\theta_1^*, \theta_2^*, \theta_3^*) = (10, 28, 8/3)\). In all cases, we set \(b_t = t^{-0.1}\) for all \(t \in \mathbb{N}\), \(a_t = 0.01\) when \(t \leq 100\) and \(a_t = t^{-0.75}\) for \(t > 100\).

As we observe from Fig. 4, all MLEnKBF variants learn the true value of \(\theta^* = 8\). The fluctuations again can be seen more clearly in the first two variants, as they are stochastic, i.e. contain either or both of \(Q\) and \(R\). This is not the case for (F3) whose dynamics are completely deterministic. In general the results are more stable that those conducted for the Lorenz 63 model, as the function \(f: \mathbb{R}^{d_x} \to \mathbb{R}^{d_x}\), does not contain \(\theta\) as a coefficient of the process \(X_t\).

\[dY_t = CX_t dt + R^{1/2}dV_t,\]

such that

\[f_i(X_t) = (X_t(i + 1) - X_t(i - 2))X_t(i - 1) - X_t(i) + \theta,\]

where again \(X_t(i)\) is the \(i^{th}\) component of \(X_t\), and we assume that \(X_t(-1) = X_t(d_x - 1), X_t(0) = X_t(d_x)\) and that \(X_t(d_x + 1) = X_t(1)\). We specify our parameter values as \(Q^{1/2} = \sqrt{2}Id\) and \(R^{1/2} = 0.5 Id\). The parameter \(\theta\) is the external force in the system, while \((X_t(i + 1) - X_t(i - 2))X_t(i - 1)\) is the advection term and \(-X_t(i)\) is the damping term. In Fig. 4, we show the results for the parameter estimation of \(\theta\) using the cases (F1)-(F3). We set the target level to be \(L = 9\), and the start level \(l_0 = 7\). In (F1) and (F2), we specify the initial state as follows: We set \(X_0(1) = 8.01\) and \(X_0(i) = 8\) for \(1 < i \leq d_x\). In (F3) case, to prevent the matrix \(P_0^{N,l}\) from being equal to zero, we set \(X_0 \sim \mathcal{N}(8 1, 0.05 Id)\). The number of samples on each level are the same as in (4.2).

5 Conclusion

The purpose of this work was to apply MLMC strategies for normalizing constant (NC) estimation. In particular our aim was to extend the work of Crisan et al. (2022), which used EnKBF, to its multilevel counterpart which is the MLEnKBF (Chada et al. 2022). As stated, our motivation is primarily in the linear setting, for which the optimal filter is the Kalman–Bucy filter, where we could provide propagation of
Fig. 4 Lorenz 96 results: The outcomes of running Algorithm 2 for the estimation of \( \theta \) in the cases (F1) (left), (F2) (middle) and (F3) (right). The black curve is the average of 6 independent runs and the shaded area is the mean ± the standard deviation. The initial value of \( \theta \) is 10.

The dashed line represents the true value \( \theta^* = 8 \). In all cases, we set \( b_t = t^{-0.1} \) for all \( t \in \mathbb{N} \), \( a_t = 0.03 \) when \( t \leq 50 \) and \( a_t = t^{-0.75} \) for \( t > 50 \).

This work naturally leads to different fruitful and future directions of work to consider. The first potential direction is to aim to provide the same analysis for other MLEnKBFs, such as the deterministic and deterministic-transport EnKBF (Bishop and Del Moral 2020). This is of interest as from the numerical results, they suggest alternative rates for (F3), different to that of the other variants. One could also provide an unbiased estimation of the NC (Rischard et al. 2018), which has connections to MLMC (Chada et al. 2021; Vihola 2018). In order to do so, one would require a modified multilevel analysis of the EnKBF, where one has uniform upper bounds, with respect to levels \( l = 1, \ldots, L \). As stated, a comparison of both normalizing estimators using both the MLEnKBF and MLPF would be interesting, but to make such a comparison, one could exploit advanced methodologies, such as in Beskos et al. (2021), and appropriate examples in low dimensions. Finally it would be of interest to develop theory for the multilevel estimator \( \overline{U}_t^{ML} \). We have only tested this computationally which seems to match the rates attained for the i.i.d. (ideal) ML estimator. To do so, one requires more sophisticated mathematics to hold initially for the problem-setting in Chada et al. (2022).

Acknowledgements This work was supported by KAUST baseline funding.

Appendix

A Analysis for EnKBF NC estimator

For the appendices, Appendix A will cover the propagation of chaos result, which is required for the variance of the single-level EnKBF NC estimator. We will then proceed to Appendix B which discusses various discretization biases of the diffusion process related to both EnKBF and the NC estimator. Finally our main theorem is proved in Appendix
C. All of our results will be specific to the vanilla variant of the ENKBF, $\mathbf{F}(1)$.

Before proceeding to our results, we will introduce the following assumptions which will hold from herein, but not be added to any statements. For a square matrix, $B$ say, we denote by $\mu(B)$ as the maximum eigenvalue of $\text{Sym}(B)$.

1. We have that $\mu(A) < 0$.
2. There exists a $C < +\infty$ such that for any $(k, l) \in \mathbb{N}_0^2$ we have that
   \[
   \max_{(j_1, j_2) \in \{1, \ldots, d_i\}^2} |P_{k\Delta_i}(j_1, j_2)| \leq C. \tag{A.1}
   \]

We note that 1. is typically used in the time stability of the hidden diffusion process $X_t$, see for instance (Del Moral and Tugaut 2018). In the case of 2. we expect that it can be verified under 1., that $S = C I$ with $C$ a positive constant and some controllability and observability assumptions (e.g. (Del Moral and Tugaut 2018, eq. (20))). Under such assumptions, the Riccati equation has a solution and moreover, by Del Moral and Tugaut (2018, Proposition 5.3) $\mathcal{P}_t$ is exponentially stable w.r.t. the Frobenius norm; so that this type of bound exists in continuous time.

Throughout the appendix we will make use of the $C_q$-inequality. For two real-valued random variables $X$ and $Y$ defined on the same probability space, with expectation operator $\mathbb{E}$, suppose that for some fixed $q \in (0, \infty)$, $\mathbb{E}[|X|^q]$ and $\mathbb{E}[|Y|^q]$ are finite, then the $C_q$-inequality is
\[
\mathbb{E}[|X + Y|^q] \leq C_q \left( \mathbb{E}[|X|^q] + \mathbb{E}[|Y|^q] \right),
\]
where $C_q = 1$, if $q \in (0, 1)$ and $C_q = 2^{q-1}$ for $q \in [1, \infty)$. In order to verify some of our claims for the analysis, we will rely on various results derived in Chada et al. (2022). For convenience-sake we will state these below, which are concerned with various $\|\cdot\|_q$-bounds, where

**Lemma A.1** For any $q \in [1, \infty)$ there exists a $C < +\infty$ such that for any $(k, l, j) \in \mathbb{N}_0^3 \times \{1, \ldots, d_i\}$:
\[
\mathbb{E}[|Y_{(k+1)\Delta_i} - Y_{k\Delta_i}|(j)|^q]^{1/q} \leq C \Delta_i^{1/2}.
\]

**Lemma A.2** For any $(q, t, k, l) \in [1, \infty) \times \mathbb{N}_0^3$ there exists a $C < +\infty$ such that for any $(j, N) \in \{1, \ldots, d_i\} \times \{2, 3, \ldots\}$:
\[
\mathbb{E}[|m^{N}_{t+k\Delta_i}(j) - m_{t+k\Delta_i}(j)|^q]^{1/q} \leq \frac{C}{\sqrt{N}}.
\]

**Lemma A.3** For any $(q, k, l) \in (0, \infty) \times \mathbb{N}_0^2$ there exists a $C < +\infty$ such that for any $N \geq 2$ and $i \in \{1, \ldots, N\}$:
\[
\mathbb{E}[|\xi^{i}_{k\Delta_i}(j)|^q]^{1/q} \leq C,
\]
where $\xi^{i}_{k\Delta_i}$ is defined through (2.11).

We now present our first result for the single-level EnKBF NC estimator, which is presented as an $\|\cdot\|_q$-error bound.

