Information barriers for noisy Lagrangian tracers in filtering random incompressible flows

This content has been downloaded from IOPscience. Please scroll down to see the full text.
2014 Nonlinearity 27 2133
(http://iopscience.iop.org/0951-7715/27/9/2133)
View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 216.165.95.76
This content was downloaded on 12/08/2014 at 17:27

Please note that terms and conditions apply.
Information barriers for noisy Lagrangian tracers in filtering random incompressible flows

Nan Chen, Andrew J Majda and Xin T Tong

Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA

E-mail: tong@cims.nyu.edu

Received 12 February 2014, revised 26 June 2014
Accepted for publication 2 July 2014
Published 1 August 2014

Recommended by K Ohkitani

Abstract
An important practical problem is the recovery of a turbulent velocity field from Lagrangian tracers that move with the fluid flow. Here, the filtering skill of moving Lagrangian tracers in recovering random incompressible flow fields defined through a finite number of random Fourier modes is studied with full mathematical rigour. Despite the inherent nonlinearity in measuring noisy Lagrangian tracers, it is shown below that there are exact closed analytic formulas for the optimal filter for the velocity field involving Riccati equations with random coefficients for the covariance matrix. This mathematical structure allows a detailed asymptotic analysis of filter performance, both as time goes to infinity and as the number of noisy Lagrangian tracers, \( L \), increases. In particular, the asymptotic gain of information from \( L \)-tracers grows only like \( \ln L \) in a precise fashion; i.e., an exponential increase in the number of tracers is needed to reduce the uncertainty by a fixed amount; in other words, there is a practical information barrier. The proofs proceed through a rigorous mean field approximation of the random Ricatti equation. Also, as an intermediate step, geometric ergodicity with respect to the uniform measure on the period domain is proved for any fixed number \( L \) of noisy Lagrangian tracers. All of the above claims are confirmed by detailed numerical experiments presented here.

Keywords: Lagrangian tracers, filtering, information barrier, optimal filter

Mathematics Subject Classification: 60G35, 60H30, 62M20, 70H03, 93E11, 94A15

(Some figures may appear in colour only in the online journal)
1. Introduction

Lagrangian tracers are drifters or floaters that follow a parcel of fluid’s movement. Serving as moving observations, Lagrangian tracers gather real-time information as they travel, which is of particular significance for data collection in the centre of the ocean since Eulerian measurements are only available near the shores [1,2]. One key application of these Lagrangian data, which is also the main concern of this paper, is the inference of the current state of the velocity field.

Data assimilation with Lagrangian tracers has attracted considerable attention over the years. Various filtering approaches have been applied to Lagrangian data in different modelling contexts. This includes assimilation methods based on Taylor expansion [3], particle filter methods [4] and the ensemble Kalman filter (EnKF) [5–9]. These approximate filters produce satisfactory numerical results in the recovery of the velocity field. Related numerical experiments also shed light on the dependence of filter performance over different factors, such as tracers’ trajectories, the number of tracers and initial deployment of the tracers [6,7].

In spite of the achievements mentioned above, little is known about filters for Lagrangian tracers in theory, as all the analyses were carried out by numerical experiments. The nonlinear dynamics of Lagrangian tracers make it difficult to establish a theoretical framework, which in turn impedes systematic understanding. One goal of the present work is to introduce random incompressible flow models which, despite the nonlinearity in measuring noisy Lagrangian tracers, have mathematical structure with exact closed analytic formulas for an optimal filter for the velocity field.

The second motivation of this paper lies in understanding the uncertainty reduction with Lagrangian tracers. Without any observations, the equilibrium distribution of the underlying dynamics provides the least biased estimation of the current flow. Unfortunately, the strong turbulent nature brings about a large uncertainty in the equilibrium distribution among many practical issues. Intuitively, the observations of the tracers’ trajectories help to reduce the uncertainty. This uncertainty reduction can be quantified using information theory, as is well documented in the literature [10–13].

In principle, adding more tracers is expected to help in gaining more information about the underlying flow. However, adding more tracers may not be an effective way to reduce the uncertainty. On the one hand, this will inevitably increase the maintenance and computational cost of the filter. On the other hand, as recently revealed in various scenarios for prediction with model error [10–12], there might exist an information barrier for our filter. An information barrier indicates that a certain level of information is extremely difficult to acquire in practice; therefore, adding tracers produces little effective information return. Evidently, identifying and understanding a potential information barrier is of importance for the optimal arrangement of the filter.

In the following, we consider a simplified set-up with a $d$-dimensional incompressible random flow modelled by a finite number of Fourier modes with random amplitudes:

$$
\vec{v}(\vec{x}, s) = \sum_{k \in K} \hat{v}_k(s) \exp(i\vec{k} \cdot \vec{x}) \vec{r}_k. \tag{1.1}
$$

Since the whole velocity field $\vec{v}$ is real-valued, we require that the modes $\hat{v}_k(s)$ with nonzero wavenumbers form conjugate pairs. The $\vec{r}_k$ are vectors orthogonal to the wavenumber $\vec{k}$, due to the incompressibility of the whole velocity, i.e. $\nabla \cdot \vec{v} = 0$. Each $\hat{v}_k(s)$ follows an Ornstein–Uhlenbeck (OU) process,

$$
d\hat{v}_k(s) = -d_k \hat{v}_k(s) ds + f_k(s) ds + \sigma_k dW^k_s(s). \tag{1.2}
$$
where \( f_k(s) \) is a prescribed external forcing. The joint dynamics of all these Fourier coefficients serve as the signal process in our filtering problem. This model can be regarded as a stochastic version of many canonical deterministic fluid models, see [14] and [15] chapter 5.

On the other hand, the observation process is given by the trajectories of \( L \) Lagrangian tracers involving some noise, i.e. noisy Lagrangian tracers. This observation process is modelled by the following stochastic differential equations (SDEs):

\[
d\vec{X}_l(s) = \vec{v}(\vec{X}_l(s), s)ds + \sigma_x dW_x^l(s) = \sum_{k \in K} \hat{v}_k(s) \exp(i\vec{k} \cdot \vec{X}_l(s))\vec{r}_k ds + \sigma_x dW_x^l(s), \quad l = 1, \ldots, L. \quad (1.3)
\]

Here is the key mathematical observation forming the basis for the analysis presented in the remainder of the paper. Although the evolution of \( \vec{X}_l(s) \) is nonlinear, it depends linearly on the signal processes \( \hat{v}_k(s) \). Therefore, our signal-observation system (1.1) and (1.3) fits into the category of ‘conditional Gaussian process’, introduced by [16]. Such signal-observation processes are linear once the observations are given. With this conditional property, the conditional distribution of the signal with given observations, i.e. the posterior distribution, is Gaussian, as long as the signal process is initially distributed as Gaussian. Moreover, the posterior mean and covariance of this distribution follow explicit nonlinear differential equations with random coefficients, which can be seen as a generalization of those of a Kalman–Bucy filter.

Two significant advantages over other filters for Lagrangian data are possessed by our filter produced by this conditional Gaussian theory. First, it is an optimal or exact filter, as there are no approximation errors appearing as with the EnKF for non-Gaussian models. Second, its evolution involves rather elementary formulas, which enables detailed theoretical analysis and simple numerical algorithms. These advantages will be exploited to study the filtering skill and explore the information barriers in the remainder of the paper.

In order to give an asymptotic characterization of the filter for \( L \to \infty \), the assumption that the tracers are uniformly deployed initially is assumed. This is the most interesting setting in practice, because intuitively it generates a more complete observation, which is also numerically verified in [7] under a different setting. Moreover, theorem 3.1 proved below establishes ergodicity of the uniform distribution for a finite number of tracers in the periodic setting; in other words, the uniform distribution is the attracting equilibrium distribution for the tracers’ projection on the periodic domain. From this, we can establish formal mean field dynamics of the random filter and use these to perform asymptotic analysis.

Here, we mention qualitatively the main results of this paper. Theorems have been proven to support these points.

- The posterior covariance matrix approaches a deterministic matrix \( R_L \), with \( R_L \) being a diagonal matrix which scales as \( L^{-1/2} \), see theorem 3.3, part (i);
- The posterior mean, i.e. the maximum likelihood estimator produced by the filter, converges to the true value of the signal, see theorem 3.3, part (ii) and (iii);
- The total uncertainty reduction gained by the observations, being measured either in relative entropy or mutual information, asymptotically increases as \( \frac{1}{2} |K| \ln L \), where \(|K|\) is the number of modes included; in other words, we gain only \( \frac{1}{4} \) nat of information for each mode with each additional order of magnitude of \( L \), see corollary 3.4.

Judging from the last item, the marginal information gained from each additional tracer increases very slowly with the number of tracers, \( L \). This will form an information barrier in practice, when only a reasonable amount of tracers are available.

The rest of this paper is organized as follows. In section 2, we set up the random incompressible model, establish the filter using conditional Gaussian theory and set up the
uncertainty quantification with explicit formulas. Section 3 states our theoretical results. The major lines of proof are in section 4 with additional verification done in the appendix. The corresponding numerical experiments, showing the uncertainty reduction and information barrier, are included in section 5. We conclude in section 6.

2. Set-up and terminology

2.1. Random incompressible flow model

To set up the intuition behind (1.1), consider a two-dimensional version of this incompressible random flow model with \( \vec{k} = (k_1, k_2)^T \),

\[
\vec{v}(\vec{x}, s) = \hat{v}_{0,1}(s) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \hat{v}_{0,2}(s) \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \sum_{|k_i| \neq |k_j| \in \mathbb{N}, k \neq \vec{0}} \hat{v}_k(s) \exp(i\vec{k} \cdot \vec{x}) \frac{i\vec{k}^\perp}{|k|},
\]

(2.1)

where \( \vec{k}^\perp = (-k_2, k_1)^T \) and \( |k_1| \lor |k_2| \) denote the maximum value of \( |k_1| \) and \( |k_2| \). The first two real modes represent a random background mean sweep and the remaining complex-valued modes form conjugate pairs and represent a superposition of random plane waves. (2.1) can be easily generalized to any higher dimension, and has a general simplified form as (1.1):

\[
\sum_{k \in K} \hat{v}_k(s) \exp(i\vec{k} \cdot \vec{x})\vec{r}_k,
\]

as long as the following properties of (2.1) are preserved:

(i) \( \vec{r}_k \) is orthogonal to the wavenumber \( \vec{k} \); when there are multiple \( k_s \) corresponding to the same wavenumber \( \vec{k} \), we assume \( \vec{r}_k \) are orthogonal to each other without loss of generality;

(ii) The modes with wavenumber \( \vec{k} = \vec{0} \) are real-valued;

(iii) The complex-valued modes come into conjugate pairs, as \( \hat{v}_k = \hat{v}_k^*, \vec{r}_k = \vec{r}_k^* \), where \( -\vec{k} \) stands for the conjugate index of \( \vec{k} \).

For the sake of conciseness, we leave the lengthy but simple model description in appendix A.1.

In the following, we group all the signal processes, \( \{\hat{v}_k(s)\}_{k \in K} \), into a \( |K| \)-dimensional vector process

\[
U_s = \begin{bmatrix} \vdots \\
\hat{v}_k(s) \\
\vdots 
\end{bmatrix}.
\]

There exists an obvious bijection between \( \vec{v}(\cdot, s) \) and \( U_s \): the latter can be obtained by taking the Fourier transform of the former, while the former can be written as

\[
\vec{v}(\vec{x}, s) = \sum_{k \in K} \hat{v}_k(s) \exp(i\vec{k} \cdot \vec{x})\vec{r}_k = P_X(\vec{x}) \cdot U_s,
\]

(2.2)

with \( P_X(\vec{x}) = [\cdots, \exp(i\vec{k} \cdot \vec{x})\vec{r}_k, \cdots] \) being a \( d \times |K| \)-dimensional matrix.

Since each of its components follows (1.2), \( U_s \) is the solution to the following linear SDE:

\[
dU_s = -\Gamma U_s ds + F_s ds + \Sigma_u dB_s,
\]

(2.3)

where \( \Gamma \) is a strictly positive diagonal matrix, \( F_s \) represents the deterministic forcing and \( B_s \) is a canonical \( |K| \)-dimensional real-valued Wiener process. In order to preserve the conjugate pairs, \( \Sigma_u \) will have \( \sigma_u \) as the \( (k, k) \)th term for the real-valued modes, and \( (k, -k) \) \( 2 \times 2 \) sub-block being \( \frac{1}{\sqrt{2}}\sigma_u \begin{bmatrix} 1 & \ldots \\
\ldots & 1 
\end{bmatrix} \) for the complex conjugate pairs; the other components are 0. Note that \( \Sigma = \Sigma_u \Sigma_u^* \) is a real diagonal matrix with the \( (k, k) \)th entry being \( \sigma_k^2 \). A further
Nonlinearity 27 (2014) 2133 N Chen et al

explanation of \( \Sigma_u \)’s form can be found in appendix A.1. Here, we assume \( F_s \) is periodic with period \( T \).

Without any observations, the least biased status of the velocity field, or equivalently \( U_s \), is given by its time-dependent equilibrium distribution, also known as the climatology in climate science. This statistical attractor of \( U_s \) can be computed by applying the Fokker–Planck equation to (2.3). Since (2.3) is linear, the equilibrium distribution is Gaussian

\[
\pi_{\text{att}} = \mathcal{N}(\vec{m}_{\text{att}}, \vec{R}_{\text{att}}),
\]

where \( \vec{m}_{\text{att}} \) has the same period as \( F_s \) and satisfies

\[
d\vec{m}_{\text{att}} = -\Gamma \vec{m}_{\text{att}} \, ds + F_s \, ds.
\]

(2.4)

The covariance has no dependence of the deterministic forcing and is the solution of the system

\[
0 = -\Gamma R_{\text{att}} - R_{\text{att}} \Gamma^* + \Sigma_u \Sigma_u^*.
\]

(2.5)

Making use of the fact that \( \Gamma \) is diagonal, one can easily solve

\[
\vec{m}_{\text{att}} = \int_{-\infty}^{t} \exp(-\Gamma(t-s)) F_s \, ds, \quad R_{\text{att}} = \frac{1}{2} \Gamma^{-1} \Sigma_u \Sigma_u^*.
\]

(2.6)

2.2. Filtering with noisy Lagrangian tracers

Given one realization of the velocity field, \( \vec{v} \geq 0 \), the trajectory of each one of the \( L \) noisy tracers can be modelled as the solution to the following SDE:

\[
d\vec{X}_l(s) = \vec{v}(\vec{X}_l(s), s) \, ds + \sigma_x dW_x^l(s)
\]

\[
= P_X(\vec{X}_l(s)) U_s \, ds + \sigma_x dW_x^l(s), \quad l = 1, \ldots, L.
\]

(2.7)

The \( \{W_x^l(s)\}_{l \leq L} \) are two-dimensional real-valued independent Wiener processes. They describe the random instrumental error over the tracers. In the language of filtering, they are called the observation errors.

