Dynamic Data-Driven Adaptive Observations in a Vortex Flowfield

Ryne Beeson*, Hoong Chieh Yeong*, and N. Sri Namachchivaya *
*Department of Aerospace Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Summary. The emergence of sensing concepts that capitalize on the rapidly increasing availability of sensor controllable degrees of freedom, ranging from sensor operating modes to physical control of the platforms carrying the sensors, require new optimal control theory-based strategies for data assimilation of high-dimensional complex systems. In this paper, we consider the application of information theoretic concepts for constructing useful optimization problems for a controllable tracer in a two vortex flowfield; formulations and their effectiveness are demonstrated through numerical simulations, with insightful comments regarding future improvements.

Introduction

A motivating application for the work contained in this paper, pertains to estimation of states for climate modeling. Accurate climate models can only be obtained by combining the models with data. Contrary to standard initial value problems, we do not have access to the initial state of these dynamical systems. Instead, we have a known initial probability density function for the initial state and treat the model states as realizations of a random variable. Lower dimensional climate models can comprise of resolved climate modes and noise terms that account for the interaction of the resolved and unresolved non-climate modes. The problem of climate prediction requires efficient new algorithms to estimate the present and future state of the climate models, based upon corrupted, distorted, and possibly partial observations of the climate modes and fast evolving non-climate modes. While perfect determination of the state is impossible under these noisy observations, it may still be desirable to obtain probabilistic estimates of the state conditioned on the information that is available.

Nonlinear filtering is well-established for the problem of estimation of partially observed states from noisy observations. But advancements in sensing and data collection technology, allow for the possibility of improving the observations available for filtering. Sensors can be designed or manipulated such that the observation obtained contains more useful information for estimating the signal process. In this paper, we consider a realistic problem of controlling a sensor platform (e.g. an unmanned aerial vehicle) to collect the most useful information at each observation time. We will consider the case of discrete, sparse observation in time. A control strategy should then be selected to extremize a certain cost functional that is useful for improving the data assimilation process.

We will begin the paper by introducing the equations of motion for the nonlinear vortex flowfield; specifically, considering the case of an extended state system where the states of the controllable tracer are appended to the states of the vortices. We recall the pertinent equations for the ensemble Kalman filter (EnKF), a Monte Carlo based filter, and make clear the observation process that we will use. Next, we demonstrate the general filtering behavior for Monte Carlo based filters like the EnKF or sequential importance sampling (SIS) particle filter (PF) when applied with modest filtering parameters to the uncontrolled nonlinear two vortex problem; the results, providing motivation for the remainder of the paper. We will then propose cost functionals in terms of information theoretic concepts and describe how these can be efficiently computing when using the EnKF of an SIS PF. Lastly, we provide numerical simulations motivating further investigation and understanding of appropriate information theoretic cost functionals.

Vortex Flowfield

A planar vortex flowfield model is used in this paper as a low-dimensional nonlinear stochastic test-bed, by which we can validate our sensor-theoretic strategies. The first random point vortex method to simulate viscous incompressible flow was introduced in [6] and [15, 1] have shown that a stochastic vortex dynamical model approximates the evolution of vorticity with viscosity, in the same way that the deterministic vortex dynamics simulates the Euler equations. Detection of vortices in data assimilation [5] and observability of vortex models using both Eulerian and Lagrangian tracers have been investigated [11]. The i-th vortex center dynamics in a viscous flowfield is governed by a Langevin equation [1]; here we write an equivalent Itô stochastic differential equation (SDE) form:

$$dX_{i,t} = \frac{1}{2\pi}\sum_{k=1}^{n}\frac{\Gamma_k}{\|X_{k,t}-X_{i,t}\|_2^2}\mathbb{J}(X_{k,t}-X_{i,t})dt + \sqrt{\sigma_x}dB_{i,t}, \qquad \mathbb{J}\equiv\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}, \qquad X_{i,0}=x\in\mathbb{R}^2, \tag{0.1}$$

where we define $0/0 \equiv 0$, $X_{i,t} \in \mathbb{R}^2$, $\Gamma_k \geq 0$ is the strength of the k-th vortex; any vortex with $\Gamma_k = 0$ is termed a Lagrangian tracer, and B_i is the signal noise, a standard Brownian motion with each B_i independent, and $\sigma_x > 0$. To define the dynamics succinctly, let

$$X_t \equiv (X_{1,t}^1, X_{1,t}^2, \ldots, X_{n,t}^1, X_{n,t}^2),$$

where superscripts indicate the components of each vortex center. We will augment our state dynamics to include control and introduce an observation process that will be used in our filtering equations. The control augmented state dynamics

are:

$$dX_{i,t} = f_i(X_t) + b_i(u_{i,t}) + \sqrt{\sigma_x} dB_{i,t} \quad \text{and} \quad u_{i,t} \in \mathcal{U} \tag{0.2}$$