**Proposition A.1** For any $(q, t, k_1, l) \in [1, \infty) \times \mathbb{N}_0^3$ there exists a $C < +\infty$ such that for any $N \in \{2, 3, \ldots\}$ we have:
\[
\mathbb{E}\left[\left\|\mathbb{U}_{t+k_1\Delta_i}(Y) - \mathbb{U}_{t+k_1\Delta_i}(Y)\right\|_q^{1/q}\right] \leq \frac{C}{\sqrt{N}}.
\]

**Proof** Let us first consider $\mathbb{U}_{t+k_1\Delta_i}(Y) - \mathbb{U}_{t+k_1\Delta_i}(Y)$, which for every $l \in \mathbb{N}_0$, we can decompose through a martingale remainder-type decomposition,
\[
\mathbb{U}_{t+k_1\Delta_i}(Y) - \mathbb{U}_{t+k_1\Delta_i}(Y) = M^l_{t+k_1\Delta_i}(Y) + R^l_{t+k_1\Delta_i}, \tag{A.2}
\]
such that
\[
M^l_{t+k_1\Delta_i}(Y) = \sum_{k=0}^{t\Delta_i^{-1} + k_1-1} (C m^N_{k\Delta_i} - R^{-1}[Y_{(k+1)\Delta_i} - Y_{k\Delta_i}]) - \sum_{k=0}^{t\Delta_i^{-1} + k_1-1} (C m_{k\Delta_i} - R^{-1}[Y_{(k+1)\Delta_i} - Y_{k\Delta_i}]),
\]
\[
R^l_{t+k_1\Delta_i} = \frac{\Delta_i}{2} \sum_{k=0}^{t\Delta_i^{-1} + k_1-1} (m^N_{k\Delta_i}, S^N_{k\Delta_i}) + \frac{\Delta_i}{2} \sum_{k=0}^{t\Delta_i^{-1} + k_1-1} (m_{k\Delta_i}, S_{k\Delta_i}).
\]
We can decompose the martingale term from (A.2) further through
\[
M^l_{t+k_1\Delta_i}(1) = M^l_{t+k_1\Delta_i}(1) + R^l_{t+k_1\Delta_i}(1),
\]
where, by setting $k = t\Delta_i^{-1} + k_1 - 1$,
\[
M^l_{t+k_1\Delta_i}(1) = \sum_{k=0}^{t\Delta_i^{-1} + k_1-1} \sum_{j_1=1}^{d_i} \sum_{j_2=1}^{d_i} \sum_{j_3=1}^{d_i} C(j_1, j_2, j_3)
\]
\[
\times [(Y_{(k+1)\Delta_i} - Y_{k\Delta_i})(j_3) - C X_{k\Delta_i}(j_3) D^{N}_{\Delta_i}],
\]
\[
R^l_{t+k_1\Delta_i}(1) = \Delta_i \sum_{k=0}^{t\Delta_i^{-1} + k_1-1} \sum_{j_1=1}^{d_i} \sum_{j_2=1}^{d_i} \sum_{j_3=1} S^{N}_{j_1, j_2, j_3}
\]
\[
\times R^{-1}(j_1, j_3) C X_{k\Delta_i}(j_3). \tag{A.3}
\]
In order to proceed we construct a martingale associated with the term of \( M_t^j \). Let us first begin with the \( M_t^j(1) \) term (A.3), where we construct the filtration \((\Omega, \mathcal{F}, \mathcal{F}_{k\Delta_t}, \mathbb{P})\) for our discrete-time martingale \((M_t^j(1), \mathcal{F}_{k\Delta_t})\).

Then by holding Hölder’s inequality

\[
\mathbb{E}[|M_{t+k\Delta_t}(1)|^{q}]^{1/q} = \mathbb{E}
\left[
\sum_{k=0}^{t\Delta_t^{-1}+k\Delta_t-1} \sum_{j_1=j_2=1}^{j_1} \sum_{j_1=1}^{j_1} C(j_1, j_2)
\{m_{k\Delta_t}(j_2) - m_{k\Delta_t}(j_2)\} R^{-1}(j_1, j_3)
\right]^{1/q}
\leq \mathbb{E}
\left[
\sum_{k=0}^{t\Delta_t^{-1}+k\Delta_t-1} \sum_{j_1=j_2=1}^{j_1} \sum_{j_1=1}^{j_1} C(j_1, j_2)
\{m_{k\Delta_t}(j_2) - m_{k\Delta_t}(j_2)\} R^{-1}(j_1, j_3)
\right]^{1/2q}
\times \mathbb{E}
\left[
\sum_{k=0}^{t\Delta_t^{-1}+k\Delta_t-1} \sum_{j_1=j_2=1}^{j_1} \sum_{j_1=1}^{j_1} C(j_1, j_2)
\sum_{j_1=1}^{j_1} (|Y(k+1)\Delta_t - Y(k)\Delta_t|) (j_3) - C X_{k\Delta_t}(j_3) \Delta t
\right]^{2q}^{1/2q}
\leq \Delta \times T_1 \times T_2.
\] (A.5)

For \( T_1 \) we can apply the Minkowski inequality and Lemma A.2 to yield

\[
T_1 \leq \sum_{k=0}^{t\Delta_t^{-1}+k\Delta_t-1} \sum_{j_1=j_2=1}^{j_1} \sum_{j_1=1}^{j_1} C(j_1, j_2) R^{-1}(j_1, j_3)
\left\{m_{k\Delta_t}(j_2) - m_{k\Delta_t}(j_2)\right\}^{2q}^{1/2q}
\leq \frac{C}{\sqrt{N}}.
\]

For \( T_2 \) we know that the expression \( \{Y(k+1)\Delta_t - Y(k)\Delta_t\}(j_3) - C X_{k\Delta_t}(j_3) \Delta t \) is a Brownian motion increment, using the formulae (2.1)–(2.2). Therefore by using the Burkholder–Davis–Gundy inequality, along with Minkowski, for \( \tilde{q} = 2q \), we have

\[
T_2 \leq \sum_{j_3=1}^{j_3} \mathbb{E}
\left[
\sum_{k=0}^{t\Delta_t^{-1}+k\Delta_t-1} |V(k+1)\Delta_t - V(k)\Delta_t|^{2q}^{1/\tilde{q}}
\left\{|Y(k+1)\Delta_t - Y(k)\Delta_t|^{2q}^{1/\tilde{q}}
\right\}
\leq \sum_{j_3=1}^{j_3} \mathbb{E}
\left[
\sum_{k=0}^{t\Delta_t^{-1}+k\Delta_t-1} |V(k+1)\Delta_t - V(k)\Delta_t|^{2q}^{1/\tilde{q}}
\right]
\leq \frac{C}{\sqrt{N}} \mathbb{E}[|Q(k\Delta_t)|^{2q/1}]
\]

Then using that fact that \( \mathbb{E}[|V(k+1)\Delta_t - V(k)\Delta_t|^{2q}] = O(\Delta_t^{\tilde{q}/2}) \), and with the summation it is of order \( O(\Delta_t^{1/2-1/\tilde{q}}) \), we can conclude \( T_2 \) is of order \( O(1) \), as we have used the case of when \( q \geq 1 \).

For the \( R_t^j(1) \) term, it follows similarly to \( M_t^j(1) \), where we require the use of Lemma A.2.

Again we make use of the Minkowski and Hölder inequality, and Lemma A.2,

\[
\mathbb{E}[|R_{t+k\Delta_t}(1)|^{q}]^{1/q} = \mathbb{E}
\left[
\sum_{k=0}^{t\Delta_t^{-1}+k\Delta_t-1} \sum_{j_1=j_2=1}^{j_1} \sum_{j_1=1}^{j_1} C(j_1, j_2)
\{m_{k\Delta_t}(j_2) - m_{k\Delta_t}(j_2)\} R^{-1}(j_1, j_3)
\right]^{1/q}
\leq \sum_{k=0}^{t\Delta_t^{-1}+k\Delta_t-1} \sum_{j_1=j_2=1}^{j_1} \sum_{j_1=1}^{j_1} C(j_1, j_2) R^{-1}(j_1, j_3)
\left\{m_{k\Delta_t}(j_2) - m_{k\Delta_t}(j_2)\right\}^{2q}^{1/2q}
\times \mathbb{E}
\left[
\sum_{k=0}^{t\Delta_t^{-1}+k\Delta_t-1} \sum_{j_1=j_2=1}^{j_1} \sum_{j_1=1}^{j_1} C(j_1, j_2)
\left\{m_{k\Delta_t}(j_2) - m_{k\Delta_t}(j_2)\right\}^{2q}^{1/2q}
\left\{X_{k\Delta_t}(j_3)\Delta_t\right\}^{2q}^{1/2q}
\right]
\leq \sum_{k=0}^{t\Delta_t^{-1}+k\Delta_t-1} \sum_{j_1=j_2=1}^{j_1} \sum_{j_1=1}^{j_1} \frac{C}{\sqrt{N}} \mathbb{E}[|X_{k\Delta_t}(j_3)\Delta_t|^{2q}^{1/2q}].
\] (A.6)