We will also group all \( \vec{X}_l(s) \) into an observation vector \( \vec{X}_s \) of dimension \( Ld \times 1 \):

\[
\vec{X}_s := \begin{bmatrix} 
\vec{X}_1(s) \\
\vdots \\
\vec{X}_L(s) 
\end{bmatrix}.
\]

(2.8)

Based on (2.6) and (2.2), \( \vec{X}_s \) follows

\[
d\vec{X}_s = P_X(\vec{X}_s) U_s \, ds + \sigma_x dW_X(s),
\]

(2.9)

where \( P_X(\vec{X}_s) \) is a \( Ld \times |K| \) matrix-valued function of \( \vec{X}_s \) defined as follows with \( W_X(s) \):

\[
P_X(\vec{X}_s) = \begin{bmatrix} 
P_X(\vec{X}_1(s)) & \cdots & \exp(i\vec{k} \cdot \vec{X}_1(s)) \vec{r}_k & \cdots \\
\vdots & \ddots & \vdots & \vdots \\
P_X(\vec{X}_L(s)) & \cdots & \exp(i\vec{k} \cdot \vec{X}_L(s)) \vec{r}_k & \cdots 
\end{bmatrix}, \quad W_X(s) = \begin{bmatrix} 
W_1^X(s) \\
\vdots \\
W_L^X(s) 
\end{bmatrix}.
\]

(2.10)

From the previous discussion, filtering \( \vec{v}(\vec{x}, t) \) with \( (\vec{X}_1(s), \ldots, \vec{X}_L(s))_{s \leq t} \) is equivalent to filtering \( U_s \) using \( \vec{X}_{s \leq t} \), with their joint evolution described by (2.3) and (2.7).

The nonlinearity of this problem comes from the tracers, as the evolution of \( \vec{X}_s \) depends nonlinearly on itself in (2.7). However, once its realization is fixed, both (2.3) and (2.7) will be linear in \( U_s \). Moreover, the initial distribution of \( U_0 \) is taken as \( \pi_0^{\text{att}} \), which is Gaussian.

2137
Therefore our problem fits tightly into the condition of theorem 12.7 of [16]. Since this theorem is of pivotal importance, we attach one simplified version of it as theorem A.1. Based on its statement, conditioned on the observation of \( X_{s,t} \), \( U_t \) is distributed as \( \pi_{t|t} = \mathcal{N} (\tilde{m}_t, R_t) \).

The posterior mean, \( \tilde{m}_t \), and covariance, \( R_t \), are solutions to the following:

\[
\frac{dR_t}{dt} = -\Gamma R_t - R_t \Gamma^* + \Sigma - \sigma_x^{-2} R_t P(t) R_t \frac{dR_t}{dt}, \\
\frac{d\tilde{m}_t}{dt} = -\Gamma \tilde{m}_t dt + F_t dt + \sigma_x^{-2} R_t P_x(X_t) (dX_t - P_x(X_t) \tilde{m}_t dt),
\]

where we denote \( \Sigma := \Sigma_x \Sigma_x^* \) and \( P(t) = P_x^*(X_t) P_x(X_t), \) which has explicit random matrix entries as

\[
[P(t)]_{j,k} = \sum_{l=1}^L \exp((\bar{k} - \bar{j}) \cdot \tilde{X}_l(t))(r_j^* r_k).
\]

It is worth noting that since \( P_x \) is a periodic function, it is equivalent to observe either \( \tilde{X}_l(t) \) or its projection on torus \( \mathbb{T}^d = \{0, 2\pi \}^d \).

Without lost of generality, we assume the initial locations of the tracers \( \tilde{X}_l(0), \ldots, \tilde{X}_L(0) \) are distributed independently. Then conditioned on the realization of \( \tilde{v}_{t \geq 0}, \tilde{X}_l(s), \ldots, \tilde{X}_L(s) \) are conditionally independent processes. Their joint law conditioned on the realization of \( \tilde{v}_{t \geq 0} \) will be denoted as \( \mathbb{P}_{\tilde{v}_{t \geq 0}} \) and the expectation taken with it denoted as \( \mathbb{E}_{\tilde{v}_{t \geq 0}} \). Regarding regularity, \( \mathbb{P}_{\tilde{v}_{t \geq 0}} \) and other conditional laws in this paper are all regular versions of conditional probabilities; their existence relies on the fact that the underlying space is Polish, see theorem 6.3 [17]. In order to differentiate this distribution from the joint law of \( \tilde{v}_{t \geq 0} \) and \( (\tilde{X}_l(s))_{0 \leq s, l \leq L} \), we will call the latter the super-ensemble probability as it includes all realizations of \( \tilde{v}_{t \geq 0} \), and denote it as \( \mathbb{P} \).

### 2.3. Uncertainty reduction

Denote by \( \pi_{t|t} \) the posterior distribution of the velocity field with observation from tracers, which is a Gaussian distribution with mean \( \tilde{m}_t \) and covariance \( R_t \). One practical question is how much uncertainty is reduced in \( \pi_{t|t} \), compared with the unfiltered velocity field, meaning the absence of observations. Here, the prior distribution is given by the equilibrium distribution \( \pi_{t\text{att}} \), with mean \( \tilde{m}_t\text{att} \) and covariance \( R_t\text{att} \). The empirical information theory, as in [11,12,18,19], is applied to measure this uncertainty reduction.

In the information-theoretic framework, one natural way to measure the lack of information in one probability density \( q \), when the real distribution is \( p \), is through the relative entropy

\[
\mathcal{P}(p, q) = \int p \ln \frac{p}{q}.
\]

Despite the lack of symmetry in its arguments, the relative entropy \( \mathcal{P}(p, q) \) provides an attractive framework for assessing the uncertainty reduction as well as measuring model error due to its two ‘distance-like’ properties: (i) \( \mathcal{P}(p, q) \) is always positive unless \( p = q \), and (ii) it is invariant under general nonlinear change of variables. These properties make relative entropy an ideal measurement for uncertainty quantification purposes.

When both \( p \sim \mathcal{N}(\tilde{m}_p, R_p) \) and \( q \sim \mathcal{N}(\tilde{m}_q, R_q) \) are Gaussian, the relative entropy has an explicit formula [20,21]:

\[
\mathcal{P}(p, q) = \left[ \frac{1}{2} (\tilde{m}_p - \tilde{m}_q)^T R_q^{-1} (\tilde{m}_p - \tilde{m}_q) \right] + \frac{1}{2} \left[ \text{tr}(R_p R_q^{-1}) - N - \ln \det(R_p R_q^{-1}) \right],
\]

where \( N \) is the dimension of both distributions. The first term in brackets in (2.12) is called the ‘signal’ and measures the lack of information in the mean weighted by model covariance,
whereas the second term in brackets is called the ‘dispersion’ and involves only the covariance ratio.

For our purposes, of key importance is the so-called Bayesian-update interpretation of relative entropy. It states that if \( p = \pi_{t|t} \) is a posterior distribution conditioned on the observation \( X_{s \leq t} \) and \( q \) is the corresponding prior distribution, which is the case for \( q = \pi_{t \text{att}} \), then \( \mathcal{P}(\pi_{t|t}, \pi_{t \text{att}}) \) measures the additional information beyond \( \pi_{t \text{att}} \) gained by having observed \( X_{s \leq t} \); namely, the uncertainty reduction.

The above discussions and formulae also apply to complex-valued random vectors, as long as their conjugates are integrated properly. The relative entropy of a random variable \( X + Yi \) can be naturally defined as the relative entropy for the real-valued pair \((X, Y)\). More discussion and verifications of formula (2.12) in this setting can be found in appendix A.2.

Moreover, as the \( U_t \) has its complex entries come in as conjugate pairs, we can apply (2.12) with its mean and covariance under \( \pi_{t|t} \) and \( \pi_{t \text{att}} \):

\[
\mathcal{P}(\pi_{t|t}, \pi_{t \text{att}}) = \frac{1}{2} \left[ (\vec{m}_t - \vec{m}_{t \text{att}})^* R_{\text{att}}^{-1} (\vec{m}_t - \vec{m}_{t \text{att}}) \right] \quad \text{Signal}
\]

\[
+ \frac{1}{2} \left[ \text{tr}(R_t R_{\text{att}}^{-1}) - |K| - \ln \det(R_t R_{\text{att}}^{-1}) \right] \quad \text{Dispersion.} \quad (2.13)
\]

Another important quantity that also characterizes the amount of information of \( U_t \) that \( X_{s \leq t} \) can provide is their mutual information. This quantity is also widely used for other scenarios of uncertainty quantifications [22–25]. For two random variables \( X, Y \) with joint density \( p \) and marginal density \( p_X, p_Y \), the mutual information is defined as:

\[
I(X; Y) := \int p(dx, dy) \ln \frac{p(dx, dy)}{p_X(dx)p_Y(dy)}.
\]

In our scenario, we can show that \( I(U_t, X_{s \leq t}) = \mathbb{E}(\mathcal{P}(\pi_{t|t}, \pi_{t \text{att}})) \), i.e. the ‘super-ensemble’ expectation value of the relative entropy over all the observations [24]. With direct computation in lemma A.3, we also have a concise formula for this quantity:

\[
I(U_t, X_{s \leq t}) = \frac{1}{2} \ln \det R_{\text{att}} - \frac{1}{2} \mathbb{E} \ln \det R_t. \quad (2.14)
\]

Note that the bijection between \( U_t \) and \( \vec{v}_s \), so (2.14) is also \( I(\vec{v}_s, (\vec{X}_l(s))_{s \leq l, l \leq L}) \).

Note that the relative entropy \( \mathcal{P}(\pi_{t|t}, \pi_{t \text{att}}) \) is stochastic since it is based on a single realization of \( X_{s \leq t} \), while the mutual information, which averages with respect to all the realizations, is deterministic.

The final remark is that both the relative entropy and mutual information can be applied to any marginal parts of \( U_t \). As these marginal parts also have Gaussian distribution, it suffices to replace the vectors, matrices and dimension in (2.13) and (2.14) with a corresponding sub-vector, sub-matrix and dimension. One can read corollary 3.5 and its proof for a detailed explanation of this.

3. Statements of results

Our results consist of three parts.

The first issue we address before analyzing the filter is the locations of the tracers. This impacts dramatically how the filter performs. Indeed, if all the tracers are clustered in one location, then intuitively we should have little inference of the flow at a distance away. Then in this situation, the filter would perform very differently comparing to an ideal situation when tracers are uniformly distributed over the whole area. References such as [5, 7] also address similar issues in a different setting.
Considering the process for the noisy Lagrangian tracer trajectory given in (1.3), our first result indicates that the ideal situation is probably the most interesting case:

**Theorem 3.1.** For a.s. realization of \( \tilde{v}(\tilde{x}, s) \), \( s \geq 0 \) formulated by (1.1), the uniform distribution, \( \lambda \), on the d-dimensional torus \( \mathbb{T}^d := (0, 2\pi]^d \), is the unique invariant measure for the process \( \tilde{X}_l(t) := \tilde{X}_l(t) \mod \mathbb{T}^d \), for any fixed \( l \). Moreover, the tracer’s projection on \( \mathbb{T}^d \) is approximately ergodic under the conditional distribution for given \( \tilde{v}_l \geq 0 \), i.e. there exists a suitable constant \( q < 0 \) such that:

\[
\limsup_{n \to \infty} \frac{1}{n} \ln \| P_{\tilde{v}_l > 0}(\tilde{X}_l(n) \in \cdot) - \lambda \|_{TV} \leq q.
\]

Here, \( P_{\tilde{v}_l > 0}(\tilde{X}_l(n) \in \cdot) \) is the conditional distribution of \( \tilde{X}_l(n) \) for given \( \tilde{v}_l \geq 0 \). The choice of \( q \) is independent of \( \tilde{v}_l \geq 0 \) and the distribution of \( \tilde{X}_l(0) \). The notation \( \| \cdot \|_{TV} \) stands for total variation distance. Recall that the total variation distance between two measures \( P \) and \( Q \) is defined as:

\[
\| P - Q \|_{TV} = \sup_{|f| \leq 1} \mathbb{E}_P f - \mathbb{E}_Q f,
\]

with the supremum runs over all measurable functions \( f \), \( |f| \leq 1 \).

Since the tracers are mutually independent, by theorem 3.1, the distribution of the configuration of their projection on \( \mathbb{T}^d \) will approach \( \lambda^\otimes L := \lambda \otimes \cdots \otimes \lambda \) in a finite amount of time. It is therefore more interesting and theoretically convenient to assume the tracers are distributed at their steady state in the beginning.

On the other hand, in the view of (2.9) and (2.10), our filter process \((\tilde{m}_t, R_t)\) is a Markov process. Also by lemma A.4, \((\tilde{m}_t, R_t)\) are bounded in probability. Therefore even if the tracers are not distributed as \( \lambda^\otimes L \), we can shift the time until the point that \( \lambda^\otimes L \) is approximately reached. In other words, assuming the following loses little generality:

**Assumption 3.2.** The projection of the initial location of each tracer on \( \mathbb{T}^d \) is independently and uniformly distributed on \( \mathbb{T}^d \). The prior distribution is chosen as \( U_0 \sim \mathcal{N}(\tilde{m}_0, R_0) \) with \( |\tilde{m}_0| \leq M, m \sqrt{L}^{-1} I \leq R_0 \leq M I \) for some constants \( 0 < m < M \).

Under assumption 3.2, our second result discusses the limiting behaviour of the posterior mean and covariance \( \tilde{m}_L \) and \( R_L \) for large \( L \). Recall that a sequence of random variables \( X_L \) converges to \( a \) in probability, written as \( X_L \Rightarrow a \) in \( P \), if for any \( \epsilon > 0 \), \( P(|X_L - a| \geq \epsilon) \to 0 \) as \( L \to \infty \). We will say \( X_L \) approaches \( Y_L \) in \( P \) if \( X_L - Y_L \to 0 \) in \( P \). We will say a deterministic sequence \( x_L \) grows asymptotically as \( f(L) \) if their ratio goes to 1 as \( L \to \infty \). We will use \( \| \cdot \| \) to denote the \( L^2 \) norm of a matrix.

**Theorem 3.3.** When filtering a random incompressible flow \( \tilde{v}(\tilde{x}, s) \) described by (1.1) using \( L \) Lagrangian tracers under assumption 3.2, the following hold as \( L \to \infty \):

(ii) For any fixed \( t > 0 \) and a.s. realization of \( \tilde{v}_l \geq 0 \), the posterior covariance \( R_L \) in the conditional distribution with given \( \tilde{v}_l \geq 0 \):

\[
\| R_L^{-1} \| \| R_t - R_L \| \to 0 \quad \text{in } P_{\tilde{v}_l \geq 0}.
\]

The sequence \( R_L \) are deterministic diagonal matrices with

\[
[R_L]_{k,k} = \frac{\sigma_k^2}{d_k + \sqrt{d_k^2 + L\sigma^2_\epsilon^2 \sigma_k^2 |r_k|^2}},
\]

(3.1)
(ii) For any fixed \( t > 0 \), the difference between the posterior mean and the signal, \( \tilde{m}_t - U_t \), converges to 0 in \( \mathbb{P} \), the super-ensemble probability that integrates all realization of \( \tilde{v}_{s>0} \).