Here f_i is the nonlinear drift vector field given in Equation (0.1), and b_i is a nonlinear function of the control u_i, a candidate control solution from the admissible control set $\mathcal{U}$. Since we will only be interested in controlling the tracers, if X_i is a vortex, then we impose the constraint that $b_i = 0$ for all t.
The observation process is defined as follows:

$$dY_t = h(X_t)dt + \sqrt{\sigma_y} dW_t, \qquad Y_0 = 0 \in \mathbb{R}^d$$

Here h is the sensor function, W_t is the sensor noise, a standard Brownian motion independent of signal noise processes B_i, and $\sigma_y > 0$. If considering random initial conditions for X_0 and Y_0, then these are to also be assumed independent of each other and the signal and observation noise processes.
For the specific problem studied below, we will consider a two vortex and single controllable tracer system. Hence $X \in \mathbb{R}^6$ and we consider nearly perfect observation of the tracer position and velocity, giving $h : \mathbb{R}^6 \to \mathbb{R}^4$. Therefore we never directly observe the states of the vortices.
We will consider discrete (sparse) in-time observations; a natural consideration for climate prediction problems. Therefore let us adopt a notation of using subscript k to indicate the index of an enumeration of a sequence of times $\{t_k\}$, that uniformly partition the full simulation time $[0, T]$; where Δt will designate the difference between any two adjacent times in the sequence (e.g. $\Delta t = |t_k - t_{k-1}|$). With this notation, a sequence of observations and controls would be $\{y_k\}$ and $\{u_k\}$ respectively.

Filtering Theory and the Ensemble Kalman Filter

Nonlinear Filtering Theory

The objective of nonlinear filtering is to solve for the normalized conditional measure [4], $\pi_t(\varphi)$, acting on a continuous bounded functional, $\varphi(x)$, conditioned on the observation process Y_t; succinctly,

$$\pi_t(\varphi) \equiv \mathbb{E}_{\mathbb{Q}} \left[\varphi(X_t) \,|\, \mathcal{Y}_t \right] \quad \text{with} \quad \mathcal{Y}_t \equiv \sigma(\{Y_s \mid s \in [0, t]\}),$$

where $\mathcal{Y}_t$ is the sigma-algebra generated by Y_s up to time t and $\mathbb{Q}$ is the probability measure of an abstract probability space $(\Omega, \mathcal{F}, \mathbb{Q})$. The conditional measure π_t then represents the 'best' estimate of $\varphi(x)$ given all observation information up to time t. As a conditional expectation, π_t is a contraction map and in the case of L_2 functions, is an orthogonal projection onto a Hilbert subspace spanned by the past observations. With a change of measure via a Girsanov transformation [10], one arrives at an evolution equation describing the motion of the unnormalized conditional measure, such that in the case that the new measure $\mathbb{P}$ is absolute continuous with respect to Lebesgue measure, this equation can be described in terms of density; a linear partial differential equation - the Zakai equation:

$$d\rho_t^{\varphi}(x) = \mathcal{G}\rho_t^{\varphi}(x)dt + \rho_t^{\varphi}(x)h^* dY_t \tag{0.3}$$

$\mathcal{G}$ is the generator of the signal process, $(\cdot)^*$ indicates the adjoint, and Y_t is a standard Brownian motion under $\mathbb{P}$. Closed form solutions of Equation (0.3) rarely exists, requiring restrictive conditions for their existence. The best known example of a closed form solution is for the case of linear signal and observation processes driven by standard Brownian motion, resulting in the linear Kalman filter. Sequential Monte Carlo based methods are commonly employed for the general nonlinear filtering problem, for example the ensemble Kalman filter or particle filters [9, 3, 17]. For brevity, we focus on the EnKF in this paper and all results present are for this type of filter algorithm; we will provide pertinent remarks regarding application with sequential importance sampling PF algorithms.