For the final term of (A.6) we can show it is of order \( O(1) \), using the Cauchy–Schwarz and Jensen’s inequality

\[
\mathbb{E}[|X_{k\Delta_t}(j_3)\Delta_t|^{2q}^{1/2q}]
\leq \Delta t \mathbb{E}
\left[
\int_{k\Delta_t}^{(k+1)\Delta_t} \left|X_{s}(j_3)\Delta_t\right|^{2q}^{1/2q}
\right]
\leq \Delta t \mathbb{E}
\left[
\int_{k\Delta_t}^{(k+1)\Delta_t} \left|X_{s}(j_3)\Delta_t\right|^{2q}^{1/2q}
\right]
\leq \Delta t \mathbb{E}
\left[
\int_{k\Delta_t}^{(k+1)\Delta_t} \left|X_{s}(j_3)\Delta_t\right|^{2q}^{1/2q}
\right]
\leq \frac{C}{\sqrt{N}} \Delta t
\]

therefore combing this with the summation in (A.6), the above quantity is of order \( O(1) \), resulting in \( \mathbb{E}[|R_{t+k\Delta_t}(1)|^{q}]^{1/q} \leq \frac{C}{\sqrt{N}} \). All that is left is the \( R_{t+k\Delta_t}(1) \) term. Before proceeding we can express, or rewrite, the \( R_{t+k\Delta_t}(1) \) term as

\[
R_{t+k\Delta_t} = -\Delta t \sum_{k=0}^{t\Delta_t^{-1}+k\Delta_t-1} \left\langle m_{k\Delta_t}, m_{k\Delta_t} \right\rangle .
\]
By taking its associated $L_q$-bound, from Minkowski’s inequality and Lemma A.2 - Lemma A.3, we have

$$
\mathbb{E}[|R^l_{i+1}\Delta t|^q]^{1/q} \\
= \mathbb{E}
\left[
-\frac{\Delta t}{2} \sum_{k=0}^{t \Delta t - 1} \sum_{j=1}^{d_k} \sum_{j=1}^{d_k} (m^N_{k \Delta t}(j) - m_{k \Delta t}(j)) S(j, j) m^N_{k \Delta t}(j)
\right]^{1/q} \\
\leq \frac{\Delta t}{2} \sum_{k=0}^{t \Delta t - 1} \sum_{j=1}^{d_k} \sum_{j=1}^{d_k} |S(j, j)|^{1/q} \\
\leq \frac{C}{\sqrt{N}}.
$$

Finally by using the Minkowski inequality, we can deduce that

$$
\mathbb{E}
\left[
|\bar{U}_{t+\Delta t}^N - \bar{U}_{t+\Delta t}^l|^{q}\right]^{1/q} \\
= \mathbb{E}
\left[
|\bar{M}_{t+\Delta t}^N(1) + \bar{R}_{t+\Delta t}^N(1) + \bar{R}_{t+\Delta t}^l|^{q}\right]^{1/q} \\
\leq \mathbb{E}[|\bar{M}_{t+\Delta t}^N(1)|^{q}]^{1/q} \\
+ \mathbb{E}[|\bar{R}_{t+\Delta t}^N(1)|^{q}]^{1/q} + \mathbb{E}[|\bar{R}_{t+\Delta t}^l|^{q}]^{1/q} \\
\leq \frac{C}{\sqrt{N}}.
$$

\[ \square \]

### B Analysis for discretized diffusion process

In this appendix we consider deriving analysis for the discretized i.i.d. particle system

$$
\xi_{t+\Delta t}^i = (I + A \Delta t) \xi_t^i + Q^{1/2} [\bar{W}_{t+\Delta t}^i - \bar{W}_t^i] \\
+ P_{k \Delta t} C^T R^{-1} (Y_{t+\Delta t} - Y_k) \\
- \left[C \xi_t^i \Delta t + R^{1/2} [\bar{V}_{t+\Delta t}^i - \bar{V}_t^i]\right],
$$

and the discretized NC estimator. We recall, in the limit as $N \to \infty$, the i.i.d. system coincides with discretized Kalman–Bucy diffusion whose mean, for $(k, l) \in \mathbb{N}_0 \times \mathbb{N}_0$, is defined by

$$
m^l_{(k+1)\Delta t} = m^l_{k \Delta t} + A m^l_{k \Delta t} \Delta t + U^l_{k \Delta t} \left(Y_{(k+1)\Delta t} - Y_k\right) - C m^l_{k \Delta t} \Delta t.
$$

We note in this appendix our results will use the notation $\bar{X}$ for the Kalman–Bucy diffusion, to keep it consistent with (Chada et al. 2022). However these results also hold for the i.i.d. system (B.1). We require additional lemmas from Chada et al. (2022), which are discretization bias results for the discretized Kalman–Bucy diffusion. We state these as follows. The notation of the equations are modified for the multilevel, which we will discuss later.

**Lemma B.1** For any $T \in \mathbb{N}$ fixed and $t \in [0, T]$ there exists a $C < +\infty$ such that for any $(l, j, j_1) \in \mathbb{N}_0 \times \{1, \ldots, d_l\}^2$:

$$
|P_l(j_1, j_2) - P^l_{t \Delta t}(j_1, j_2)| \leq C \Delta t.
$$

**Lemma B.2** For any $T \in \mathbb{N}$ fixed and $t \in [0, T - 1]$ there exists a $C < +\infty$ such that for any $(l, j, k_1) \in \mathbb{N}_0 \times \{1, \ldots, d_l\} \times \{0, 1, \ldots, \Delta t^{-1}\}$:

$$
\mathbb{E}
\left[
\left|\bar{X}_{t+\Delta t}(j) - \bar{X}_{t+\Delta t}(j_1)\right|^2
\right] \leq C \Delta t^2.
$$

We now present our first conditional bias result, which will be the weak error of the Kalman–Bucy diffusion. This weak error will be analogous to the strong error of Lemma B.2, which was not proved, or provided, in Chada et al. (2022). However this result will be required for the current and succeeding appendix.

**Lemma B.3** For any $T \in \mathbb{N}$ fixed and $t \in [0, T - 1]$ there exists a $C < +\infty$ such that for any $(l, k_1) \in \mathbb{N}_0 \times \{1, \ldots, d_l\} \times \{0, 1, \ldots, \Delta t^{-1}\}$:

$$
\mathbb{E}
\left[
\left|\bar{X}_{t+\Delta t}(j) - \bar{X}_{t+\Delta t}(j_1)\right|
\right] \leq C \Delta t.
$$
Proof As before we can separate the above expression in different terms,

\[ \mathbb{E}\left[ \overline{X}_{t+k_1\Delta t}(j) - \overline{X}'_{t+k_1\Delta t}(j) \right] = T_1 + T_2 + T_3, \]

such that, for \( t' = \lfloor \frac{t}{\Delta t} \rfloor \Delta t, t \in \mathbb{R}^+ \), we have

\[ T_1 = \mathbb{E}\left[ \int_0^{t+k_1\Delta t} \left( \sum_{j_1=1}^{d_s} A(j, j_1) \overline{X}_s(j_1) - \overline{X}'_{t}(j_1) \right) \right], \]

\[ T_2 = \mathbb{E}\left[ \int_0^{t+k_1\Delta t} \left( \sum_{j_1=1}^{d_s} \sum_{j_2=1}^{d_s} P_s(j, j_1) \mathcal{C}(j_1, j_2) \overline{X}_s(j_2) - \overline{X}'_{t}(j_2) \right) \right], \]

\[ T_3 = \mathbb{E}\left[ \sum_{j_1=1}^{d_s} \sum_{j_2=1}^{d_s} \int_0^{t+k_1\Delta t} \left( P_s(j, j_1) - P_{t'}(j, j_1) \right) \mathcal{C}(j_1, j_2) d\overline{X}_s(j_2) \right]. \]

where \( \mathcal{C} = \mathcal{C} C \) with \( \mathcal{C} = C^T R^{-1} \). Now let us consider each individual term first for \( T_1 \), that (A.1) we have the following bound

\[ |T_1| \leq C \int_0^{t+k_1\Delta t} \max_{j_1 \in \{1, \ldots, d_s\}} \mathbb{E}\left[ \overline{X}_s(j) - \overline{X}'_{t}(j) \right] ds. \] (B.3)

For \( T_2 \), we can apply Lemma B.1, using the fact that \( \max_{j_2 \in \{1, \ldots, d_s\}} \mathbb{E}[\mathcal{C}(\overline{X}'_{t}(j_2)) \right] \leq C \) we have

\[ |T_2| \leq C \Delta t. \] (B.4)

Similarly for \( T_3 \), we can use Lemma B.1 and Lemma A.1 which provides the bound

\[ |T_3| \leq C \Delta t. \] (B.5)

Thus combining (B.3)-(B.5) leads to

\[ \max_{j_1 \in \{1, \ldots, d_s\}} \mathbb{E}\left[ \overline{X}_{t+k_1\Delta t}(j) - \overline{X}'_{t+k_1\Delta t}(j) \right] \leq C \Delta t \]

Finally by applying Grönwall's lemma, leads to the desired result. □

We now proceed with our result of the discretized NC estimator, which is the strong error, through the following lemma.