(iii) There exists a constant time \( s_0 \), such that for a.s. fixed realization of \( \tilde{v}_{s>0} \) \( t > s_0 \), \( \tilde{m}_t - U_t \) converges to 0 in \( \mathbb{P}_{\tilde{v}_{s>0}} \).

In other words, asymptotically \( \pi_{it} \) can be approximated by \( \mathcal{N}(U_t, R_L) \). We only need to plug in the formulas for relative entropy (2.13) and mutual information (2.14) to have formally the following corollary on the asymptotic behaviour of uncertainty reduction:

**Corollary 3.4.** Under the condition of theorem 3.3, the following hold as \( L \to \infty \):

(i) The dispersion part of the relative entropy in (2.13) approaches the following in \( \mathbb{P}_{\tilde{v}_{s>0}} \):

\[
\frac{1}{2} \sum_{k \in K} \left( \ln \left( \frac{1}{2} \left( 1 + \sqrt{1 + L \sigma_k^{-2} d_k^{-2} \sigma_k^{-2} |r_k|^2} \right) - 1 \right) \right).
\]

(ii) The signal part in (2.13) will converge to \( \frac{1}{2} (U_t - \tilde{m}^{\text{at}})^* R^{\text{at}}_{\text{att}}^{-1} (U_t - \tilde{m}^{\text{at}}) \) in \( \mathbb{P} \) for \( t > 0 \) and \( \mathbb{P}_{\tilde{v}_{s>0}} \) for a.s. \( \tilde{v}_{s>0} \) when \( t > s_0 \), with \( s_0 \) defined by theorem 3.3 part (iii).

(iii) The mutual information between the velocity field \( \tilde{v}_t \) and the observations \( (\tilde{X}_t(s))_{s \in l, l \leq L} \) grows asymptotically as \( \frac{1}{2} |K| \ln L \).

Notice that under \( \mathcal{N}(U_t, R_L) \), each component is independent as \( R_L \) is diagonal, so there is the following version of corollary 3.4 for marginal processes:

**Corollary 3.5.** Under the condition of theorem 3.3, let \( \tilde{K} \subset K \) be a subset that does not separate any conjugate pairs. Denoting \( \tilde{U}_t = [\cdots, \tilde{v}_k(t), \cdots]^T \tilde{K} \), then the following holds as \( L \to \infty \):

(i) The dispersion part of the relative entropy of \( \tilde{U}_t \), approaches the following in \( \mathbb{P}_{\tilde{v}_{s>0}} \):

\[
\frac{1}{2} \sum_{k \in \tilde{K}} \left( \ln \left( \frac{1}{2} \left( 1 + \sqrt{1 + L \sigma_k^{-2} d_k^{-2} \sigma_k^{-2} |r_k|^2} \right) - 1 \right) \right).
\]

(ii) The signal part of the relative entropy of \( \tilde{U}_t \), will converge to \( \frac{1}{2} (\tilde{U}_t - \tilde{m}^{\text{at}})^* \tilde{R}^{\text{at}}_{\text{att}}^{-1} (\tilde{U}_t - \tilde{m}^{\text{at}}) \) in \( \mathbb{P} \) for \( t > 0 \) and \( \mathbb{P}_{\tilde{v}_{s>0}} \) for a.s. \( \tilde{v}_{s>0} \) when \( t > s_0 \), with \( s_0 \) defined by theorem 3.3 part (iii).

(iii) The mutual information between the velocity field \( \tilde{v}_t \) and the observations \( (\tilde{X}_t(s))_{s \in l, l \leq L} \) grows asymptotically as \( \frac{1}{2} |\tilde{K}| \ln L \).

Our final remark is that all the statements of the filter \( \pi_{it} \), theorem 3.3 and corollary 3.4–3.5 will hold trivially if the observations of the velocity field are done at regularly spaced fixed points other than noisy Lagrangian tracers. See section 4.5 for details.

### 4. Proof of theorems

#### 4.1. Conditional geometric ergodicity of tracers

In this section we will prove theorem 3.1. We will use the following notation:

\[
\mathbb{P}_{\tilde{v}_{s>0}}^x (\cdot) := \mathbb{P}_{\tilde{v}_{s>0}} (\tilde{X}_t(n) \in \cdot | \tilde{X}_t(0) = x), \quad \forall x \in \mathbb{T}^d.
\]

Recall that \( \tilde{X}_t(n) \) is the process associated with the noisy Lagrangian tracer equation’s projection on \( \mathbb{T}^d \):

\[
\tilde{X}_t(t) = \tilde{X}_t(0) + \int_0^t \tilde{v}(\tilde{X}_s, s) ds + \int_0^t \tilde{a}_s dW_t^x(s) \mod \mathbb{T}^d.
\]
We also denote the supremum velocity in a velocity field as:

\[ |\vec{v}|_\infty := \sup_x |\vec{v}(x, s)|. \]

We write \( p : \mathbb{R}^d \to T^d \) as the canonical covering map and we denote \( p^{-1}x \) as the unique inverse image of \( x \) in \((0, 2\pi)^d\).

Also recall that a coupling of two measures is a joint measure with its marginal distributions being the two measures, respectively. The essence of the following proof is constructing a coupling between the trajectories of two tracers \( \vec{X}_t(s) \) starting from two different locations but transported by the same realization of \( \vec{v}_{s,N} \). By Goldstein’s theorem (section 14 [26]), we have

\[ \|\mathbb{P}_v^x - \mathbb{P}_v^y\|_{TV} \leq 2\mathbb{P}_{\vec{v}_{s,N}}(\text{two tracers are coupled into one trajectory from time } n), \]

so if a coupling exists such that two tracers are eventually coupled, we will have our claim. In a rough sense, we will construct this coupling one time step by one time step. The key argument in (4.2) below implies two tracers can be coupled at the end of one time step with a probability bounded away from 0 no matter where they start. Hence, the probability that the tracers are not coupled will decay geometrically in time. In order to show (4.2), we will show it is measured theoretically equivalent to the case when \( \vec{v}_s \) is trivial using the Girsanov theorem as long as \( \vec{v}_s \) is bounded in \( L^2 \).

**Proof.** To show the invariance for the measure \( \lambda \), noting that \( \vec{v}(x, s) \) is spatially periodic, it suffices for us to verify that locally the density \( f \equiv 1 \) is a solution to the Fokker–Plank equation for (2.6). Using the Fokker–Plank equation on \( T^d \), see section 8.1 [27], the evolution of a density under \( \mathbb{P}_{\vec{v}_{s,N}} \) is

\[ \frac{df}{dt} = -\sum_{i=1}^d \frac{\partial}{\partial x_i} [\vec{v}(x, t) f(x, t)] + \frac{\sigma^2}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} [\delta_{ij} f(x, t)]. \]

When \( f \equiv 1 \), due to \( \nabla \cdot \vec{v}_t \equiv 0 \), the right-hand side above is 0, hence \( f \equiv 1 \) is a time-independent invariant solution of the equation above. The uniqueness of the invariant measure is proved by the geometric ergodicity argument below.

To get the geometric ergodicity of the tracer, notice that \( \vec{v}(x, s) \) is formulated as in (2.2), and \( U_s \) is an OU process with strictly positive damping, so \( \vec{v}_s \) is ergodic. By Birkhoff’s ergodic theorem, if we denote the periodicity of \( \pi_t^{m0} \) as \( T \), then

\[ \lim_{i\to\infty} \frac{1}{T} \int_0^T |\vec{v}_s|_\infty ds = \frac{1}{T} \int_0^T \mathbb{E}_{\pi_t^{m0}}[P_X(\tilde{x}) \cdot U_s]_\infty ds \quad \text{P-a.s.} \]

Let \( N \) be the twice the value of the right-hand side, and denote

\[ k(\vec{v}, n) := \# \{ k : 0 \leq k \leq n - 1, \int_k^{k+1} |\vec{v}(s)|_\infty^2 ds < N \}. \]

Then using \( \int_0^n |\vec{v}|_\infty^2 ds \geq 2N(n - k(\vec{v}, n)) \), we can conclude

\[ \lim_{n \to \infty} \frac{k(\vec{v}, n)}{n} \geq \frac{1}{2} \quad \text{P-a.s.} \] \hspace{1cm} (4.1)

We fix a \( \vec{v}_{s,N} \) such that the event of (4.1) takes place. We will first prove the following claim: there exists a uniform \( c > 0 \) such that for any \( x \) and \( y \) in \( T^d \),

\[ \|\mathbb{P}_{\vec{v}_{s,N}}^{x} - \mathbb{P}_{\vec{v}_{s,N}}^{y}\|_{TV} \leq 2 - c \mathbb{1}_{\int_0^1 |\vec{v}(s)|_\infty^2 ds < N}. \] \hspace{1cm} (4.2)

This claim is trivial for \( \{ \int_0^1 |\vec{v}(s)|_\infty^2 ds \geq N \} \), so we assume the opposite holds. Consider the following joint processes on \( \mathbb{R}^d \),

\[ \vec{X}_s = p^{-1}x + \sigma_x W_s, \quad \vec{Y}_s = p^{-1}y + \sigma_y W'_s, \]

2142
where $W_s$ and $W'_t$ are two d-dimensional canonical Wiener processes. We denote their laws respectively as $Q_1$ and $Q_2$. Since $p^{-1}x$ and $p^{-1}y$ lie in $(0, 2\pi]^d$, so $|p^{-1}x - p^{-1}y| \leq 2\sqrt{d\pi}$. Then since $X_t$ and $Y_t$ are distributed as $N(p^{-1}x, \sigma_1^2 I_d)$, $N(p^{-1}y, \sigma_2^2 I_d)$, with $|p^{-1}x - p^{-1}y| \leq 2\sqrt{d\pi}$ we can find a universal $a > 0$ such that
\[
\|N(p^{-1}x, \sigma_1^2 I_d) - N(p^{-1}y, \sigma_2^2 I_d)\|_{TV} \leq 2 - a, \quad \forall x, y \in \mathbb{T}^d.
\]
Due to the relation between coupling and total variation, see theorem (I.5.2) in [26], we can find a maximal coupling $Q_{1,2}^{x,y}$ of the random variables $\tilde{X}_1, \tilde{Y}_1$: $Q_{1,2}^{x,y}(\tilde{X}_1 = \tilde{Y}_1) \geq a$. Then conditioned on the value of $\tilde{X}_1, \tilde{Y}_1$, we can extend it to a coupling, $Q_{1,2}^{x,y}$, of processes $(\tilde{X}_s)_{s \leq 1}, (\tilde{Y}_s)_{s \leq 1}$, explicitly defined as:
\[
Q_{1,2}^{x,y}(\tilde{X}_s \leq 1, \tilde{Y}_s \leq 1) = \int_{v \in \mathbf{z}} Q^{x,y}(\tilde{X}_1 = u, \tilde{Y}_1 = v) Q_{1,2}^{x,y}(\tilde{X}_s \leq 1, \tilde{Y}_s \leq 1 | u, v).
\]
Then as $Q_{1,2}^{x,y}$ is the marginal distribution of $(\tilde{X}_1, \tilde{Y}_1)$ for $Q_{1,2}^{x,y}$:
\[
Q_{1,2}^{x,y}(\tilde{X}_1 = \tilde{Y}_1) = Q_{1,2}^{x,y}(\tilde{X}_1 = \tilde{Y}_1) \geq a, \quad \forall x, y \in \mathbb{T}^d.
\]
(4.3)

Now denote:
\[
z_0(t) := \exp(\mathcal{M}_t - \frac{1}{2} \mathcal{M}_0), \quad \mathcal{M}_t = \sigma_1^{-1} \int_0^t \tilde{w}(\tilde{X}_s, s) dW_s + \tilde{v}(\tilde{Y}_s, s) dW'_s.
\]
(4.4)

Since $\int_0^t (|\tilde{w}(\tilde{X}_s, s)|^2 + |\tilde{v}(\tilde{Y}_s, s)|^2) ds \leq 2 \int_0^t |\tilde{w}_s|^2 ds < 2N$, the Novikov condition (corollary 3.5.13 [28]), holds for $z_0(t)$, meaning $z_0(t)$ is a martingale. By the Girsanov theorem, under measure $\mu$ defined by $d\mu := z_0(1) dQ_{1,2}^{x,y}$, the processes
\[
\tilde{W}_t = W_t - \sigma_1^{-1} \int_0^t \tilde{w}(\tilde{X}_s, s) ds, \quad \tilde{W}'_t = W'_t - \sigma_1^{-1} \int_0^t \tilde{v}(\tilde{Y}_s, s) ds
\]
are canonical Wiener processes. The $\tilde{X}_1, \tilde{Y}_1$ on the other hand are solutions to the equations:
\[
\tilde{X}_t = p^{-1}x + \int_0^t \tilde{w}(\tilde{X}_s, s) ds + \sigma_1 \tilde{W}_s, \quad \tilde{Y}_t = p^{-1}y + \int_0^t \tilde{v}(\tilde{Y}_s, s) ds + \sigma_1 \tilde{W}'_s.
\]
So $\tilde{X}_s$ and $\tilde{Y}_s$ can be seen as lifts of the processes described by (2.6) starting from point $x$ and $y$. In other words, the joint law of $(p\tilde{X}_1, p\tilde{Y}_1)$ is a coupling of $P^x_{1,1}$ and $P^y_{1,1}$. Therefore
\[
\|P^x_{1,1} - P^y_{1,1}\|_{TV} = \sup_{|f| \leq 1} P^x_{1,1} f - P^y_{1,1} f = \sup_{|f| \leq 1} E_{\mu}(f(\tilde{X}_1) - f(\tilde{Y}_1)) \leq 2\mu(\tilde{X}_1 \neq \tilde{Y}_1).
\]
(4.5)

On the other hand, let $M = 2\sigma_1^{-1}\sqrt{N/d}$, since under $Q_{1,2}^{x,y}$ for $t \leq 1$,
\[
(Q_{1,2}^{x,y})_t = E_{\tilde{X}_t} \mathcal{M}_t = \sigma_1^{-2} \mathcal{M}_0 \int_0^t (|\tilde{w}(\tilde{X}_s, s)|^2 + |\tilde{v}(\tilde{Y}_s, s)|^2) ds \leq 2\sigma_1^{-2} \int_0^t |\tilde{w}_s|^2 ds \leq 2\sigma_1^{-2} N.
\]
Using the definition of $z_0(1)$, (4.4), and Chebyshev’s inequality,
\[
Q_{1,2}^{x,y}(z_0(1) \leq \exp(-M - \sigma_2^{-2} N)) \leq Q_{1,2}^{x,y}(M_t \leq M) = \frac{(\mathcal{M}_1)_t}{M^2} = \frac{1}{2}\frac{a}{\sigma_1^2}.
\]
From this and (4.3) we derive
\[
\mu(\tilde{X}_1 = \tilde{Y}_1) = \int_1 \chi_{\tilde{X}_1 = \tilde{Y}_1} z_0(1) \cdot dQ_{1,2}^{x,y}
\]
\[
\geq Q_{1,2}^{x,y}(\tilde{X}_1 = \tilde{Y}_1, z_0(1) \geq \exp(-M - \sigma_2^{-2} N)) \exp(-M - \sigma_2^{-2} N)
\]
\[
\geq (Q_{1,2}^{x,y}(\tilde{X}_1 = \tilde{Y}_1) - Q_{1,2}^{x,y}(z_0(1) \leq \exp(-M - \sigma_2^{-2} N)) \exp(-M - \sigma_2^{-2} N)
\]
\[
\geq \frac{1}{2}\frac{a}{\sigma_1^2} \exp(-M - \sigma_2^{-2} N).
\]
Based on (4.5), \( c = a \exp(-M - \sigma^2 N) > 0 \) satisfies our claim (4.2).