Ensemble Kalman Filter

The EnKF can be viewed as a Monte Carlo approximation of the classical Kalman filter. It uses the full nonlinear dynamics for the prediction step of the algorithm and mimics the linear Gaussian update step of the Kalman filter; using the Jacobian of the sensor function, $\nabla_x h$, and forming a sample prior error covariance, C, in lieu of propagating an error covariance matrix. The EnKF is numerically advantageous for large data assimilation problems where the assumption of Gaussian statistics is appropriate. The assumptions made for the update step of the EnKF also allow for convenient simplifications and reformulations of various entropy relations, as will be seen later in the paper.
The pertinent equations of the EnKF for our two vortex, one tracer problem with discrete-time observations (see [9] for further details) are as follows. At time k, an ensemble of $N_s \in \mathbb{N}$ particles $\{q_{k|k}^j \in \mathbb{R}^6\}_{j=1}^{N_s}$ represents the posterior density $p(x_k|y_k)$, which is assumed to be Gaussian with mean $X_k = \sum q_{k|k}^j / N_s$. A prior density, $p(x_{k+1}|y_k)$ is then constructed at time $k+1$, by advection of the ensemble $\{q_{k|k}^j\}$ under the nonlinear dynamics given by Equation (0.2). The mean of the prior density, $X_{k+1|k} = \sum q_{k+1|k}^j / N_s$ is the best estimate of the signal process at time $k+1$ with information up to time k. The sample prior error, $A_{k+1} \in \mathbb{R}^{N_s \times 6}$ is defined as:

$$A_{k+1} \equiv \left[q_{k+1|k}^1 - X_{k+1|k} \,\middle|\, \dots \,\middle|\, \dots \,\middle|\, q_{k+1|k}^{N_s} - X_{k+1|k} \right]^T$$

The sample prior error covariance, $C \in \mathbb{R}^{6\times 6}$, is then a non-negative symmetric matrix defined as:

$$C_{k+1} \equiv \frac{1}{N-1} A_{k+1}^T A_{k+1} \tag{0.4}$$

The innovation covariance is defined as:

$$S_{k+1} \equiv R_{k+1} + \nabla_x h\big|_{X_{k+1|k}} C_{k+1} (\nabla_x h)^T\big|_{X_{k+1|k}} \tag{0.5}$$

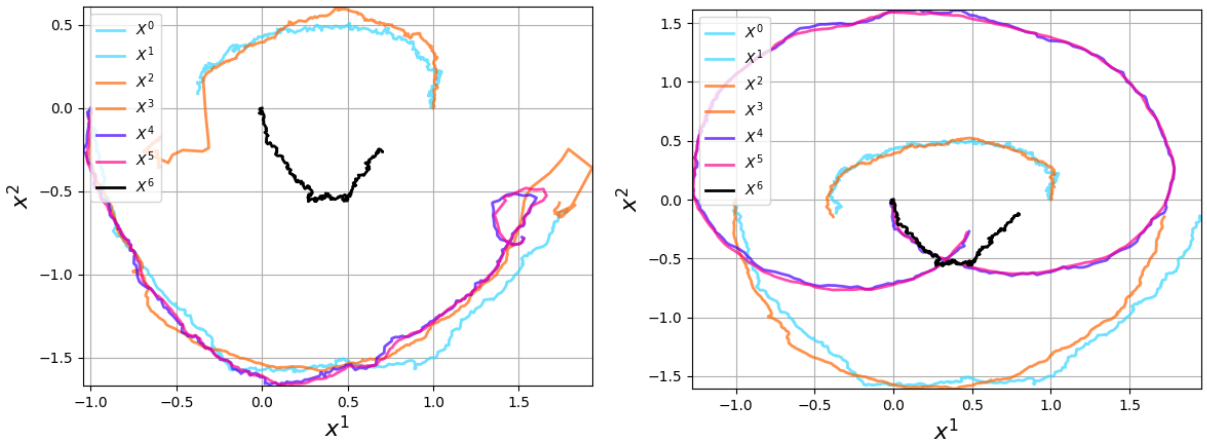
where $R_{k+1} = R \in \mathbb{R}^{4\times 4}$ is the observation noise covariance, a symmetric positive definite matrix. We assume R to be of the form $\sigma_y I_{4\times 4}$, where I is the identity matrix. The sample Kalman gain is defined as $K_{k+1} \equiv C_{k+1}(\nabla_x h)^T\big|_{X_{k+1|k}} S_{k+1}^{-1}$ and the innovation for the j-th member of the ensemble is $y_{k+1} - h(q_{k+1|k}^j)$. Then the update of each member of the ensemble $\{q_{k+1|k+1}^j\}$ is:

$$q_{k+1|k+1}^j = q_{k+1|k}^j + K_{k+1}\left[y_{k+1} - h(q_{k+1|k}^j)\right]$$

The posterior density $p(x_{k+1}|y_{k+1})$ is now approximated by $\{q_{k+1|k+1}^j\}$ and the mean $X_{k+1|k+1} = \sum q_{k+1|k+1}^j / N_s$; completing one iteration of the EnKF algorithm.