**Lemma B.4** For any \( T \in \mathbb{N} \) fixed and \( t \in [0, T - 1] \) there exists a \( C < +\infty \) such that for any \((l, k_1) \in \mathbb{N}_0 \times \{0, 1, \ldots, \Delta^{-1}_T\}\

\[ \mathbb{E}\left[ \left( \left| \overline{U}_{t+k_1\Delta t}(Y) - \overline{U}_{t+k_1\Delta t}(Y) \right| \right)^2 \right] \leq C \Delta_t^2. \]

**Proof** Let us first recall that,

\[ \overline{U}_{t+k_1\Delta t}(Y) = \sum_{k=0}^{t\Delta^{-1}_T+k_1-1} \langle C m_{k \Delta t}, R^{-1}(Y_{k+1\Delta t} - Y_{k\Delta t}) \rangle - \frac{\Delta t}{2} \sum_{k=0}^{t\Delta^{-1}_T+k_1-1} \langle m_{k \Delta t}, S m_{k \Delta t} \rangle, \]

\[ \overline{U}_{t+k_1\Delta t}(Y) = \int_0^{t+k_1\Delta t} \left[ \langle C m_s, R^{-1}dY_s \rangle - \frac{1}{2} \langle m_s, S m_s \rangle ds \right]. \]

In order to proceed we again consider a martingale-remainder type decomposition. Therefore by setting \( t' = \lfloor \frac{t}{\Delta t} \rfloor \Delta t, t \in \mathbb{R}^+ \), and expanding on the angle brackets, we have

\[ M_{t+k_1\Delta t}(1) = \sum_{j_1=1}^{d_s} \sum_{j_2=1}^{d_s} \sum_{j_3=1}^{d_s} \int_0^{t+k_1\Delta t} R^{-1}(j_1, j_3) C(j_1, j_2) \left[ m_s(j_2) - m_{t'}^{l_2}(j_2) \right] R^{1/2}(j_1, j_3) dV_s(j_3), \]

\[ R_{t+k_1\Delta t} = \sum_{j_1=1}^{d_s} \sum_{j_2=1}^{d_s} \int_0^{t+k_1\Delta t} S(j_1, j_2) \left[ m_s(j_1)m_s(j_2) - m_{t'}^{l_1}(j_1)m_{t'}^{l_2}(j_2) \right] ds, \]

\[ R_{t+k_1\Delta t}(1) = \sum_{j_1=1}^{d_s} \sum_{j_2=1}^{d_s} \sum_{j_3=1}^{d_s} \int_0^{t+k_1\Delta t} R^{-1}(j_1, j_3) C(j_1, j_2) \left[ m_s(j_1) - m_{t'}^{l_1}(j_1) \right] C X_s(j_3) ds, \]

where we have used the formula for the observational process (2.1), combined both remainder terms into one, and taken the scaled Brownian motion \( V_t \). Let us first consider the remainder term of \( R_{t+k_1\Delta t}(1) \). Through Jensen's inequality we have

\[ \mathbb{E}\left[ R_{t+k_1\Delta t}(1) \right]^2 = \mathbb{E}\left[ \sum_{j_1=1}^{d_s} \sum_{j_2=1}^{d_s} \sum_{j_3=1}^{d_s} \int_0^{t+k_1\Delta t} R^{-1}(j_1, j_3) C(j_1, j_2) \left[ m_s(j_2) - m_{t'}^{l_2}(j_2) \right] C X_s(j_3) ds \right]^2 \]

\[ \leq \sum_{j_1=1}^{d_s} \sum_{j_2=1}^{d_s} \sum_{j_3=1}^{d_s} \sum_{j_4=1}^{d_s} R^{-2}(j_1, j_3) C^2(j_1, j_2) \int_0^{t+k_1\Delta t} \mathbb{E}\left[ \left( m_s(j_2) - m_{t'}^{l_2}(j_2) \right) C X_s(j_4) \right]^2 ds. \]
Then by using $d_1^2 d_2^2$ applications of the $C_2$-inequality we get

$$\mathbb{E}[R_{t+k_1 \Delta t}(1)]^2 \leq C \left( \sum_{j_1=1}^{d_1} \sum_{j_2=1}^{d_2} \int_0^{t+k_1 \Delta t} S(j_1, j_2) \right)^2$$

$$\mathbb{E}[X_s(j_1)]^2 ds$$

We know that $\max_{j_2 \in \{1, \ldots, d_1\}} \mathbb{E}[X_s(j_4)]^2 \leq C$, therefore all we need is to bound $|m_s(j_1) - m'_t(j_1)|^2$. Therefore by using the fact that $\mathbb{E}[X_s(j)] = m_s(j)$ and $\mathbb{E}[X'_t(j)] = m'_t(j)$, we can use the weak error, i.e. Lemma B.3, to conclude that $\mathbb{E}[R_{t+k_1 \Delta t}(1)]^2 \leq C \Delta t^2$.

Now to proceed with $R_{t+k_1 \Delta t}(Y)$, we can split the difference of the mean term

$$m_s(j_1) m_s(j_2) - m'_t(j_1) m'_t(j_2)$$

$$= m_s(j_1) m_s(j_2) - m'_t(j_1) m'_t(j_2)$$

$$+ m'_t(j_1) m_s(j_2) - m'_t(j_2) m'_t(j_2)$$

$$= (m_s(j_1) - m'_t(j_1)) m_s(j_2) + m'_t(j_1) (m_s(j_2) - m'_t(j_2)).$$

(B.6)

Therefore one can substitute (B.6) into $R_Y(Y)$, and by Jensen’s and the $C_2$-inequality, results in

$$\mathbb{E}[R_{t+k_1 \Delta t}(1)]^2 \leq C \max_{j_2 \in \{1, \ldots, d_1\}} \mathbb{E}[X_s(j_2)]^2 ds,$$

$$\text{Through the same substitution as before, and using the weak error, i.e. Lemma B.3, we have that} \mathbb{E}[R_{t+k_1 \Delta t}(1)]^2 \leq C \Delta t^2.$$  

Lastly we have the martingale term $M_{t+k_1 \Delta t}(1)$. As before we can apply Jensen’s inequality

$$\mathbb{E}[M_{t+k_1 \Delta t}(1)]^2 = \mathbb{E} \left[ \sum_{j_1=1}^{d_1} \sum_{j_2=1}^{d_2} \int_0^{t+k_1 \Delta t} R^{-1}(j_1, j_2) C(j_1, j_2) \right]$$

$$\mathbb{E}[m_s(j_2) - m'_t(j_2)] ds.$$

Then by using the Ito isometry and $d_1^2 d_2^2$ applications of the $C_2$-inequality, we have

$$\mathbb{E}[M_{t+k_1 \Delta t}(1)]^2 \leq C \int_0^{t+k_1 \Delta t} \mathbb{E} \left[ \sum_{j_1=1}^{d_1} \sum_{j_2=1}^{d_2} \int_0^{t+k_1 \Delta t} R^{-1}(j_1, j_2) C(j_1, j_2) \right]$$

$$\mathbb{E}[m_s(j_2) - m'_t(j_2)] ds.$$  

and, as before, by using Lemma B.3 as done for $R_{t+k_1 \Delta t}$, we can conclude that $\mathbb{E}[M_{t+k_1 \Delta t}(1)]^2 \leq C \Delta t^2$. Therefore by combining all terms and a further application of the $C_2$-inequality three times

$$\mathbb{E} \left[ \left( \overline{U}_{t+k_1 \Delta t}(Y) - \overline{U}'_{t+k_1 \Delta t}(Y) \right)^2 \right]$$

$$\leq \mathbb{E} \left[ M_{t+k_1 \Delta t}(1) \right]^2 + \mathbb{E} \left[ R_{t+k_1 \Delta t}(1) \right]^2 + \mathbb{E} \left[ R_{t+k_1 \Delta t}(1) \right]^2$$

$$\leq C \Delta t^2.$$

□

C Analysis for i.i.d. MLEnKBF NC estimator

We now discuss the analysis, related to the variance, of both the NC estimator using the EnKBF and the i.i.d. MLEnKBF. This will lead onto the proof of our main result, presented as Theorem 3.1. We note that in our notations, we extend the case of the discretized EnKBF, to the discretized MLEnKBF, by adding superscripts $l$ as above. Specifically the analysis now considers the i.i.d. couple particle system

$$\zeta^{i,l}_{(k+1) \Delta t} = \zeta^{i,l}_{k \Delta t} + A \zeta^{i,l}_{k \Delta t} + Q^{1/2} \left[ \overline{W}_{(k+1) \Delta t} - \overline{W}_{k \Delta t} \right]$$

$$+ P^{X^{i,l}}_{k \Delta t} C R^{-1} \left[ \left( Y_{(k+1) \Delta t} - Y_{k \Delta t} \right) \right]$$
Here we use the notation: for a $\text{C.1}$ MSE bound on $\text{EnKBF NC}$ estimator within the $\text{NC}$ estimator.