The next step is applying (4.2) at each time step and using the Markov property to form an induction. Denote
\[
Q^\lambda_{1,n} = \mathbb{P}^\lambda_{1,n-1} \land \lambda, \quad \tilde{P}^\lambda_{1,n} = \mathbb{P}^\lambda_{1,n-1} - Q^\lambda_{1,n}, \quad \tilde{\lambda}_{n-1} = \lambda - Q^\lambda_{1,n-1}.
\]
In other words, the density of \( Q^\lambda_{1,n-1} \) is the minimum of the ones of \( \mathbb{P}^\lambda_{1,n-1} \) and \( \lambda \); it represents the maximum volume of measure that \( \mathbb{P}^\lambda_{1,n-1} \) and \( \lambda \) can be coupled with. Notice that,
\[
\mathbb{P}^\lambda_{1,n-1}(\mathbb{P}^\lambda_{1,n}) = \tilde{\lambda}_{n-1}(\mathbb{P}^\lambda_{1,n}) = \frac{1}{2} \mathbb{P}^\lambda_{1,n-1} - \lambda \|TV.
\]
Then by letting
\[
\mathbb{P}^\lambda_{1,n-1}(dx, dy) = \begin{cases} 1 \mathbb{P}^\lambda_{1,n-1}(dy) \tilde{\lambda}_{n-1}(dy) & \tilde{\lambda}_{n-1}(\Omega) > 0; \\
0 & \tilde{\lambda}_{n-1}(\Omega) = 0,
\end{cases}
\]
one can verify that \( \mathbb{P}^\lambda_{1,n-1} \) is a coupling of \( \mathbb{P}^\lambda_{1,n-1} \) and \( \tilde{\lambda}_{n-1} \); that is, its marginal distribution is \( \mathbb{P}^\lambda_{1,n-1} \) and \( \tilde{\lambda}_{n-1} \) in the \( x \) and \( y \) direction.

Notice that by the Markov property, using the shift operator \( \theta : (\theta^i v)(s) = \tilde{v}(s + n) \), the following equation holds by conditioning on the value of \( \tilde{X}_i(n-1) \),
\[
\mathbb{P}^\lambda_{1,n-1} f = \int \mathbb{P}^\lambda_{1,n-1}(dy) \mathbb{P}^\lambda_{1,n-1} f.
\]
Therefore for any measurable \( f, |f| \leq 1 \), making use of the fact that \( \lambda \) is invariant under \( \mathbb{P}^\lambda_{1,n-1} \),
\[
\mathbb{P}^\lambda_{1,n-1} f - \lambda f = \int \mathbb{P}^\lambda_{1,n-1}(dy) \mathbb{P}^\lambda_{1,n-1} f - \lambda \mathbb{P}^\lambda_{1,n-1} f
\]
\[
= \int \mathbb{P}^\lambda_{1,n-1}(dy) \mathbb{P}^\lambda_{1,n-1} f - \tilde{\lambda}_{n-1}(dz) \mathbb{P}^\lambda_{1,n-1} f
\]
\[
= \int \mathbb{P}^\lambda_{1,n-1}(dy, dz) (\mathbb{P}^\lambda_{1,n-1} f - \mathbb{P}^\lambda_{1,n-1} f)
\]
\[
\leq \frac{1}{2} \mathbb{P}^\lambda_{1,n-1} - \lambda \|TV(2 - c1_{\lambda_{n-1}}(\theta^i v)_{x < \lambda})).
\]
We apply (4.6) in the last step. By taking the supreme over all measurable \( f, |f| \leq 1 \), we have
\[
\mathbb{P}^\lambda_{1,n-1} f - \lambda \|TV \leq (1 - c \frac{1}{2} 1_{\lambda_{n-1}}(\theta^i v)_{x < \lambda}) \|\mathbb{P}^\lambda_{1,n-1} - \lambda \|TV.
\]
Using induction, \( \mathbb{P}^\lambda_{1,n-1} f - \lambda \|TV \leq 2(1 - \frac{1}{2})^k(\tilde{\lambda}) \). Due to the Markov property on the initial condition,
\[
\mathbb{P}^\lambda_{1,n}(\tilde{X}_i(n) \in A) = \int \mathbb{P}(\tilde{X}_i(0) \in dx) \mathbb{P}^\lambda_{1,n}(A) dx.
\]
So applying Jason’s inequality,
\[
\|\mathbb{P}^\lambda_{1,n}(\tilde{X}_i(n) \in \cdot) - \lambda \|TV \leq \int \mathbb{P}(\tilde{X}_i(0) \in dx) (\mathbb{P}^\lambda_{1,n} \mathbb{P}^\lambda_{1,n-1} - \lambda \|TV \leq 2(1 - c \frac{1}{2})^k(\tilde{\lambda}) \).
\]
Recalling (4.1), we have
\[
\limsup_{n \to \infty} \frac{1}{n} \ln \|\mathbb{P}^\lambda_{1,n}(\tilde{X}_i(n) \in \cdot) - \lambda \|TV \leq - \frac{1}{2} \ln \left(1 - \frac{c}{2}\right) < 0.
\]
\[\square\]
(iii) Defining models. For example, [31] shows the geometric ergodicity for a particle model governed by the Stokes’ law while the underlying velocity field is random; its method can be easily applied to show the geometric ergodicity of the pair \((U_s, X_s)\). Theorem 3.1 here considers a rather different problem, since the statement takes place in the conditional distribution, the propagation of \(\tilde{X}_l(s)\) is inhomogeneous and depends on the realization of the underlying velocity \(\tilde{v}\). In a sense, the coupling argument used here is essentially verifying the minorization condition in [29], playing the same role as lemma 2.3 in [30] for noisy Lagrangian tracers.

4.2. The mean field dynamics of the filter

In this section we will prove theorem 3.3. Let us first briefly explain the intuition. Since the posterior covariance \(R_t\) follows its own dynamics (2.9) but enters in evolution of \(\bar{m}_t\), through (2.10), it is intuitive to study \(R_t\) first. Recall from (2.9) that \(R_t\) is the solution to a Riccati equation with random coefficients:

\[
dR_t = [-\Gamma R_t - R_t \Gamma^* + \Sigma - \sigma^2 \bar{R} P(t) R_t] dt.
\]

The randomness comes from the tracers \(X_t\), which enter the evolution by the second order coefficient

\[
P(t) = \sum_{l=1}^{L} P_X(\tilde{X}_l(t)) P_X(\tilde{X}_l(t)).
\] (4.7)

The second equality above comes from the fact that \(P_X(X_t)\) is a concatenation of \(P_X(\tilde{X}_l(t))\) as in (2.8). The randomness of \(P(t)\) also makes \(R_t\) a random process. However, under assumption 3.2, by theorem 3.1, \(\tilde{X}_l(t), \ldots, \tilde{X}_L(t)\) are i.i.d. samples from the uniform distribution \(P_{\tilde{v}_{l}, \rho}\). Therefore the value of \(P(t)\) as a sum of i.i.d. random variables will not deviate too far from its mean in \(P_{\tilde{v}_{l}, \rho}\) when \(L\) is large. So intuitively, \(R_t\) will stay close to the solution of the formal mean field average of (2.9) as

\[
dR_t = [-\Gamma R_t - R_t \Gamma^* + \Sigma - \sigma^2 L R_t M R_t^*] dt, \quad M := \mathbb{E}_{\tilde{v}} P_X(X) P_X(X).
\] (4.8)

On the other hand, matrix Riccati equations are known to be a contraction for Hermitian matrices. Therefore, \(R_t\) will also converge to \(R_L\) in time, which is the stable point of (4.8); it will also be shown in the proof of theorem 3.3 part i) that this convergence is exponentially fast with a rate which scales as \(\sqrt{L}\). Using this accelerating convergence, \(R_t\) approaches \(R_L\) very rapidly.

Since we will deal with variables that depend on \(L\), we will say some sequence \(x_L\) is of the order \(f(L)\), written as \(x_L \sim f(L)\), if \(M f(L) \geq x_L \geq m f(L)\) for some constants \(M, m > 0\) as \(L \to \infty\).

The next lemma explicitly computes the value of \(M\) and \(R_L\), and also gives probabilistic control over the deviation of \(P(t)\) from \(L M\).

Lemma 4.2. (i) \(M\) in (4.8) is a diagonal matrix with the \(k\)-th entry \(|r_k|\).

(ii) The stable point of (4.8), \(R_L\), is a diagonal matrix with the \(k\)-th entry being

\[
[R_L]_{k,k} = \frac{\sigma^2_k}{d_k + \sqrt{d_k^2 + \lambda \sigma^2_k |r_k|^2}}.
\]

(iii) Defining \(\overline{P}_L = P(s) - L M\), \(q_L := L^{-1} \sup_{s \leq t} \|\overline{P}_L\|\), then for any \(q > 1/2\), \(L^{1-q} q_L \to 0\) in \(P_{\tilde{v}_{l}, \rho}\) for a.s. \(\tilde{v}_{l} \geq 0\).
Proof. It suffices for us to prove each part with a fixed realization of \( \nu_{i \geq 0} \) so that the claim of theorem 3.1 holds.

**Part (i)** Recall that by (2.11), \( P(t) \) has each of its matrix entries explicitly as:

\[
(P(t))_{j,k} = \sum_{l=1}^{L} \exp(i(\vec{k} - \vec{j}) \cdot \vec{X}(t))(\vec{r}_l^j \vec{r}_l^k).
\]

When \( j = k \), the quantity above is a constant, \( L|\vec{r}_l|^2 \). Hence \( M \) has the described diagonal terms.

When \( \vec{j} = \vec{k}, j \neq k \), the quantity above is 0 as \( \vec{r}_l \perp \vec{r}_k \). (This is why we require \( \vec{r}_l \) to be orthogonal to each other if they share the same wavenumber as in section 2.)

When \( j \neq k \), notice that if we let \( Y_l = \exp(i(\vec{k} - \vec{j}) \cdot \vec{X}(t)) \), then using \( \vec{X}(t) \sim \lambda \) under \( \mathbb{P}_{\nu_{i \geq 0}} \), we have

\[
\mathbb{E}_{\nu_{i \geq 0}} Y_l = \frac{1}{|\mathcal{S}|} \int_{\mathcal{S}} \exp(i(\vec{k} - \vec{j}) \cdot x) \, dx = 0.
\]

In summary, \( M \) is diagonal.

**Part (ii)** Since (4.8) has all matrices being diagonal, we can solve the equation in each diagonal entry, which is solving \( 0 = -2d_k x + \sigma_k^2 - L\sigma_k^2|\vec{r}_l|^2 x^2 \). Directly solving these equations gives (ii).

**Part (iii)** Notice that \( \tilde{P}_s := P(s) - LM \) has its diagonal terms being 0 and its off-diagonal terms the same as \( P(s) \). In order to show our claim of \( q_L \), it suffices to show \( L^{-q}([\tilde{P}_s]_{j,k})_{s \leq t} \) converge to the constant 0 process in \( \mathbb{P}_{\nu_{i \geq 0}} \), because by the following we have our claim proved:

\[
L^{-q} q_L = \sup_{s \leq t} L^{-q} \| \tilde{P}_s \| \leq |K| \max_{j \neq k} \sup_{1 \leq r} L^{-q} \| \tilde{P}_s \|_{j,k} \xrightarrow{L \to \infty} 0 \quad \text{in} \quad \mathbb{P}_{\nu_{i \geq 0}}.
\]

Since \( z_{j}^{l} := L^{-q}([\tilde{P}_s]_{j,k} \) are continuous in \( s \), it suffices to verify the conditions of theorem 7.5 of [32]. This verification is done by lemma A.5. □

**Proof of theorem 3.3, part (i).** Fix any \( 0 < c_0 \leq \frac{1}{2} \min_k |\vec{r}_k|^2 \). In the following, if we say some constant is uniform, then that constant works for all \( c_0 \leq \frac{1}{2} \min_k |\vec{r}_k|^2 \) because only a bounded set of \( c_0 \) is considered. We introduce this technicality because we may decrease \( c_0 \) to a smaller value which depends on these constants in part (iii) of this theorem. For part (i) and (ii), any fixed \( c_0 \leq \frac{1}{2} \min_k |\vec{r}_k|^2 \) will suffice.

Recall that

\[
q_L := L^{-1} \sup_{s \leq t} \| P(s) - LM \|,
\]

one has

\[
L(M - q_L I) \leq P(s) \leq L(M + q_L I), \quad s \leq t.
\]

We will focus on the event \( \{q_L \leq c_0 \} \), which by lemma 4.2 (iii) has \( \mathbb{P}_{\nu_{i \geq 0}} \) probability goes to 1.

By the comparison theorem of Riccati equations (theorem 4.1.4 of [33]), if we let \( R_{s,t} \) be the solution of

\[
dR_{s,t} = [-\Gamma R_{s,t} - R_{s,t} \Gamma + \Sigma - L\sigma_s^{-2}R_{s,t}(M \mp q_L I)R_{s,t}]dt, \quad R_{s,s} = 0;
\]

then \( R_t \) is bounded by the two: \( R_{s,t} \leq R_t \leq R_{s,s}, \quad s \leq t \).

On the other hand, since \( R_{s,t} \) follows (4.9), according to theorem 8.5 [34], they will converge in time to the stable points of (4.9), respectively, denoted by \( R_{s,t} \), which are the solutions to the algebraic Riccati equations:

\[
0 = -\Gamma R_{s,t} - R_{s,t} \Gamma + \Sigma - L\sigma_s^{-2}R_{s,t}(M \mp q_L I)R_{s,t}.
\]

2146
or explicitly:

$$R_{\pm} = \begin{bmatrix} \cdots & \sigma_k^2 \vdots & \vdots \\ d_k + \sqrt{d_k^2 + L\sigma_k^{-2} (\bar{r}^2_k)^+} \\ \vdots & \cdots \end{bmatrix}.$$  

By theorem 8.5 of [34], when $q_L \leq c_0$ the rate that $R_{\pm,t}$ converges to $R_{\pm}$ in Thompson’s metric is at least

$$\|L\sigma_k^{-2} (M - c_0I)^{1/2} \Sigma(M - c_0I)^{1/2}\| \geq \frac{1}{2} L\sigma_k^{-2} \min_k |\bar{r}_k| \sigma_k^2 =: \mu_L^2.$$  

In other words: $p(R_{\pm,t}, R_{\pm}) \leq \exp(-2\mu_L t)p(R_0, R_{\pm})$. According to [34], for two positive semidefinite Hermitian matrices $A, B$, Thompson’s metric is defined as:

$$p(A, B) := \ln \inf\{ t : A \leq tB, B \leq tA \}.$$  

Since $ML^{-1/2} \leq R_{\pm,0} \leq ML$, when $q_L \leq c_0$ there is a uniform constant $D_0$ such that $p(R_0, R_{\pm}) \leq |\ln(D_0)\sqrt{L}|$. Therefore $p(R_{\pm,t}, R_{\pm}) \leq |\ln(D_0)\sqrt{L}| \exp(-2\mu_L t)$. Note that if $p(A, B) \leq a$ for two Hermitian matrices, then by definition of Thompson’s metric, $(e^{-a} - 1)B \leq A - B \leq (e^{a} - 1)B$, which implies $\|A - B\| \leq (e^{a} - 1)\|B\|$. Hence, $\|R_{\pm,t} - R_{\pm}\| \leq \|R_{\pm}\|[\exp(\ln(D_0)\sqrt{L}) \exp(-2\mu_L t)] - 1].$

Notice the following bound holds for all $0 \leq x \leq 1, a > 0$,

$$\exp(ax) - 1 = ax + \frac{1}{2}(ax)^2 + \frac{1}{3!}(ax)^3 + \ldots \\
\leq x[a + \frac{1}{2}a^2 + \frac{1}{3}a^3 + \ldots] \leq e^a x.$$  

Assuming $q_L \leq c_0$, as $\|R_{\pm}\| \sim 1/\sqrt{L}$ we have for a uniform constant $D_1$,

$$\|R_{\pm,t} - R_{\pm}\| \leq D_0\|R_{\pm}\|\sqrt{L} \exp(-2\mu_L t) = D_1 \exp(-2\mu_L t).$$  

On the other hand, according to the formulation of $R_{\pm}$, which is $C^1$ in $q_L$, there is a uniform constant $C_1$ such that $\|R_{\pm} - R_{\pm}\| \leq C_1 \sqrt{L}^{-1} q_L$ when $q_L \leq c_0$. Then from $R_{\pm,t} - R_{\pm,t} \leq R_{\pm,t} - R_{\pm}$, we can verify that when $q_L \leq c_0$,

$$\|R_{\pm,t} - R_{\pm}\| \leq \|R_{\pm,t} - R_{\pm}\| \vee \|R_{\pm,t} - R_{\pm}\| \leq C_1 \sqrt{L}^{-1} q_L + D_1 \exp(-2\mu_L t).$$  

On the other hand, as $\|R_{\pm,1}\| \sim \sqrt{L}$, there are some constants $\tilde{C}_1, \tilde{D}_1$ so that for any $\xi > 0$,

$$\mathbb{P}_{\tilde{V}_{\geq 0}}(\|R_{\pm,1}\| > \xi) \leq \mathbb{P}_{\tilde{V}_{\geq 0}}(q_L \leq c_0) + \mathbb{P}_{\tilde{V}_{\geq 0}}(\tilde{C}_1 q_L + \tilde{D}_1 \sqrt{L} \exp(-2\mu_L t) > \xi),$$  

both probabilities on the right go to 0 by lemma 4.2 and $\mu_L \sim \sqrt{L}$ as $L \to \infty$, therefore we conclude our claim. □

Since $R_t := \mathbb{E}(U_t - \bar{m}) \cdot (U_t - \bar{m})^*|X_{t\leq t})$, using its convergence in the super-ensemble probability $\mathbb{P}$ easily leads to the following proof of theorem 3.3, part (ii).

**Proof of theorem 3.3, part (ii).** Because $\mathbb{E}(U_t - \bar{m})^2|X_{t\leq t}) = \text{tr}(R_t)$, by part (i) of theorem 3.3, the law of total probability, i.e. $\mathbb{E}Y = \mathbb{E}_{\tilde{V}_{\geq 0}} Y$, and the dominated convergence theorem, for any $\xi > 0$,

$$\mathbb{P}(\|R_t\| > \xi) = \mathbb{E}_{\tilde{V}_{\geq 0}}(\|R_t\| > \xi) \to 0, \quad \text{as} \quad L \to \infty.$$  

2147
In order to show the convergence of \( P \) applied:

\[
\mathbb{P}(|U_t - \bar{m}_t| \geq \xi) \leq \mathbb{P}(|\text{tr}(R_t) \geq \xi^4) + \mathbb{P}(|\text{tr}(R_t) \leq \xi^4, |U_t - \bar{m}_t| \geq \xi)
\]

By letting \( L \to \infty \) then \( \xi \to 0 \), we have shown the claim.

\[ \square \]

### 4.3. Convergence of posterior mean

In order to show the convergence of \( \bar{m}_t \) towards \( U_t \) in \( \mathbb{P}_{\nu \geq 0} \), we need to do some computations. The intuition can be revealed by replacing \( dX_t \) in (2.10) by its formulation (2.7), so we rewrite (2.10) as

\[
d\bar{m}_t = [-\Gamma \bar{m}_t + F_t + \sigma_t^{-2} R_t P(t)(U_t - \bar{m}_t)]dt + \sigma_t^{-1} R_t P(t)(X_t)dW_t(t). \tag{4.12}
\]

This can be seen as a linear process with random perturbation. Its damping coefficient is \( \Gamma + \sigma_t^{-2} R_t P(t) \), which by part (i) of theorem 3.3 can be seen as \( \Gamma + \sigma_t^{-2} L R_t L M \) plus some deviations. As \( L \) increases, \( \Gamma + \sigma_t^{-2} L R_t L M \) is of the order \( \sqrt{L} \), which is also the order of force that will apply to \( \bar{m}_t \), if it deviates from \( U_t \).

**Proof of theorem 3.3, part (iii).** First fix a realization of \( \nu_{\geq 0} \) such that part (i) holds. We simplify the notation of (4.12) by defining a new \( d \)-dimensional vector process \( w_t \):

\[
dw_t := \frac{1}{\sqrt{L}} P^*_{(X)}dW(t) = \frac{1}{\sqrt{L}} \begin{bmatrix} \vdots \\
\sum_{l=1}^{L} e^{i(j-\bar{j})\bar{\bar{x}}(i,t)} \bar{P}^*_{(X)} dW^*_{k} \end{bmatrix}.
\]

The quadratic covariation of \( w_t \) is

\[
\langle dw_t, dw_t \rangle = \frac{1}{L} \left[ \sum_{l=1}^{L} e^{i(j-\bar{j})\bar{\bar{x}}(i,t)} \bar{P}^*_{(X)} dW^*_{k} \right] dt = \frac{1}{L} P(t)dt. \tag{4.13}
\]

Notice that \( L^{-1} P(t) \) has each matrix entry of norm at most 1,

\[
\langle dw_t, dw_t \rangle \leq L^{-1} \text{tr}(P(t)) \leq |K|.
\]

Subtracting the equation for \( dU_t \) in (2.3) from \( d\bar{m}_t \) in (4.12), we have the evolution of the difference \( \bar{n}_t = U_t - \bar{m}_t \):

\[
d\bar{n}_t = -(\Gamma + \sigma_t^{-2} R_t P(t))\bar{n}_t dt + \Sigma_i dB_i + \sigma_t^{-1} \sqrt{L} R_t dw_t.
\]

Let \( \Psi(s, t) \) be the fundamental matrix group generated by \( D_t := -d_0 I - \sigma_t^{-2} R_t P(t) \), that is if \( \bar{x}_t = D_t \bar{x}_t \), then \( \bar{x}_s = \Psi(s, t)\bar{x}_t \), \( \forall s \leq t \). Then by the Duhamel formula,

\[
\bar{n}_t = \Psi(0, t)\bar{n}_0 - \int_{0}^{t} \Psi(s, t) \Sigma_i dB_i + \sigma_t^{-1} \sqrt{L} \int_{0}^{t} \Psi(s, t) R_t dw_s. \tag{4.14}
\]

Recall that all of the derivation of the proof for part (i) holds as long as

\[
q_L = L^{-1} \sup_{s \leq t} \| P(t) - LM \| \leq c_0,
\]
where \(c_0\) is any small fixed number. We define the stopping time \(\tau := \inf\{s \geq 0, L^{-1} \| \tilde{\Pi}_s \| \geq c_0\}\). Then the event \(B := \{\tau \leq t\}\) indicates \(q_L \leq c_0\). By (4.14) and the union bound,
\[
\mathbb{P}_{\tilde{\nu}_{\Delta \Omega}}(\tilde{\nu}_i \geq 3\epsilon) \leq \mathbb{P}_{\tilde{\nu}_{\Delta \Omega}}(B^{+}) + \mathbb{P}_{\tilde{\nu}_{\Delta \Omega}}(|\Psi(0, t) \tilde{\nu}_i| \geq \epsilon, B)
\]
\[
\geq \epsilon, B) + \mathbb{P}_{\tilde{\nu}_{\Delta \Omega}}\left( \left| \int_{0}^{t} \Psi(s, t) \Sigma_{s} d\nu_{s} \right| \geq \epsilon, B \right)
\]
\[
+ \mathbb{P}_{\tilde{\nu}_{\Delta \Omega}}\left( \left| \sigma_{s}^{-1} \sqrt{\ell} \int_{0}^{t} \Psi(s, t) R_{s} d\nu_{s} \right| \geq \epsilon, B \right).
\]
In order to show \(\tilde{n}_i \rightarrow 0\) in \(\mathbb{P}_{\tilde{\nu}_{\Delta \Omega}}\), it suffices to show for any \(\epsilon > 0\), each of the four probabilities on the right above goes to 0. Using lemma 4.2, we know \(\mathbb{P}_{\tilde{\nu}_{\Delta \Omega}}(B^{+}) \rightarrow 0\) as \(L \rightarrow \infty\).

In order to bound the remaining terms, let us bound \(||\Psi(s, t)\|\) when the event \(\{\tau \geq t\}\) holds first. Let \(x_t = \Psi(s, t)x_s\), then \(x_t = -(d_0 + \sigma_x^{-2} R_s P(t)) x_s\). Recall that based on the derivation in the proof of part (i), (4.10) holds when \(\{t \leq \tau\}\). Notice this implies the following hold for \(s \leq t\) and some uniform constants \(D_1, C_1\) and \(C_1\):
\[
\| R_{s} \| \leq D_1 \exp(-2\mu_L s) + c_0 C_1 \sqrt{\ell}^{-1}, \quad \| R_{L} \| \leq \tilde{C}_1 \sqrt{\ell}^{-1}, \quad \| \tilde{P}_s \| \leq c_0.
\]
Hence there are uniform constants \(C_2, D_2\) such that when \(s \leq t \leq \tau\),
\[
\frac{d}{d s} |x_{s}|^2 = \text{Re}(2x_{s}^{*} x_{s}) = -2x_{s}^{*}(\Gamma + \sigma_x^{-2} L R_L M + \sigma_x^{-2} L \text{Re}(\tilde{R}_s M + (R_L + \tilde{R}_s) \tilde{P}_s)x_s
\]
\[
\leq -2x_{s}^{*}(\Gamma + \sigma_x^{-2} L R_L M)x_s + 2\sigma_x^{-2} L \left[ D_1 \exp(-2\mu_L s) + c_0 C_1 \sqrt{\ell}^{-1}\right]\| M \|
\]
\[
+ c_0(\tilde{C}_1 \sqrt{\ell}^{-1} + c_0 C_1 \sqrt{\ell}^{-1} + D_1 \exp(-2\mu_L s)) \| x_s \|^2
\]
\[
\leq -2x_{s}^{*}(\sigma_x^{-2} L R_L M)x_s + 2[c_0 C_2 \sqrt{\ell} + L D_2 \exp(-2\mu_L s)]\| x_s \|^2.
\]
Since \(R_L M\) is a diagonal matrix with all real positive entries, the following holds for an \(a > 0\):
\[
\sigma_x^{-2} x_s^{*} L R_L M x_s \geq |x_s|^2 \min_k \frac{\sigma_x^{-2} L a |\tilde{r}_k|^2}{d_k + \sqrt{d_k^2 + \sigma_x^{-2} L a |\tilde{r}_k|^2}} \geq |x_s|^2 a \sqrt{\ell} \sigma_x |\tilde{r}_k|.
\]
If we let \(d_L = \frac{1}{2} a \sqrt{\ell} \sigma_x |\tilde{r}_k|\) and choose a \(c_0 \leq \frac{1}{2} C_2^{-1} a\), then we can simplify the inequality of \(\frac{d}{d s} |x_{s}|^2\) above by another bound:
\[
\frac{d}{d s} |x_{s}|^2 \leq -2(d_L - L D_2 \exp(-2\mu_L s))\| x_s \|^2.
\]
This choice of \(c_0\) is legitimate if we recall that in the first line of the proof of part (i), \(c_0\) can be any fixed positive small number.

Using Gronwall’s inequality and \(\mu_L \sim \sqrt{\ell}\), one has with some \(D_3\)
\[
|x_s| \leq \exp\left(-d_L (t - s) + L D_2 \int_{s}^{t} \exp(-2\mu_L r) \, dr\right)|x_s|
\]
\[
\leq \exp(-d_L (t - s) + D_3 \sqrt{\ell} \exp(-2\mu_L s))\| x_s \|,
\]
in other words up to \(t \leq \tau\).
\[
\| \Psi(s, t) \| \leq \exp(-d_L (t - s) + D_3 \sqrt{\ell} \exp(-2\mu_L s)) =: \psi(s, t).
\]
Let \(s_0\) be a constant larger than \(\limsup 2D_3 \sqrt{\ell}/d_L < \infty\), then if \(s > \frac{1}{2} s_0\),
\[
d_L s \geq D_3 \sqrt{\ell}, \quad \sqrt{\ell} \exp(-2\mu_L s) \leq 1.
\]
So when \( t > s_0 \), the second probability in (4.15) is bounded by an indicator
\[
\mathbb{P}_{\tilde{u}_{>0}}(\Psi(0, t)\tilde{n}_0 \geq \epsilon, t \leq t) \leq 1_{\exp(-D_1t)}(\geq t),
\]
which goes to 0 as \( L \to \infty \). For the third probability in (4.15), observe that with integration by parts
\[
\int_0^t \Psi(s, t) \Sigma_d(B_s - B_t) = \Psi(0, t)\Sigma_dB_t + \int_0^t \Psi(s, t) D_s \Sigma_d[B_s - B_t]ds.
\]
It suffices for us to show \( \int_0^t \Psi(s, t) \|D_s\| \|B_s - B_t\|ds \to 0 \) when \( t \leq t \). This computational verification will be done in lemma A.6, part (i) using
\[
\|D_s\| = \|\Gamma + \sigma_x^{-2} R, P(t)\|
\]
\[
\leq \|\Gamma\| + \sigma_x^{-2} \|R, L\| + \bar{P}_t
\]
\[
\leq \|\Gamma\| + \sigma_x^{-2} (\|\bar{R}, t\| + \|R, L\|)(\|L\| + c_0)
\]
\[
= AL \exp(-\muLt) + \sqrt{L}B
\]
for some constants \( A, B \).