We will use the fact that the i.i.d. system coincides with the Kalman–Bucy Diffusion $\mathcal{X}_t$, in the limit of $N \to \infty$. This implies the mean and covariance are defined through the Kalman–Bucy filter and the Riccati equations, which allows us to use the results from Appendix B, for the process $\zeta^{\text{C} \Delta_t}$.

### C.1 MSE bound on $\text{EnKBF NC}$ estimator

Here we use the notation: for a $d_x$-dimensional vector $x$ denote $\|x\|_2 = (\sum_{j=1}^{d_x} x_j^2)^{1/2}$.

**Proposition C.1** For any $T \in \mathbb{N}$ fixed and $t \in [0, T - 1]$ there exists a $C < +\infty$ such that for any $(l, N, k_1) \in \mathbb{N}_0 \times \{2, 3, \ldots\} \times \{0, 1, \ldots, \Delta^{-1}_{1} - 1\}$:

$$
\mathbb{E} \left[ \left\| \hat{U}_{t+k_1\Delta_t}^{N,l} - \hat{U}_{t+k_1\Delta_t}^{l} \right\|_2^2 \right] \leq C \left( \frac{1}{N} + \Delta^2 \right).
$$

**Proof**

Using the $C_2$–inequality one has

$$
\mathbb{E} \left[ \left\| \hat{U}_{t+k_1\Delta_t}^{N,l} - \hat{U}_{t+k_1\Delta_t}^{l} \right\|_2^2 \right] 
\leq C \mathbb{E} \left[ \left\| \hat{U}_{t+k_1\Delta_t}^{N,l} - \hat{U}_{t+k_1\Delta_t}^{l} \right\|_2^2 \right] 
+ \mathbb{E} \left[ \left\| \hat{U}_{t+k_1\Delta_t}^{l} - \hat{U}_{t+k_1\Delta_t}^{l} \right\|_2^2 \right].
$$

(C.1)

The first term on the R.H.S. can be controlled by standard results for i.i.d. sampling (recall that $\zeta^{\text{C} \Delta_t} | \mathcal{F}_{t+k_1\Delta_t}$ are i.i.d. Gaussian with mean $m_{t+k_1\Delta_t}$ and covariance $P_{t+k_1\Delta_t}$), that is

$$
\mathbb{E} \left[ \left\| \hat{U}_{t+k_1\Delta_t}^{N,l} - \hat{U}_{t+k_1\Delta_t}^{l} \right\|_2^2 \right] \leq C \frac{1}{N}.
$$

(C.2)

The formula in (C.2) can be proved by using the formulae for the NC estimators, in the usual integral form, and through a simple application of the general Minkowski inequality.

Note that it is crucial that (A.1) holds, otherwise the upper-bound can explode as a function of $l$. For the right-most term on the R.H.S. of (C.1) by Jensen’s inequality and Lemma B.4:

$$
\mathbb{E} \left[ \left\| \hat{U}_{t+k_1\Delta_t}^{l} - \hat{U}_{t+k_1\Delta_t}^{l} \right\|_2^2 \right] \leq C \Delta^2.
$$

(C.3)

So the proof can be concluded by combining (C.1), (C.2) and (C.3).

### C.2 Variance of i.i.d. MLEnKBF NC estimator

**Proposition C.2** For any $(t, q) \in \mathbb{N}_0 \times [1, \infty)$, there exists a $C < +\infty$ such that for any $(l, N, k_1) \in \mathbb{N} \times \{2, 3, \ldots\} \times \{0, 1, \ldots, \Delta^{-1}_{1} - 1\}$:

$$
\mathbb{E} \left[ \left\| \hat{U}_{t+k_1\Delta_t}^{N,l} - \hat{U}_{t+k_1\Delta_t}^{l} \right\|_2^2 \right] 
- \mathbb{E} \left[ \left\| \hat{U}_{t+k_1\Delta_t}^{N,l} - \hat{U}_{t+k_1\Delta_t}^{l} \right\|_2^2 \right] 
\leq \frac{C \Delta_{1}^{1/2}}{\sqrt{N}}. \quad (C.4)
$$

As before, in order to proceed we will make use of a martingale–remainder decomposition. Recall that, for level $s \in \{l - 1, l\}$, we have

$$
\hat{U}_{t+k_1\Delta_t}^{N,s} - \hat{U}_{t+k_1\Delta_t}^{s} = M_{t+k_1\Delta_t}^{l} + R_{t+k_1\Delta_t}^{l}.
$$

Therefore substituting into the LHS of (C.4), and using Minkowski’s inequality, results in

$$
\mathbb{E} \left[ \left\| \hat{U}_{t+k_1\Delta_t}^{N,l} - \hat{U}_{t+k_1\Delta_t}^{l} \right\|_2^2 \right] 
- \mathbb{E} \left[ \left\| \hat{U}_{t+k_1\Delta_t}^{N,l} - \hat{U}_{t+k_1\Delta_t}^{l} \right\|_2^2 \right] 
\leq \frac{C \Delta_{1}^{1/2}}{\sqrt{N}}. \quad (C.5)
$$

**Proof**

$$
\mathbb{E} \left[ \left\| \hat{U}_{t+k_1\Delta_t}^{N,l} - \hat{U}_{t+k_1\Delta_t}^{l} \right\|_2^2 \right] 
- \mathbb{E} \left[ \left\| \hat{U}_{t+k_1\Delta_t}^{N,l} - \hat{U}_{t+k_1\Delta_t}^{l} \right\|_2^2 \right] 
\leq \frac{C \Delta_{1}^{1/2}}{\sqrt{N}}. \quad (C.5)
$$
\[
\sum_{j_1=1}^{d_x} \sum_{j_2=1}^{d_x} \sum_{j_3=1}^{d_{l+x}} \sum_{j_4=1}^{d_{l+x}} C(j_1, j_2) C(j_3, j_4) R^{-1}(j_1, j_3) X_{t_{j_1}}(j_4) \sum_{k=0}^{t_{j_2}+k_{l+x}} \mathcal{C}(j_1, j_2, k_{l+x}) \left( m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) - m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) \right) R^{-1}(j_1, j_3) C X_{t_{j_1}}(j_3) \Delta_{t_{j_1}}^{-1} \left| q \right|^{1/q}.
\]

Then through the generalized Minkowski inequality
\[
\mathbb{E} \left[ \left( |R_t(1) - R_t^{-1}(1)| \right) q^{1/q} \right] \leq C \sum_{j_2=1}^{d_x} \sum_{j_1=1}^{d_x} \sum_{j_3=1}^{d_{l+x}} \sum_{j_4=1}^{d_{l+x}} C(j_1, j_2) C(j_3, j_4) R^{-1}(j_1, j_3) \times \mathbb{E} \left[ \int_0^{t_{j_2}+k_{l+x}} \left( m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) X_{t_{j_1}}(j_4) \right) \left( m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) X_{t_{j_1}}(j_4) \right) ds \right]^{q^{1/q}}
\]
\[
\leq C \sum_{j_2=1}^{d_x} \sum_{j_1=1}^{d_x} \mathbb{E} \left[ \int_0^{t_{j_2}+k_{l+x}} \left( m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) X_{t_{j_1}}(j_4) \right) ds \right]^{q^{1/q}}
\]
\[
\leq C \sum_{j_2=1}^{d_x} \sum_{j_1=1}^{d_x} \int_0^{t_{j_2}+k_{l+x}} \left( m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) X_{t_{j_1}}(j_4) \right) ds \right]^{q^{1/q}}
\]
\[
=: T_1 + T_2,
\]

where we have used \( \tau^s = \lfloor \frac{s}{\Delta_t} \rfloor \Delta_t \) for \( t \in \mathbb{R}^+ \), and
\[
\bar{m}_{k_{l+x}}^{N,s} = m_{k_{l+x}}^{N,s} - m_{k_{l+x}}^{N,s}, \quad s \in \{0, 1, \ldots, t_{j_2} - 1\}. \tag{C.6}
\]