To bound the last probability in (4.15), we define \( y_t := \sigma_x^{-1} \int_0^t \Psi(s, t)\sqrt{\bar{L}}R_sdw_s \). Due to Chebyshev’s inequality, it suffices for us to show \( \mathbb{E}_{\tilde{u}_{>0}}(|y_t|^2 1_{\tau \leq t}) \to 0 \). Notice that \( y_t \) follows:
\[
dy_t = -\left(\Gamma + \sigma_x^{-2} R, P(t)\right)y_t dt + \sigma_x^{-1} \sqrt{\bar{L}}R_sdw_s.
\]
Following the derivation of (4.15), we have the same bound for \( \text{Re}(y_t^\tau \left(\Gamma + \sigma_x^{-2} R, P(t)\right)y_t) \) when \( \tau \geq t \); applying it and the bound for the quadratic variation in (4.13) in the following application of the Itô formula:
\[
d|y_t|^2 = -2\text{Re}(y_t^\tau \left(\Gamma + \sigma_x^{-2} R, P(t)\right)y_t dt + R_s dw_s^t, dw_s^s)R_s + 2y_t^\tau R_s dw_s.
\]
\[
\leq -2[d_L - LD_3 \exp(-2\muLt)]|y_t|^2 dt + |K||R_s|^2 dt + 2\text{Re}(y_t^\tau R_s dw_s).
\]
If we let \( z_t \) be the solution with \( z_0 = 0 \) and
\[
dz_t = -2[d_L - LD_3 \exp(-2\muLt)]z_t dt + |K||R_s|^2 dt + 2\text{Re}(y_t^\tau R_s dw_s).
\]
This is well defined as one can easily verify that \( y_t \) is a \( L^2 \)-integrable process, and by lemma A.4 \( R_s \) is uniformly bounded from above. Then by the comparison principle, \( z_t \geq |y_t|^2 1_{\tau \leq t} \), \( \mathbb{P}_{\tilde{u}_{>0}}\text{-a.s.} \). Moreover, notice that \( w_t \) is a martingale, using
\[
z_t = \int_0^t \Psi(s, t)\|K||R_s|^2 ds + 2\int_0^t \Psi(s, t)\text{Re}(y_t^\tau R_s dw_s),
\]
we have:
\[
\mathbb{E}_{\tilde{u}_{>0}}(|y_t|^2 1_{\tau \geq t}) \leq \mathbb{E}z_t \leq |K| \int_0^t \Psi(s, t)\|R_s|^2 ds.
\]
It remains to show the quantity above converges to 0; this is rather elementary since \( \|R_s\|^2 = \|R, L + \bar{R}_t\|^2 \leq B \), then apply lemma A.6 (ii) to verify this.

4.4. Asymptotic uncertainty reduction

Proof of corollary 3.4. With clear knowledge of the asymptotic behaviour \( R_s \) and \( m_t \), the information theory part is quite elementary.
Part (i) Recall that in (2.13), the dispersion part is
\[-\frac{1}{2} \ln(\det R_t) + \frac{1}{2} \ln(\det R_{\text{att}}) - \frac{|K|}{2} + \frac{1}{2} \text{tr}(R_t R_{\text{att}}^{-1}).\]
Since the climatological covariance $R_{\text{att}}$ is a constant matrix, $\|R_t\| \to 0$ in $P_{\tilde{E}_{\tilde{G}\tilde{R}}}$ by theorem 3.3 (i), asymptotically the trace part vanishes. Applying the result of theorem 3.3 (i) to the following decomposition:
\[\ln(\det R_t) = \ln(\det(I + R_t^{-1}(R_t - R_{\text{att}}))) + \ln(\det R_{\text{att}})\]
the first term goes to 0 in $P_{\tilde{E}_{\tilde{G}\tilde{R}}}$, Hence the dispersion part asymptotically converges to
\[-\frac{1}{2} \ln(\det R_{\text{att}}) + \frac{1}{2} \ln(\det R_{\text{att}}) - \frac{|K|}{2} = \frac{1}{2} \sum_{k \in K} \left( \ln \left( \frac{1}{2} (1 + \sqrt{1 + L \sigma_k^{-2} d_k^{-2} \sigma_k^{-2} |\tilde{r}_k|^2}) - 1 \right) \right)\]
Part (ii) The signal part of the relative entropy is $\frac{1}{2} (\tilde{m}_t - \tilde{m}_{\text{att}}) R_{\text{att}}^{-1} (\tilde{m}_t - \tilde{m}_{\text{att}})$, by our results in (ii) and (iii), $\tilde{m}_t - U_t$ converges to 0 in two different ways. Hence it is straightforward to have our claim.
Part (iii) Since the mutual information is
\[I(U_t, X_s) = \frac{1}{2} \ln(\det R_t) - \frac{1}{2} \mathbb{E} \ln(\det R_t)\]
Again we decompose $\ln(\det R_t) = \ln(\det(R_t^{-1} R_t)) + \ln(\det R_{\text{att}})$. On one hand by theorem 3.3 part (i), $\ln(\det(R_t^{-1} R_t)) \to 0$ in $P$, and notice that according to lemma A.4, there is some constant $U$ so that $\ln(\det(R_t^{-1} R_t)) \leq U \ln L$ a.s. Then for any fixed $\xi > 0$,
\[\mathbb{E} \ln(\det(R_t^{-1} R_t)) \leq \xi + \mathbb{P}(\|\ln(\det(R_t^{-1} R_t)) \geq \xi\) U \ln L.\]
The right-hand side is of an order less than $\ln L$. On the other hand,\[-\frac{1}{2} \ln(\det R_{\text{att}}) \xrightarrow{L \to \infty} 1.\]
Putting the two parts together we see our claim holds. \square

Proof of corollary 3.5. For marginal processes, it suffices to redo the previous derivation but using sub-matrices. Let us add a tilde to symbols to indicate sub-vectors of indexes in $\tilde{K}$ or sub-matrices of indexes in $\tilde{K} \times \tilde{K}$. We shift the indexes so the ones in $\tilde{K}$ come ahead. As the mean and covariance of $\tilde{U}_t$ are these sub-vectors and sub-matrices, the relative entropy is
\[\mathcal{P}(\pi_t|\tilde{K}, \pi_{\text{att}}|\tilde{K}) = \frac{1}{2} \left( (\tilde{m}_t - \tilde{m}_{\text{att}})^\dagger \tilde{R}_{\text{att}}^{-1} (\tilde{m}_t - \tilde{m}_{\text{att}}) \right) + \frac{1}{2} \left( \text{tr}(\tilde{R}_{\text{att}}^{-1}) - |\tilde{K}| - \ln(\det(\tilde{R}_{\text{att}})) \right)\]
The mutual information is
\[I(\tilde{U}_t, X_s) = \frac{1}{2} \ln(\det \tilde{R}_{\text{att}}) - \frac{1}{2} \mathbb{E} \ln(\det \tilde{R}_{\text{att}})\]
The results of theorem 3.3 indicate their marginal version holds as well, because evidently $\|R_t - R_{\text{att}}\| \geq \|\tilde{R}_t - \tilde{R}_{\text{att}}\|$; as $R_{\text{att}}$ is diagonal:
\[(\tilde{m}_t - \tilde{m}_{\text{att}})^\dagger \tilde{R}_{\text{att}}^{-1} (\tilde{m}_t - \tilde{m}_{\text{att}}) = \begin{bmatrix} \tilde{m}_t - \tilde{m}_{\text{att}} \\ 0 \end{bmatrix}^\dagger R_{\text{att}}^{-1} \begin{bmatrix} \tilde{m}_t - \tilde{m}_{\text{att}} \\ 0 \end{bmatrix} \leq (\tilde{m}_t - \tilde{m}_{\text{att}})^\dagger \tilde{R}_{\text{att}}^{-1} (\tilde{m}_t - \tilde{m}_{\text{att}});\]
if \( \tilde{R}_t \) has uni-norm eigenvector \( \tilde{u} \) for its least eigenvalue, then
\[
\| \tilde{R}_t^{-1} \| = \| \tilde{R}_t \tilde{u} \|^{-1} = \left| \begin{array}{c} \tilde{u} \\ 0 \end{array} \right|^{-1} \leq \inf_{|u| \leq 1} \| R_t u \|^{-1} = \| R_t^{-1} \|.
\]

Using this marginal version of the results, the proof of this corollary is identical to the one of corollary 3.4. \( \square \)

4.5. Comparison with regularly spaced Eulerian observations

It is interesting to consider the counterpart of theorem 3.3 and Corollary 3.4–3.5 when the observations are Eulerian, i.e. we have observation of the velocity field from \( L \) different fixed points instead of \( L \) tracers. Such problems are considered thoroughly in [15]. Assuming the velocity field is generated the same way as in (2.2), then the observation of \( L \) points instead of \( L \) tracers is simply \( P_X(\tilde{z}_t)U_t \) plus a white observation noise \( \sigma_x dW_t^l(s) \). Here we use \( Z = (\tilde{z}_1, \ldots, \tilde{z}_L) \) to represent the locations where the observations are made. One natural and simple choice of the observation sites is the regular spaced lattice points on \( \mathbb{T}^n \) and \( L = n^d \) for some \( n \), i.e.
\[
\{\tilde{z}_1, \ldots, \tilde{z}_{n^d}\} = \left\{ \frac{2\pi j}{n} : j \in \{0, \ldots, n - 1\}^d \right\}.
\]

In this special case, all the claims of theorem 3.3 and corollary 3.4–3.5 hold essentially the same. To see this, it is sufficient to notice that
\[
(P_X^s(Z)P_X(Z)_{j,k}) = \sum_{l=1}^{n^d} \exp(i(\tilde{k} - \tilde{j}) \cdot \tilde{z}_l(t)) (\tilde{g}_j^l \tilde{g}_k^l) = n^d 1_{j=k} |\tilde{g}_j|^2,
\]
using the identity \( \sum_{l=0}^{n^d-1} \exp(2i\pi j/n) = 0 \). In other words, \( P_X^s(Z)P_X(Z) = n^d M = LM \), where \( M \) is derived in lemma 4.2. Hence (4.17) is simply the mean field limit of (2.9). So following a simpler version of the same proof, the claims of theorem 3.3 and corollary 3.4–3.5 hold without change. If one changes the configuration \( \{\tilde{z}_1, \ldots, \tilde{z}_L\} \), \( P_X^s(Z)P_X(Z) \) will be a different matrix, which induces a different limit for \( R_t \) and hence also for other variables. To examine these general problems, we must have a detailed description of how the observation sites are distributed as \( L \) increases. Since this is not closely related to the main focus of this paper, we do not elaborate further in this direction.
5. Numerical experiments

In this section we numerically validate the theoretical results in section 3. We consider a $d = 2$ dimensional set-up, like (2.1) with wavenumbers in $[-2, 2]^2$, the total dimension of the signal process $U$, is $|K| = 26$.

For system parameters in (1.2), all the damping $d_k$ are set as 0.05, which corresponds to a moderately long decorrelation time $\tau_{corr} = 20$ units as occurs in many geophysical flows. The system is assumed to have no deterministic forcing for simplicity and therefore the mean at the equilibrium is zero. Equipartition energy is assumed with all $\sigma_k^2 = 0.3$ to mimic turbulence at large scale [15]. The observation noise of the tracers is set to be $\sigma_v = 0.5$ as the observation noise level is roughly half of the system noise of each mode. In the following, we study the filtering skill with different numbers of tracers $L$, ranging from 2 to 500.

The streamlines in the physical space recovered by the filter are illustrated in figure 1. The top row displays streamlines associated with the true velocity field at $T = 5$, 15 and 25, followed by the three rows showing the recovered flow field using $L = 2$, 10 and 50 tracers. In all the panels, the solid and dashed curves represent the counterclockwise and clockwise motion of the flows, respectively. Clearly, the flow field with only two tracers succeeds in recovering the general profile of the flow field. Yet, a low recovering skill is observed in certain areas, such as the flipped sign in the left top part at $T = 5$. As expected, the recovered flow field approaches the true one with the increase of $L$. However, despite the significant error reduction in the recovered flow with $L = 50$, some visible difference still exists. Even with $L$ increasing to 500 as shown in figure 1, the recovered streamlines are not perfectly matching the truth, which implies the potential information barriers.

On the other hand, figure 2 explicitly demonstrates the information barrier phenomenon, which is mathematically described by corollary 3.4 and 3.5. The reduced uncertainty in the signal part converges to its limit value as $L$ increases. As for the dispersion part, although it has a sustained growth, the growth is only at a rate of $\frac{1}{4} |K| \ln L$. With $|K| = 26$, reducing every 15 nats of uncertainty from the dispersion part requires a tenfold increase of tracers. This is illustrated in panel (h) of figure 2. This uncertainty reduction approaches the asymptotic value (3.3) from corollary 3.4, as indicated by the overlapped solid and dashed lines. In addition, the length of the transient period for the dispersion part to arrive at the equilibrium value goes down with the increase in $L$.

6. Conclusion and discussion

In this paper, we set up an optimal filter of a random incompressible flow using observations from Lagrangian tracers. The derivation of this filter is based on the conditional Gaussian theory in [16]. Despite the nonlinearity of the tracers, the posterior distribution is Gaussian with mean and covariance following elementary differential equations with random coefficients. Such exploitation of Gaussian structure has been proved to be very fruitful in many areas of data assimilation for random processes [35], and may bring further insight in building other simple filters for nonlinear systems [15].

We prove the geometric ergodicity of the tracers’ projection on $\mathbb{T}^d$. We then analyse the asymptotical behaviour of the filter assuming each tracer is distributed uniformly on $\mathbb{T}^d$ when $L \to \infty$. We show the posterior covariance will approach $R_\epsilon$, a deterministic diagonal matrix, and the posterior mean will converge to the signal process.

With clear asymptotic characterization of the filter, we can characterize the uncertainty reduction by the observations from $L$ tracers as $L \to \infty$. In terms of both relative entropy and mutual information, the uncertainty reduction grows asymptotically as $\frac{1}{4} |K| \ln L$. 

2153
**Figure 1.** Streamlines recovered with different numbers of tracers \( L \). The first row shows the streamlines of the true realization. The second, third and fourth rows show the recovered streamlines using \( L = 2, 10 \) and 50 tracers, respectively. The three columns show the comparison of streamlines at time \( T = 5 \), 15 and 25. Streamlines with positive and negative values represent the counterclockwise and clockwise direction of the velocity field. The circles represent the location of the tracers at the current time.

In other words, in order to gain the same amount of additional information, one would need exponentially more tracers. This certainly is an information barrier in practice when only a reasonable amount of tracers are available. As seen in the numerical experiments (section 5), 50 tracers produce roughly the same result as with 500 tracers. It is worth noting that [6] has a similar claim, as their result indicates that 36 tracers without any analysis, instead of 64, is optimal under their setting.

Apart from the direct conclusions listed above, we also have the following remarks on our results:

(i) The requirement that the flow is incompressible is not necessary for the derivation of the filter. Indeed, incompressibility does not show up in the filter set-up part (section 2.2); it is only required in theorem 3.1 to make the uniform distribution invariant. It is possible to generalize the ideas in this paper to flows that are not incompressible, e.g. a shallow water model. This lies beyond the scope of this paper, and will be addressed by a follow-up paper of the authors.

2154
Figure 2. Uncertainty reduction with different numbers of tracers $L$. Panels (a)–(c) show the uncertainty reduction in the signal part as a function of time (thick solid line) compared with its limit (thick dashed line) for each fixed $L$. In panel (d), the solid line with circles shows the averaged value of the absolute error in the signal part compared with its limit across time interval $T \in [0, 50]$ as a function of $L$. Panels (e)–(g) show the uncertainty reduction in the dispersion part as a function of time for each fixed $L$. In panel (h), the solid line with circles shows the averaged dispersion part across time interval $T \in [5, 50]$ as a function of $L$ and the dotted line represents the function $\frac{1}{4}|K| \ln L$. The magnifying blocks in panels (f) and (g) show the evolution in the initial one time unit.

(ii) The conditional Gaussian filtering framework will remain valid if the OU processes $\tilde{v}_k(s)$ are correlated; in other words, $\Sigma_u$ in (2.3) is replaced with a general matrix. The geometric ergodicity of the noisy tracers will also remain valid since theorem 3.1 holds as long as the underlying velocity field is incompressible and ergodic. The main reason we are not considering such a general case is that the limit $R_L$ in theorem 3.3 would be too notationally heavy to write down explicitly.

(iii) It is worth mentioning that all theoretical results, theorem 3.1, 3.3 and corollary 3.4, are stated in the conditional distribution $\mathbb{P}_{\tilde{w}_{\leq t}}$. Generally speaking, proving results in $\mathbb{P}_{\tilde{w}_{> t}}$ for
a.s. realization of $\tilde{v}_{s} \geq 0$ is more difficult than proving the same results in the super-ensemble probability $\tilde{P}$. However, in actual practical applications, often one is only observing noisy tracers in a single realization of a random flow field, so we emphasize this here. Such a.s. results in the random flow field also allow us to study the filter behaviour for rare extreme realization of the flow field.

(iv) As discussed in section 3, the configuration of the tracers’ locations impacts the performance of the filter. Non-ideal situations, e.g. clustering all tracers initially, are ruled out by the assumption 3.2. As shown in [5, 7, 8], different starting locations of tracers and their trajectories result in very different filter results at least in short periods of time. However, as the second part of this paper addresses mainly the asymptotic behaviour of the tracers, it is preferable to work under assumption 3.2.

(v) Although theorem 3.1 indicates the distribution of tracers will converge to the equilibrium in a finite amount of time, the length of this transient regime could be decades if all the tracers are disposed initially at the same location and $\sigma_x$ is insignificant compared to the scale of $\vec{v}$. In practice, in order for the filter to behave as theorem 3.3 describes, the initial location of tracers should be chosen uniformly in the area.

(vi) Theorem 3.3 part (i) indicates the posterior covariance is close to a deterministic matrix $R_L$ when $L$ is large. This fact would be used to deploy an imperfect filter with a great increase of efficiency: we can simply set $R_t$ to be $R_L$ after a period of time. This resolves a large computational burden when $|K|$ and $L$ are large, because updating the Riccati equation (2.9) is very stiff while $R_L$ is diagonal. Similar efficient cheap diagonal filters have been developed in the context of turbulent dynamical systems [15].

Acknowledgments

We thank the anonymous reviewers for pointing out the connection between the current paper and various existing results, and their helpful suggestions on the arrangement of the paper. This research is supported by the MURI award grant N-000-1412-10912, where AJM is the principal investigator, while NC and XTT are supported as graduate and postdoctoral fellows, respectively.

Appendix

Appendix A.1. A detailed formulation of the random incompressible flow model

We consider a random periodic incompressible $d$-dimensional flow as a superposition of random plane waves, formulated by (1.1):

$$\tilde{v}(\vec{x}, s) = \sum_{k \in K} \tilde{v}_k(s) \exp(ik \cdot \vec{x}) \tilde{r}_k,$$

where $K$ is some finite set and $k$ is associated with a wavenumber $\vec{k}$. The incompressible condition for a constant density flow gives $\nabla \cdot \vec{v} = 0$, which implies $k \cdot \exp(ik \cdot \vec{x}) \tilde{r}_k = 0$ for each Fourier mode $\vec{k}$. Therefore, the eigenvector $\tilde{r}_k$ is perpendicular to the wavenumber $\vec{k}$. There can be multiple $k$s sharing the same wavenumber $\vec{k}$, so we require their $\tilde{r}_k$ to be orthogonal to each other so that the contribution of each $\tilde{v}_k$ to $\vec{v}_s$ is identifiable. We model $\{\tilde{v}_k(s)\}_{k \in K}$ as a group of independent OU processes with damping $d_k > 0$:

$$d\tilde{v}_k(s) = -d_k \tilde{v}_k(s) ds + f_k(s) ds + \sigma_k dW_k^\nu(s).$$

When the wavenumbers $\vec{k} = 0$, the corresponding modes represent a random background mean sweep, with the corresponding $\tilde{v}_k(s)$, $\tilde{r}_k$ being real-valued.
For nonzero wavenumbers, in order to keep $\tilde{v}(x, s)$ being real-valued, we assume the components of (1.1) come with their conjugate pairs. We denote $-k$ as the conjugate index of $k$ with the corresponding wavenumber being $-\tilde{k}, k_{-k}, f_{-k}(s), W_{-k}(s)$ being conjugates for the ones of index $k$ and $d_k = d_{-k}, \sigma_k = \sigma_{-k}$. One way to ensure this is formulating the complex Wiener processes as follows:

$$W_k^u(s) = \frac{1}{\sqrt{2}}(B_k(s) + iB_{-k}(s)), \quad W_{-k}(s) = \frac{1}{\sqrt{2}}(B_k(s) - iB_{-k}(s)),$$

where $\{B_k(s)\}$ are independent one-dimensional real-valued Wiener processes. It is usually convenient to discuss these modes in terms of conjugate pairs. This is the reason why the matrix $\Sigma_u$ in (2.2) holds a particular form as illustrated after (2.2).

Filtering formula for conditional Gaussian process

The following is a simplified version of theorem 12.7 [16] in our context:

**Theorem A.1.** Let $(u_t, x_t)$ be a continuous diffusion-type process with

$$du_t = [-\Gamma_t u_t + F_t]dt + \Sigma_u dW_t(t),$$

$$dx_t = [P_t u_t + Q_t]dt + \Sigma_x dW_x(t),$$

with bounded $\Gamma_t, F_t, P_t, Q_t$ processes being functions of $x_t$. If $\mathbb{P}(u_0 \in \cdot | x_0) = \mathcal{N}(m_0, R_0)$, then conditioned on $x_{t \leq s}$, $\mathbb{P}(u_t \in \cdot | x_{t \leq s})$ is $\mathcal{N}(m_t, R_t)$, with $m_t, R_t$ being solutions to the following with the initial value $m_0, R_0$:

$$dm_t = [-\Gamma_t m_t + F_t]dt + R_t \Sigma_x^* \Sigma_x^{-1} [dx_t - (P_t m_t + Q_t) dt],$$

$$dR_t = [-\Gamma_t R_t - R_t \Sigma_x^* \Sigma_x^{-1} + R_t \Sigma_x^* \Sigma_x^{-1} - R_t \Sigma_x^* \Sigma_x^{-1} P_t R_t]dt.$$

**A.2. Information theory formulae for complex Gaussian random vectors**

We say a complex-valued random vector $X + Yi$ is Gaussian distributed if $(X, Y)$ as a real-valued vector is Gaussian distributed. The relative entropy of $X + Yi$ between probability density $p$ and $q$ can be naturally defined as the relative entropy of $(X, Y)$ under these two probabilities.

One caveat in dealing with the complex random vector $X + Yi$ is that its covariance matrix does not contain sufficient second moments to induce the density of $(X, Y)$. For example, $X$ and $\frac{1}{\sqrt{2}}(X + Yi)$ have the same mean and covariance if $X, Y$ are independent $\mathcal{N}(0, 1)$ random variables. Of course one can look for the covariance matrix for the joint vector $(X, Y)$, but in many cases that will be very inconvenient. For instance, in our filtering model, we will need additional formulas other than (2.10) and (2.9).

The common practice in the engineering community, e.g. [36], to overcome this setback is to consider the pair $(X + Yi, X - Yi)$, since one can show its covariance matrix is sufficient to induce the law of $(X, Y)$. Moreover, if one has the $(X + Yi, X - Yi) \sim \mathcal{N}(\bar{m}_p, R_p)$ and $\mathcal{N}(\bar{m}_q, R_q)$, the relative entropy of $(X, Y)$ under two measures can also be computed by (2.12).

As the authors cannot find a reference clearly stating this fact, we prove it using elementar algebra.

**Lemma A.2.** Suppose under measures $\mathbb{P}$ and $\mathbb{Q}$, $X, Y, Z$ are real-valued joint Gaussian vectors with $Y$ and $Z$ being of the same dimension. Suppose also that $[X, Y + Zi, Y - Zi]$ has mean $m_p, m_q$ and covariance $R_p, R_q$ under each measure respectively, then the relative entropy
between \( \mathbb{P} \) and \( \mathbb{Q} \) of \([X, Y, Z]\) is given by
\[
\mathcal{P}(\mathbb{P}, \mathbb{Q}) = \left[ \frac{1}{2} (\bar{m}_p - \bar{m}_q)^\top R_q^{-1} (\bar{m}_p - \bar{m}_q) \right] + \frac{1}{2} \left[ \text{tr}(R_p R_q^{-1}) - N - \ln \det(R_p R_q^{-1}) \right],
\]
where \( N = \dim X + \dim Y + \dim Z \).

**Proof.** Simply note that
\[
\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & \frac{1}{2} I & \frac{1}{2} I \\ 0 & -\frac{1}{2} I & \frac{1}{2} I \end{bmatrix} \begin{bmatrix} X \\ Y + Z_1 \\ Y - Z_1 \end{bmatrix}.
\]
Denote \( M \) as the matrix on the right-hand side. Then \([X, Y, Z]\) are distributed as \( \mathcal{N}(M\bar{m}_p, MR_pM^\top), \mathcal{N}(M\bar{m}_q, MR_qM^\top) \) respectively under \( \mathbb{P}, \mathbb{Q} \). So applying the formula for the real-valued Gaussian vector \([X, Y, Z]\) and cancelling out the appearance of \( M \), one can show our claim elementarily.

As an application to our filtering purpose, \( X \) consists of real modes. We pick one complex mode out of each conjugate pair, and use their real parts to form \( Y \) and the imaginary part to form \( Z \). The conjugate counterparts of these modes are naturally represented by \( Y - Z_1 \).

This is the first reason why we prefer to talk of the complex modes in conjugate pairs like the statement of corollary 3.4: their joint mean and covariance is necessary and sufficient to induce the law of them. On the other hand, in the sense of uncertainty reduction, the information regarding \( \hat{v}_k \) or \( \hat{v}_{-k} \) or both of them as a pair, are the same: \( \hat{v}_k \)’s value determines \( \hat{v}_{-k} \) and vice versa. So it is more reasonable to think of complex modes in pairs.

**Lemma A.3.** Under the set-up of section 2, the mutual information is given by
\[
I(U_t, X_{s \leq t}) = \mathbb{E}(\mathcal{P}(\pi_{t|t}, \pi_{att})) = \frac{1}{2} \ln \det R_{att} - \frac{1}{2} \mathbb{E} \ln \det R_t.
\]

**Proof.** The first line comes from an equivalent definition of mutual information, see [23] and [37] equation (3.2). That is, if we write the law of \( X \) conditioned on \( Y = y \) as \( p_{X|Y=y} \), then by disintegration \( p(dx, dy) = p_{X|Y=y}(dx) p_Y(dy) \), hence
\[
I(U_t, X_{s \leq t}) = \int p(dx, dy) \ln \frac{p_X(x|y) p_Y(dy)}{p_X(x) p_Y(dy)}
= \int p_Y(dy) \int p_X(x|y) p_X(x) \ln p_X(x|y)
= \int p_Y(dy) \mathbb{E}(p_{X|Y=y}) p_X(dx).
\]

Let \( X = U_t, Y = X_{s \leq t} \), notice \( \pi_{t|t} \) is the law of \( U_t \) conditioned on \( X_{s \leq t} \), and \( U_t \sim \pi_{att} \) under \( \mathbb{P} \), we have \( I(U_t, X_{s \leq t}) = \mathbb{E}(\mathcal{P}(\pi_{t|t}, \pi_{att})) \). Next, plug in the formula (2.13):
\[
I(U_t, X_{s \leq t}) = \frac{1}{2} \ln \det R_{att} - \frac{1}{2} \mathbb{E} \ln \det R_t + \frac{1}{2} \text{tr}(\mathbb{E} R_t R_{att}^{-1}) - \frac{|K|}{2}
+ \frac{1}{2} \text{tr}(\mathbb{E} (\bar{m}_t - \bar{m}_{att}^t)(\bar{m}_t - \bar{m}_{att}^t)^*) R_{att}^{-1}.
\]
Using the fact that \( \bar{m}_t \) is the posterior mean, by \( \mathbb{E}_{X_{s \leq t}} U_t = \bar{m}_t \) and \( \bar{m}_t, \bar{m}_{att}^t \) is \( \sigma(X_{s \leq t}) \)-adapted:
\[
\mathbb{E}_{X_{s \leq t}} (\bar{m}_t - U_t)(\bar{m}_t - \bar{m}_{att}^t)^* = [\mathbb{E}_{X_{s \leq t}} (\bar{m}_t - U_t)] (\bar{m}_t - \bar{m}_{att}^t)^* = 0.
\]
Hence
\[ R_{\text{all}} = \mathbb{E}(U_t - \tilde{m}_t^{\text{all}})(U_t - \tilde{m}_t^{\text{all}})^* \]
\[ = \mathbb{E}(U_t - \tilde{m}_t + \tilde{m}_t - \tilde{m}_t^{\text{all}})(U_t - \tilde{m}_t + \tilde{m}_t - \tilde{m}_t^{\text{all}})^* \]
\[ = \mathbb{E}(U_t - \tilde{m}_t)(U_t - \tilde{m}_t)^* + \mathbb{E}(\tilde{m}_t - \tilde{m}_t)(\tilde{m}_t - \tilde{m}_t)^* \]
\[ = ER_t + E(\tilde{m}_t - \tilde{m}_t)(\tilde{m}_t - \tilde{m}_t)^*. \]