For \( T_1 \) we can express it as
\[
T_1 = C \sum_{j_2=1}^{d_x} \sum_{j_1=1}^{d_x} \mathbb{E} \left[ \int_0^{t_{j_2}+k_{l+x}} \left( m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) - m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) \right) X_{t_{j_1}}(j_4) ds \right]^{q^{1/q}}
\]
\[
\leq C \sum_{j_2=1}^{d_x} \sum_{j_1=1}^{d_x} \int_0^{t_{j_2}+k_{l+x}} \left( m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) - m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) \right) X_{t_{j_1}}(j_4) ds \right]^{q^{1/q}}
\]
\[
=: T_1 + T_2.
\]

For \( T_2 \) we can apply the Marcinkiewicz–Zygmund and Hölder inequalities, and using the fact that means can be expressed as the expectations of (3.9)–(3.10)
\[
T_3 = C \sum_{j_2=1}^{d_x} \sum_{j_1=1}^{d_x} \int_0^{t_{j_2}+k_{l+x}} \mathbb{E} \left[ \left( m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) - m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) \right) X_{t_{j_1}}(j_4) \right]^{q^{1/q}} ds
\]
\[
\leq C \sum_{j_2=1}^{d_x} \sum_{j_1=1}^{d_x} \int_0^{t_{j_2}+k_{l+x}} \mathbb{E} \left[ \left( m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) - m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) \right) X_{t_{j_1}}(j_4) \right]^{q^{1/q}} ds
\]
\[
\leq C \sum_{j_2=1}^{d_x} \sum_{j_1=1}^{d_x} \int_0^{t_{j_2}+k_{l+x}} \mathbb{E} \left[ \left( m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) - m_{t_{j_2}+k_{l+x}}^{N,l-1}(j_2) \right) X_{t_{j_1}}(j_4) \right]^{q^{1/q}} ds.
\]

The process \( \mathbb{E}[|X_{t_{j_1}}|^{2q}] \) is of order \( \mathcal{O}(1) \) and the recursion is of order \( \mathcal{O}(\Delta_t) \), through the strong error Lemma B.2. For \( T_4 \) we know it is sufficiently small, which is of order \( \mathcal{O}(\Delta_t) \). For \( T_2 \), we use the definition of the discretized diffusion process
\[
X_{t_{j_1}} - X_{t_{j_1}-1} = \int_{t_{j_1}-1}^{t_{j_1}} A_X du + Q^{1/2} \left( W_{t_{j_1}} - W_{t_{j_1}-1} \right). \tag{C.7}
\]

As before, we know the difference of the Brownian motion increment \( \mathbb{E}\left[ |W_{t_{j_1}} - W_{t_{j_1}-1}|^{q/2} \right] \) is of order \( \mathcal{O}(\Delta_t^{q/2}) \), and, as before, \( \mathbb{E}[|X_{t_{j_1}}|^{q}] \leq C. \) Therefore the integral term of (C.7) is of order \( \mathcal{O}(\Delta_t^{q/2}) \). Finally, for \( \mathbb{E}[|m_{t_{j_2}+k_{l+x}}^{N,l-1}|^{q}] \), as it is of order \( \mathcal{O}(N^{-1/2}) \), therefore, combining all terms, we can deduce from that
\[
\mathbb{E}\left[ \left( |R_{t_{j_2}+k_{l+x}}^{N,l-1}(1) - R_{t_{j_2}+k_{l+x}}^{N,l-1}(1)| \right) \right]^{q^{1/q}} \leq \frac{C \Delta_t^{q/2}}{\sqrt{N}}.
\]

**Lemma C.2** For any \((t, q) \in \mathbb{N}_0 \times [1, \infty), \) there exists a \( C < +\infty \) such that for any \((l, N, k_1) \in \mathbb{N} \times \{2, 3, \ldots\} \times \{0, 1, \ldots, \Delta_t^{-1} - 1\):
\[ \mathbb{E}\left[ \left( M_{t+k_1\Delta_{t-1}} - M_{t+k_1\Delta_{t-1} - 1} \right) \right]^{q/2} \leq \frac{C\Delta_1^{1/2}}{N}. \]  \hspace{1cm} (C.8)

**Proof** As before, we set \( \tau_i^s = \frac{1}{\Delta_s} \Delta_s \) for \( t \in \mathbb{R}^+ \), and make use of (C.6)

\[ \mathbb{E}\left[ \left( M_{t+k_1\Delta_{t-1}} - M_{t+k_1\Delta_{t-1} - 1} \right) \right]^{q/2} \]
\[ = \mathbb{E}\left[ \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} C(j_1, j_2) \left[M_{t+k_1\Delta_{t-1}}(j_2) - M_{t+k_1\Delta_{t-1} - 1}(j_2) \right] R^{-1}(j_1, j_3) \]
\[ \times \left( Y_{(t+k_1)\Delta_{t-1} - 1} - Y_{(t+k_1)\Delta_{t-1}} \right) \left(j_3 = CX_{(t+k_1)\Delta_{t-1}}(j_3) \Delta_t \right) \]
\[ \sum_{k=0}^{t+k_1\Delta_{t-1} - 1} \left( \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} C(j_1, j_2) \left[M_{t+k_1\Delta_{t-1}}(j_2) - M_{t+k_1\Delta_{t-1} - 1}(j_2) \right] R^{-1}(j_1, j_3) \]
\[ \left( Y_{(t+k_1)\Delta_{t-1} - 1} - Y_{(t+k_1)\Delta_{t-1}} \right) \left(j_3 = CX_{(t+k_1)\Delta_{t-1}}(j_3) \Delta_t \right) \left( Y_{(t+k_1)\Delta_{t-1} - 1} - Y_{(t+k_1)\Delta_{t-1}} \right) \left(j_3 = CX_{(t+k_1)\Delta_{t-1}}(j_3) \Delta_t \right) \right) \]  \hspace{1cm} (C.9)

Then by using generalized Minkowski and Jensen’s inequality

\[ \mathbb{E}\left[ \left( M_{t+k_1\Delta_{t-1}} - M_{t+k_1\Delta_{t-1} - 1} \right) \right]^{q/2} \]
\[ = \mathbb{E}\left[ \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} C(j_1, j_2) R^{-1}(j_1, j_3) \int_0^{t+k_1\Delta_{t-1}} \left( M_{t+k_1\Delta_{t-1}}(j_2) - M_{t+k_1\Delta_{t-1} - 1}(j_2) \right) ds \right]^{q/2} \]
\[ \leq C \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \int_0^{t+k_1\Delta_{t-1}} \left( M_{t+k_1\Delta_{t-1}}(j_2) - M_{t+k_1\Delta_{t-1} - 1}(j_2) \right) ds \right]^{q/2} \]
\[ \leq \frac{d_x}{c} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \int_0^{t+k_1\Delta_{t-1}} \left( M_{t+k_1\Delta_{t-1}}(j_2) - M_{t+k_1\Delta_{t-1} - 1}(j_2) \right) ds \right]^{q/2} \]
\[ =: T_2 + T_3. \]

For \( T_2 \), we know it follows the same analysis as Lemma C.1. Therefore we can conclude that

\[ T_2 \leq \frac{C\Delta_1^{1/2}}{N}. \]

For \( T_3 \) by using the Burkholder–Davis–Gundy and Hölder inequality, as done previously from (A.5), and we use the bound from Lemma C.1, for the \( \left( M_{t+k_1\Delta_{t-1}} - M_{t+k_1\Delta_{t-1} - 1} \right) \) term. Therefore this implies that

\[ \mathbb{E}\left[ \left( \left( M_{t+k_1\Delta_{t-1}} - M_{t+k_1\Delta_{t-1} - 1} \right) \right) \right]^{q/2} \]
\[ \leq \frac{C\Delta_1^{1/2}}{N}. \]

**Lemma C.3** For any \( (t, q) \in \mathbb{N}_0 \times [1, \infty) \), there exists a \( C < +\infty \) such that for any \( (l, N, k_1) \in \mathbb{N} \times \{2, 3, \ldots\} \times \{0, 1, \ldots, \Delta_{t-1} - 1\} \):

\[ \mathbb{E}\left[ \left( \left( R_{t+k_1\Delta_{t-1}} - R_{t+k_1\Delta_{t-1} - 1} \right) \right) \right]^{q/2} \]
\[ \leq \frac{C\Delta_1^{1/2}}{N}. \]  \hspace{1cm} (C.9)