In other words, \( \mathbb{E}(\tilde{m}_t - \tilde{m}_t)(\tilde{m}_t - \tilde{m}_t)^* = R_{\text{all}} = \mathbb{E}R_t \), so
\[ \text{tr}\left( (\mathbb{E}(\tilde{m}_t - \tilde{m}_t^{\text{all}})(\tilde{m}_t - \tilde{m}_t^{\text{all}})^*)R_t^{-1}\right) = \text{tr}(I_K) - \text{tr}(\mathbb{E}R_t R_t^{-1}). \]

Putting this back to the equation of \( I(U_t, X_{t < \cdot}) \), we have shown the identity. □

### A.3. Verifications of various claims

**Lemma A.4.** There exist two diagonal matrices \( R_L^-, R^+, R_L^- \geq mL^{-1/2}I \) for some constant \( m > 0 \), such that if \( R_L^- \leq R_0 \leq M I, M \geq \|R^+\| \) then the solution to (2.9) is uniformly bounded in time by \( R_L^- \leq R_t \leq M I \) a.s. Hence \( E|\tilde{m}_t|^2 < \infty. \)

**Proof.** Consider the following deterministic processes. Let \( R_{s,t} \) be the solution to the following:
\[ \dot{R}_{s,t} = -\Gamma R_{s,t} - R_{s,t} \Gamma^* + \Sigma, \quad R_{s,0} = MI \]
and let \( R^* \) be the stable point of the dynamics above. Therefore if \( R_{s,0} = MI \geq R_s, R_{s,t} \) will be decreasing in time, so \( R_{s,t} \leq M I \).

Next, let \( R_{-,t} = R_L \) be the stable point of the following
\[ \dot{R}_{-,t} = -\Gamma R_{-,t} - R_{-,t} \Gamma^* + \Sigma - \sigma\bar{\epsilon}^{-2}L |K| R_{-,t}^2. \]

Since this is a diagonalizable matrix equation, one can easily verify that each entry of \( R_L \) is of the order \( \sqrt{L}^{-1/2} \).

Using the comparison theorem of the differential Riccati equation, theorem 4.1.4 of [33], because \( 0 \leq P(t) \leq |K|I \), we have \( R_{-,t} \leq R(t) \leq R_{s,t} \) a.s. This concludes our claim of \( R_t \).

Then note that by Young’s inequality
\[ E|\tilde{m}_t|^2 \leq 2E|\tilde{m}_t - U_t|^2 + 2E|U_t|^2 = 2E\text{tr}(R_t) + 2\text{tr}(R_{\text{all}}) < \infty. \]

□

**Lemma A.5.** For any fixed \( j \neq k, t > 0, q > \frac{1}{2} \), \( z^L_t := L^{-q}(|\tilde{P}_t|) \) is a family of processes that converges to 0 in the uniform norm of \( C([0, t]) \) under \( P_{\tilde{v}, \tilde{z}} \) when \( L \to \infty. \)

**Proof.** Recall that if \( \tilde{j} = \tilde{k} \), then \( z^L_t \equiv 0 \), this Lemma trivially holds. So we assume \( \tilde{j} \neq \tilde{k} \) in the following.

By theorem 7.5 of [32], it suffices for us to check \( z^L_t \to 0 \) in \( P_{\tilde{v}, \tilde{z}} \) for each \( x \), and also show the modulus of continuity condition for the family \( \{z^L_t\} \).

Fix \( s \), as \( \{Y_t = \exp(i(\tilde{k} - \tilde{j}) \cdot \tilde{X}_t(s))\} \) is a sequence of i.i.d. random complex variables in \( P_{\tilde{v}, \tilde{z}} \) with
\[ \mathbb{E}_{\tilde{v}, \tilde{z}} Y_t = 0, \quad \mathbb{V}ar_{\tilde{v}, \tilde{z}} Y_t = 1, \quad \mathbb{E}_{\tilde{v}, \tilde{z}} Y_t^2 = \frac{1}{|T|^2} \int_T \exp(2i(\tilde{j} - \tilde{k}) \cdot x) \, dx = 0. \]
By the central limit theorem for complex random variables,

\[ \frac{1}{\sqrt{L}} [\bar{P}_{t,k} - \bar{P}_{t,k}^0] = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} Y_i \rightarrow_N (0, 1, 0) \quad \text{for a.s. } \bar{v}_{t,k}. \]

where \( X + Y_i \sim N(0, 1, 0) \) if \( X \perp Y, X, Y \sim N(0, \frac{1}{2}) \) (see [36] for more details of complex Gaussian distribution). So for any \( q > 1/2, L^{-q}[\bar{P}_{t,k} - \bar{P}_{t,k}^0] \rightarrow 0 \) in \( \mathbb{P}_{\bar{v}_{t,k}} \) for any fixed \( s \leq t \).

Next we check the modulus continuity condition for the processes \([z^L_r]_s\). By the corollary following theorem 7.4 of [32], it suffices for us to show that for any \( \epsilon, \eta > 0 \), there are \( \delta > 0 \) and \( L_0 \) such that for any \( 0 < r < t - \delta \), and \( L \geq L_0 \)

\[ \frac{1}{\delta} \mathbb{P}_{\bar{v}_{t,k}} \left( \sup_{r \leq s \leq r + \delta} \left| z_r^L - z_r^L \right| \geq 2\epsilon \right) \leq \eta. \]

Applying Itô’s formula we have

\[ z_r^L - z_r^L = \int_r^s L^{-q} \sum_{l=1}^{L} \exp(i(\bar{k} - \bar{j}) \cdot \bar{X}_l(u)) |i(\bar{k} - \bar{j}) \cdot \bar{v}(\bar{X}_l(u), u) + \sigma_r^2 |\bar{k} - \bar{j}|^2 | \, du \]

\[ + \sigma_r \int_r^s L^{-q} \sum_{l=1}^{L} \exp(i(\bar{k} - \bar{j}) \cdot \bar{X}_l(u)) i(\bar{k} - \bar{j}) \cdot dW^q_l(u) \]

\[ = \int_r^s L^{-q} \sum_{l=1}^{L} [f_u(\bar{X}_l(u)) du + g_u(\bar{X}_l(u)) dW^q_l(u)] \]

with \( f_u, g_u \) representing the corresponding functions. So it suffices to show

\[ \frac{1}{\delta} \mathbb{P}_{\bar{v}_{t,k}} \left( \sup_{r \leq s \leq r + \delta} \left| \int_r^s L^{-q} \sum_{l=1}^{L} g_u(\bar{X}_l(u)) dW^q_l(u) \right| \leq \frac{1}{\delta} \right) \leq \eta. \]

Applying Doob’s inequality and Itô’s isometry to the martingale part of (A.1):

\[ \mathbb{E}_{\bar{v}_{t,k}} \left[ \left| \int_r^s L^{-q} \sum_{l=1}^{L} g_u(\bar{X}_l(u)) dW^q_l(u) \right|^2 \right] \leq \frac{L^{-2q}}{\delta^2} \mathbb{E}_{\bar{v}_{t,k}} \int_r^s \sum_{l=1}^{L} |g_u(\bar{X}_l(u))|^2 du \leq \frac{L^{-2q}}{\delta^2} \mathbb{E}_{\bar{v}_{t,k}} \int_r^s \sum_{l=1}^{L} |\bar{X}_l(u)|^2 du \leq \frac{L^{-1-2q}}{\delta^2} \mathbb{E}_{\bar{v}_{t,k}} \left| \bar{X}_l(u) \right|^2. \]

We can find \( L_0 \), so the right-hand side is sufficiently small when \( L \geq L_0 \).

Applying the Cauchy–Schwarz inequality to the Riemann integral part in (A.1),

\[ \sup_{r \leq s \leq r + \delta} \left| \int_r^s L^{-q} \sum_{l=1}^{L} f_u(\bar{X}_l(u)) du \right|^2 \leq \frac{L^{-2q}}{\delta^2} \int_r^s \sum_{l=1}^{L} f_u(\bar{X}_l(u))^2 \, du. \]

Using Chebyshev’s inequality,

\[ \mathbb{P}_{\bar{v}_{t,k}} \left( \sup_{r \leq s \leq r + \delta} \left| \int_r^s L^{-q} \sum_{l=1}^{L} f_u(\bar{X}_l(u)) du \right| \geq \epsilon \right) \leq \frac{\delta}{\epsilon^2} \frac{L^{-2q}}{\delta^2} \int_r^s \sum_{l=1}^{L} \mathbb{E}_{\bar{v}_{t,k}} |f_u(\bar{X}_l(u))|^2 \, du. \]
Next, note that
\[
E_{\xi>0}f_{u}(\tilde{X}_{t}(u)) = \frac{1}{|\mathbb{H}|} \int_{\mathbb{H}} \exp(i(k - \tilde{j}) \cdot x) |i(k - \tilde{j}) \cdot \tilde{v}(x, u) + \sigma_{i}^{2}|k - \tilde{j}|^{2}|dx = 1 \sum_{\tilde{n} = \tilde{j} - \tilde{k}} \tilde{n} \cdot \tilde{v}_{n}(u) = 0;
\]
the second equality invokes the decomposition (1.1) where all modes with wavenumbers not being \(\tilde{j} - \tilde{k}\) vanish in the periodic integral; the third equality holds because \((k - \tilde{j}) \cdot \tilde{r}_{n} = 0\) by definition of \(\tilde{r}_{n}\).

Since \(\tilde{X}_{t}(u)\) are conditionally independent,
\[
E_{\xi>0} \left| \sum_{l=1}^{L} f_{u}(\tilde{X}_{t}(u)) \right|^{2} = L E_{\xi>0} |f_{u}(\tilde{X}_{t}(u))|^{2} \leq LC(|v_{u}|^{2} + 1).
\]
Using this inequality, the right-hand side of (A.2) is bounded by
\[
\frac{1}{\varepsilon^{2}} e^{L} \left( \delta + \int_{t}^{r_{fix}} |v_{u}|^{2} \, du \right)
\]
which goes to 0 for large \(L\) and small \(\delta\). Hence we complete the verification of the continuous modulus condition. \(\square\)

**Lemma A.6.** If \(|\psi(s, t)| \leq C \exp(-(d_{L}(t-s) + D \sqrt{L} \exp(-\mu L s)))\), suppose \(s_{0}\) is a constant such that \(d_{L}(t-s) \geq D \sqrt{L} \exp(-\mu L s) \leq 1\) for \(s \geq s_{0}\) and \(L\) sufficiently large, then for any \(t > s_{0}\),

(i) With any fixed continuous process \(v_{s}\), if \(|v_{s}| = 0\ and \(|d(s)| \leq AL \exp(-\mu L s) + \sqrt{L} B\), then as \(L \to \infty\),
\[
\int_{0}^{t} \psi(s, t) d(s) v_{s} \, ds \to 0.
\]
(ii) With \(|d(s)| \leq B, \int_{0}^{t} \psi(s, t) d(s) \, ds \to 0\ as \(L \to \infty\).

**Proof.** For the claim (i), note
\[
AL \int_{0}^{t} \psi(s, t) \exp(-\mu L s) v_{s} \, ds \leq ACL \sup_{s \leq t} |v_{s}| t \exp(-d_{L}(t + D \sqrt{L})).
\]
If \(t \geq s_{0}\), when \(L \to \infty\), the right-hand side tends to 0.

And for \(\sqrt{L} \int_{j}^{t} \psi(s, t) v_{s}\), because \(v_{s} = 0\ and \(\tilde{v}_{s}\) is continuous, for any fixed \(\xi > 0\), there exists a constant \(\delta > 0\ such that \(|v_{s}| \leq \xi for s \geq t - \delta\). Hence we can break the integration into three parts and bound each using \(d_{L}(s) \geq D \sqrt{L} \exp(-\mu L s) \leq 1, \forall s \geq s_{0}\):
\[
\sqrt{L} \int_{j}^{t} \psi(s, t) v_{s} \, ds \leq C \sqrt{L} \xi \int_{t-\delta}^{t} \exp(-d_{L}(t-s) + 1) \, ds \leq C d_{L}^{-1} \sqrt{L} \xi.
\]
\[
\sqrt{L} \int_{s_{0}}^{t} \psi(s, t) v_{s} \, ds \leq \sqrt{L} \sup_{s \leq t} |v_{s}| \int_{s_{0}}^{t} \exp(-d_{L}(t-s) + 1) \, ds \\
\leq ed_{L}^{-1} \sqrt{L} \sup_{s \leq t} |v_{s}| \exp(-d_{L} \delta)
\]
\[ \sqrt{L} \int_0^{\xi_0} \psi(s,t)v_s \, ds \leq \sqrt{L} \sup_{s \leq t} |v_s| \int_0^{\xi_0} \exp(-d_L(t-s)) \, ds \]
\[ \leq \sqrt{L} d_{t_0}^{L-1} \sup_{s \leq t} |v_s| \exp(-d_L(t-s_0)) \, ds. \]

Recall that \( d_L \) is of the order \( \sqrt{L} \), the last two terms go to 0 when \( L \to \infty \) if \( t > s_0 \). The first is of the order \( \xi \). Since \( \xi \) is an arbitrary number, our claim is true.

For (ii), just note
\[ \int_0^t \psi(s,t) \, ds = \int_0^{\xi_0} \psi(s,t) \, ds + \int_{\xi_0}^t \psi(s,t) \, ds \leq e d_{t_0}^{L-1} + \sqrt{L} d_{t_0}^{L-1} \exp(-d_L(t-s_0)), \]
the right-hand side goes to 0 as \( L \) increases. \( \Box \)

References

[26] Lindvall T 2002 Lectures on the Coupling Methods (New York: Dover)
[27] Chirikjian G S 2009 Stochastic Models, information theory, and lie groups (Applied and Numerical Harmonic Analysis vol 1 (Berlin: Springer)
[28] Karatzas I and Shreve S 1991 Brownian Motion and Stochastic Calculus (Graduate Texts in Mathematics) (Berlin: Springer)
[34] Lawson J and Lim Y 2007 A Birkhoff contraction formula with applications to Riccati equations SIAM J. Control Optim. 46 930–51