**Proof** Again we let \( \tau_i^s = \frac{1}{\Delta_s} \Delta_s \) for \( t \in \mathbb{R}^+ \), and using the generalized Minkowski inequality

\[ \mathbb{E}\left[ \left( \left( R_{t+k_1\Delta_{t-1}} - R_{t+k_1\Delta_{t-1} - 1} \right) \right) \right]^{q/2} \]
\[ = \mathbb{E}\left[ \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} S(j_1, j_2) \int_0^{t+k_1\Delta_{t-1}} \left( m_{t+k_1\Delta_{t-1}}(j_2) - m_{t+k_1\Delta_{t-1} - 1}(j_2) \right) ds \right]^{q/2} \]
\[ \leq \frac{d_x}{c} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \int_0^{t+k_1\Delta_{t-1}} \left( m_{t+k_1\Delta_{t-1}}(j_2) - m_{t+k_1\Delta_{t-1} - 1}(j_2) \right) ds \right]^{q/2} \]
\[ = \frac{d_x}{c} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \int_0^{t+k_1\Delta_{t-1}} \left( m_{t+k_1\Delta_{t-1}}(j_2) - m_{t+k_1\Delta_{t-1} - 1}(j_2) \right) ds \right]^{q/2} \]
\[ \leq \frac{d_x}{c} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \int_0^{t+k_1\Delta_{t-1}} \left( m_{t+k_1\Delta_{t-1}}(j_2) - m_{t+k_1\Delta_{t-1} - 1}(j_2) \right) ds \right]^{q/2} \]
\[ \leq \frac{d_x}{c} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \sum_{j=1}^{d_x} \int_0^{t+k_1\Delta_{t-1}} \left( m_{t+k_1\Delta_{t-1}}(j_2) - m_{t+k_1\Delta_{t-1} - 1}(j_2) \right) ds \right]^{q/2} \]  \hspace{1cm} (C.9)
\[ T_1 = C \sum_{j_1} \sum_{j_2} \left( \sum_{l_1} \sum_{l_2} \left( \mathbb{E} \left[ \left( m_{N,t_i}^{N,l}(j_1)m_{N,t_i}^{N,l}(j_2) \right) \right] \right) \right) \]

\[ + C \sum_{j_1} \sum_{j_2} \left( \sum_{l_1} \sum_{l_2} \left( \mathbb{E} \left[ \left( m_{N,t_i}^{N,l}(j_1)m_{N,t_i}^{N,l}(j_2) \right) \right] \right) \right) \]

\[ =: T_1 + T_2. \]

For \( T_1 \) we make use of (C.6) and with the generalized Minkowski inequality

\[ T_1 = C \sum_{j_1} \sum_{j_2} \left( \sum_{l_1} \sum_{l_2} \left( \mathbb{E} \left[ \left( m_{N,t_i}^{N,l}(j_1)m_{N,t_i}^{N,l}(j_2) \right) \right] \right) \right) \]

\[ =: T_1 + T_2. \]

Similarly, \( T_4 \) can be expressed as

\[ T_4 = C \sum_{j_1} \sum_{j_2} \left( \sum_{l_1} \sum_{l_2} \left( \mathbb{E} \left[ \left( m_{N,t_i}^{N,l}(j_1)m_{N,t_i}^{N,l}(j_2) \right) \right] \right) \right) \]

\[ =: T_1 + T_2. \]

For \( T_2 \), we can rewrite it as

\[ T_2 = C \sum_{j_1} \sum_{j_2} \left( \sum_{l_1} \sum_{l_2} \left( \mathbb{E} \left[ \left( m_{N,t_i}^{N,l}(j_1)m_{N,t_i}^{N,l}(j_2) \right) \right] \right) \right) \]

\[ =: T_1 + T_4. \]

For \( T_3 \) we can use the difference of mean trick, as in (B.6),

\[ T_3 = C \sum_{j_1} \sum_{j_2} \left( \sum_{l_1} \sum_{l_2} \left( \mathbb{E} \left[ \left( m_{N,t_i}^{N,l}(j_1)m_{N,t_i}^{N,l}(j_2) \right) \right] \right) \right) \]

\[ =: T_3 + T_4. \]

We know \( \mathbb{E}[m_{N,t_i}^{N,l}] \leq C \), and \( m_{N,t_i}^{N,l}(j_1) \) is of order \( O(N^{-\frac{1}{2}}) \). The last bracket term is of order \( O(\Delta t) \), arising from the strong error. The first bracket term is the same, that appears in Lemma C.1, which is of order \( O(\frac{\Delta t}{N^2}) \).
We express this as
\[
T_2 = C \sum_{j_1=1}^{d_s} \sum_{j_2=1}^{d_s} \left( \mathbb{E} \left[ \left| \int_0^{t+k_1 \Delta t-1} T_5 + T_6 + T_7 \, ds \right|^q \right] \right)^{1/q}.
\]
where

\[
T_5 = \left( \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_1) \right) \left( \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_2) \right) - \left( \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_1) \right) \left( \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_2) \right).
\]

\[
T_6 = \left( \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_1) \right) \mathbb{E} \left[ \xi_{t,i-1}^{j} (j_2) \right] - \left( \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_1) \right) \mathbb{E} \left[ \xi_{t,i-1}^{j} (j_2) \right].
\]

\[
T_7 = \left( \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_2) \right) \mathbb{E} \left[ \xi_{t,i-1}^{j} (j_1) \right] - \left( \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_2) \right) \mathbb{E} \left[ \xi_{t,i-1}^{j} (j_1) \right].
\]

For $T_6$ and $T_7$, they can be expressed as

\[
T_6 + T_7 = \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_1) \left( \mathbb{E} \left[ \xi_{t,i-1}^{j} (j_2) \right] - \mathbb{E} \left[ \bar{\xi}_{t,i-1}^{j} (j_2) \right] \right)
\]

\[
+ \mathbb{E} \left[ \xi_{t,i-1}^{j} (j_2) \right] \frac{1}{N} \sum_{i=1}^{N} \left( \bar{\xi}_{t,i-1}^{j} (j_1) - \bar{\xi}_{t,i-1}^{j} (j_1) \right)
\]

\[
+ \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_2) \left( \mathbb{E} \left[ \xi_{t,i-1}^{j} (j_1) \right] - \mathbb{E} \left[ \bar{\xi}_{t,i-1}^{j} (j_1) \right] \right)
\]

\[
+ \mathbb{E} \left[ \xi_{t,i-1}^{j} (j_1) \right] \frac{1}{N} \sum_{i=1}^{N} \left( \bar{\xi}_{t,i-1}^{j} (j_2) - \bar{\xi}_{t,i-1}^{j} (j_2) \right).
\]

For $T_5$, we can rewrite it as

\[
T_5 = \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_2) \left( \frac{1}{N} \sum_{i=1}^{N} \left( \bar{\xi}_{t,i-1}^{j} (j_1) - \bar{\xi}_{t,i-1}^{j} (j_1) \right) \right)
\]

\[
+ \left( \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_1) \right) \frac{1}{N} \sum_{i=1}^{N} \left( \bar{\xi}_{t,i-1}^{j} (j_2) - \bar{\xi}_{t,i-1}^{j} (j_2) \right).
\]

Using these expressions, and through the generalized Minkowski and Jensen’s inequality, $T_2$ can be simplified as

\[
T_2 \leq C \sum_{j_1=1}^{d_s} \sum_{j_2=1}^{d_s} \left( \mathbb{E} \left[ \left| \int_0^{t+k_1 \Delta t-1} \left( \frac{1}{N} \sum_{i=1}^{N} \bar{\xi}_{t,i-1}^{j} (j_2) \right) \right|^q \right] \right)^{1/q}.
\]

Then by using the results from the strong and weak error of the diffusion processes, in Lemma B.2 and B.3, we reach the following bound of

\[
\mathbb{E} \left[ \left| \left( R_{t+k_1 \Delta t-1}^j - R_{t}^j \right) \right|^q \right]^{1/q} \leq \frac{C \Delta t}{\sqrt{N}}.
\]

Therefore by combining all the results from Lemma C.1 - Lemma C.3, leads to the desired result of (C.4) in Proposition C.2.

### C.3 Proof of Theorem 3.1

**Proof** Noting (2.15) one has

\[
[\hat{U}_{t}^{\text{ML}} - \hat{U}_{t}](Y) = \left[ \hat{U}_{t}^{N_0} - \hat{U}_{t}^{0} \right](Y)
\]
\begin{align*}
+ \sum_{l=1}^{L} & \left( \widehat{U}_{i}^{N,l} - \widehat{U}_{i}^{N,l-1} - \widehat{U}_{i}^{l} + \widehat{U}_{i}^{l-1} \right)(Y) + \{ \widehat{U}_{i}^{l} - \widehat{U}_{i} \}(Y).
\end{align*}

Thus, by using three applications of the $C_2$-inequality we have
\begin{align*}
& \mathbb{E} \left[ \left\| \widehat{U}_{i}^{M,L} - \widehat{U}_{i} \right\|_2^2 \right] \\
\leq & \mathcal{C} \left( \mathbb{E} \left[ \left\| \widehat{U}_{i}^{N,0} - \widehat{U}_{i}^{0} \right\|_2^2 \right] \\
+ & \mathbb{E} \left[ \left\| \sum_{l=1}^{L} \left( \widehat{U}_{i}^{N,l} - \widehat{U}_{i}^{N,l-1} - \widehat{U}_{i}^{l} + \widehat{U}_{i}^{l-1} \right)(Y) \right\|_2^2 \right] \\
+ & \mathbb{E} \left[ \left\| \{ \widehat{U}_{i}^{l} - \widehat{U}_{i} \}(Y) \right\|_2^2 \right].
\end{align*}

For the last term on the R.H.S. one can use (C.2) and for the middle term on the R.H.S., we can use (C.3). For the middle term, one has
\begin{align*}
\left\| \sum_{l=1}^{L} \left( & \widehat{U}_{i}^{N,l} - \widehat{U}_{i}^{N,l-1} - \widehat{U}_{i}^{l} + \widehat{U}_{i}^{l-1} \right)(Y) \right\|_2^2 \\
= & \sum_{l=1}^{L} \sum_{j=1}^{d_{\nu}} \left( \left( \widehat{U}_{i}^{N,l} - \widehat{U}_{i}^{N,l-1} - \widehat{U}_{i}^{l} + \widehat{U}_{i}^{l-1} \right)(Y) \right)^2(j) \\
+ & \sum_{l=1}^{L} \sum_{q=1}^{d_{\nu}} \sum_{l=1}^{L} \sum_{j=1}^{d_{\nu}} \left( \left( \widehat{U}_{i}^{N,l} - \widehat{U}_{i}^{N,l-1} - \widehat{U}_{i}^{l} + \widehat{U}_{i}^{l-1} \right)(Y) \right)(j) \\
\times & \left( \left( \widehat{U}_{i}^{N,q} - \widehat{U}_{i}^{N,q-1} - \widehat{U}_{i}^{q} + \widehat{U}_{i}^{q-1} \right)(Y) \right)(j).
\end{align*}

Then using a combination of the independence of the coupled particle systems along with Proposition C.2 the proof can be concluded. \hfill $\square$

### References

Arapostathis, A., Marcus, S.I.: Analysis of an identification algorithm arising in the adaptive estimation of Markov chains. Math. Control Signals Syst. 3, 1–29 (1990)

Bain, A., Crisan, D.: Fundamentals of Stochastic Filtering. Springer, New York (2009)

Beskos, A., Crisan, D., Jasra, A., Kantas, N., Ruzayqat, H.: Score-based parameter estimation for a class of continuous-time state space models. SIAM J. Math. Comput. (to appear) (2021)

Bhar, R.: Stochastic Filtering with Applications in Finance. World Scientific (2010)

Bishop, A.N., Del Moral, P.: On the mathematical theory of ensemble (linear-gaussian) Kalman–Bucy filtering. arXiv preprint arXiv:2006.08843 (2020)

Bishop, A.N., Del Moral, P.: On the stability of Kalman–Bucy diffusion processes. SIAM J. Control Optim. 55(6), 4015–4047 (2017)

Brosse, N., Durmus, A., Moulines, E.: Normalizing constants of log-concave densities. Elect. J. Stat. 12, 851–889 (2018)

Cerou, F., Del Moral, P., Guyader, A.: A non-asymptotic variance theorem for unnormalized Feynman–Kac particle models. Ann. Inst. Henri Poincare 4, 629–649 (2011)

Chada, N.K., Franks, J., Jasra, A., Law, K.J.H., Vihola, M.: Unbiased inference for discretely observed hidden Markov model diffusions. SIAM/ASA J. Uncertainty Quantif. 9(2), 763–787 (2021)

Chada, N.K., Jasra, A., Yu, F.: Multilevel ensemble Kalman–Bucy filters. SIAM/ASA J. Uncertainty Quantif. (to appear) (2022)

Chernov, A., Hoel, H., Law, K.J.H., Nobile, F., Tempone, R.: Multilevel ensemble Kalman filtering for spatio-temporal processes. Numer. Math. 1:1–55, (2021)

Crisan, D., Del Moral, P., Jasra, A., Ruzayqat, H.: Log-normalization constant estimation using the ensemble Kalman–Bucy filter with application to high-dimensional models. Adv. Appl. Probab. (to appear) (2022)

Crisan, D., Rozovskii, B.: The Oxford Handbook of Nonlinear Filtering. Oxford University Press, Oxford (2011)

Del Moral, P., Jasra, A., Law, K., Zhou, Y.: Multilevel sequential Monte Carlo samplers for normalizing constants. TOMACS 27, 1 (2017)

Del Moral, P., Doucet, A., Singh, S.S.: Forward smoothing using sequential Monte Carlo. arXiv preprint arxiv:1012.5390 (2010)

Del Moral, P.: Feynman-Kac Formulse. Springer, New York (2004)

Del Moral, P., Tugaut, J.: On the stability and the uniform propagation of chaos properties of ensemble Kalman–Bucy filters. Ann. Appl. Probab. 28, 790–850 (2018)

Evensen, G.: Data Assimilation: The Ensemble Kalman Filter, 2nd edn. Springer (2009)

Fossum, K., Mannseth, T., Stordal, A.S.: Assessment of multilevel ensemble-based data assimilation for reservoir history matching. Comput. Geosci. 24, 217–239 (2020)

Giles, M.B.: Multilevel Monte Carlo path simulation. Op. Res. 56, 607–617 (2008)

Giles, M.B.: Multilevel Monte Carlo methods. Acta Numer. 24, 259–328 (2015)

Gelman, A., Meng, X.-L.: Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. Stat. Sci. 1, 163–185 (1998)

Hoel, H., Law, K.J.H., Tempone, R.: Multilevel ensemble Kalman filtering. SIAM J. Numer. Anal. 54(3), 1813–1839 (2016)

Hoel, H., Shaimerdenova, G., Tempone, R.: Multi-index ensemble Kalman filtering arXiv preprint arXiv:2104.07263 (2021)

Jasra, A., Kamatani, K., Law, K.J.H., Zhou, Y.: Multilevel particle filters. SIAM J. Numer. Anal. 55(6), 3068–3096 (2017)

Jasra, A., Kamatani, K., Osei, P.P., Zhou, Y.: Multilevel particle filters: normalizing constant estimation. Stat. Comput. 28, 47–60 (2018)

Jazwinski, A.: Stochastic processes and filtering theory, vol. x 63. Academic Press, New York (1970)

Kostov, S., Whiteley, N.: An algorithm for approximating the second moment of the normalizing constant estimate from a particle filter. Methodol. Comput. Appl. Probab. 19, 799–818 (2017)

Le Gland, F., Mevel, M.: Recursive identification in hidden Markov models. Proc. 36th IEEE Conf. Dec. Contr. pp. 3468–3473 (1997)

Lorenz, E.N.: Deterministic nonperiodic flow. J. Atmos. Sci. 20(2), 130–141 (1963)

Lorenz, E.N.: Predictability: A problem partly solved. Proc. ECMWF Seminar on predictability I, 1–18 (1996)

Majda, A., Wang, X.: Non-linear Dynamics and Statistical Theories for Basic Geophysical Flows. Cambridge University Press (2006)

Oliver, D., Reynolds, A.C., Liu, N.: Inverse Theory for Petroleum Reservoir Characterization and History Matching, 1st edn. Cambridge University Press (2008)

Reich, S.: Data assimilation: the Schrödinger perspective. Acta Numer. 28, 635–711 (2019)

Poyiadjis, G., Doucet, A., Singh, S.S.: Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. Biometrika 98, 65–80 (2011)
Sakov, P., Oke, P.R.: A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. Tellus A. \textbf{60}(2), 361–371 (2008)

Spall, J.C.: Introduction to Stochastic Search and Optimization, 1st edn. Wiley, New York (2003)

Spall, J.C.: Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE Trans. Auto. Control. \textbf{37}, 332–341 (1992)

Rischard, M., Jacob, P.E., Pillai, N.: Unbiased estimation of log normalizing constants with applications to Bayesian cross-validation. arXiv preprint \texttt{arxiv:1810.01382} (2018)

Vihola, M.: Unbiased estimators and multilevel Monte Carlo. Op. Res. \textbf{66}, 448–462 (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.