EnKF - No Control

We provide numerical solutions of two test cases that capture the typical behavior that may be seen when applying the EnKF to the two vortex, one tracer problem with modest EnKF filter parameters. Similar behavior is seen with SIS PF. Let us denote the true signal process as $\widetilde{X}_k$. We will assume perfect knowledge of the signal states at $t = 0$; that is $X_0 = \widetilde{X}_0$. We will use an Euler-Maruyama integration scheme with integration step size of $\delta t = 0.01$ over a time interval of $[0, T = 30]$. We consider discrete-time observation with observation step size $\Delta t = 10\delta t$ and define $\sigma_x^2 =$ 3E-2, $\sigma_y^2 =$ 5E-4. Lastly, the number of particles used in the simulation will be $N_s = 10$, the first vortex with strength $\Gamma_1 = 1$ will have initial condition $(-1, 0)$ and the second vortex with strength $\Gamma_2 = 2$ will have initial condition $(0, 1)$.
Figure (1a) shows the EnKF behavior for the aforementioned filter parameters for a single (typical) realization when the tracer starts with perfect knowledge of the vortices position and itself with the same initial conditions as the $\Gamma_1 = 1$ vortex; that is (-1, 0). Figure (1a) shows the states up to $t = 27$, with noticeable filter divergence occurring. In the deterministic setting, if the tracer starts with the same initial conditions as the Γ_k vortex, then it will have the same state as Γ_k for all t. Therefore the tracer dynamics are only dependent on the position of the $\Gamma_{j\neq k}$ vortex; implying that the problem becomes a single vortex - single tracer scenario. In the stochastic setting, with probability one the tracer will be perturbed off of the Γ_k vortex. As can be seen from Equation (0.1), the normalizer goes as the square of the distance and therefore the vector field is singular at the location of a vortex. So although being close to a vortex should result in a dimensional reduction, it results in a typical lose of filtering robustness and eventual filter divergence. We note here that due to numerical difficulties, and our use of a fixed step integration scheme, we set the forcing term, in Equation (0.1), associated with the Γ_k vortex to zero if the tracer has an l_2 distance of 1E-1 from the vortex center. Therefore the filtering result shown in Figure (1a) is actually nicer than it would be in reality.



(a) Tracer with initial position $(-1, 0)$, the same as the vortex $\Gamma_1 = 1$

(b) Tracer with initial position $(0, 0)$

Figure 1: Trajectory paths: vortex signals X^0, X^1, vortex estimates X^2, X^3, tracer signal X^4, tracer estimate X^5, and vortex signal mean X^6. The vortex with $\Gamma_1 = 1$ has initial state (-1, 0). The second vortex with $\Gamma_2 = 2$ has initial state (0, 1).

Figure (1b) shows the EnKF behavior, again for the aforementioned parameters, for a single (typical) realization of the tracer starting at the origin $(0, 0)$ and with perfect initial knowledge of the vortex positions. The signal path of the vortices in Figure (1b) is the same as that shown in Figure (1a), but here we show up to time $t = T = 30$. We will use continue to use these two configurations and always the same signal path in the results to follow. Figure (1b) shows that estimation of the unobserved signal dynamics appears superior when starting with this geometry, as direct interference with a vortex is unlikely. Yet, it is quite clear that even with this geometry, the EnKF estimate begins to diverge late in the simulation.

EnKF - Regulator Control

In Figures (1a) and (1b), we have also plotted the mean of the vortices. In this subsection, we show numerically that a control solution that stations the tracer near the mean of the vortices results in robust and accurate estimation. First let us describe the general control strategies that we consider in this paper. Then we will define the control solution used for regulation of the tracer about the mean of the vortices and lastly provides comments on the solution behavior using such a control strategy.

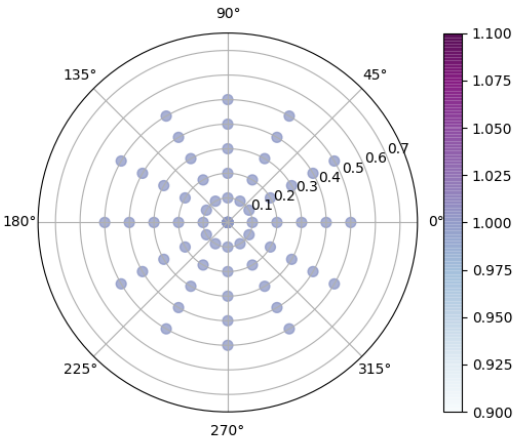
Let $N_o \equiv T/\Delta t$ be the number of discrete observations for a given simulation, $M_\theta, M_G \in \mathbb{N}$, $\overline{G} > 0$, and define the following sets:

$$\Theta \equiv \{2\pi n/M_\theta \,|\, n \in [0, 1, \ldots, M_\theta - 1]\} \quad \text{and} \quad G \equiv \{n\overline{G}/M_G \,|\, n \in [0, 1, \ldots, M_G]\}$$

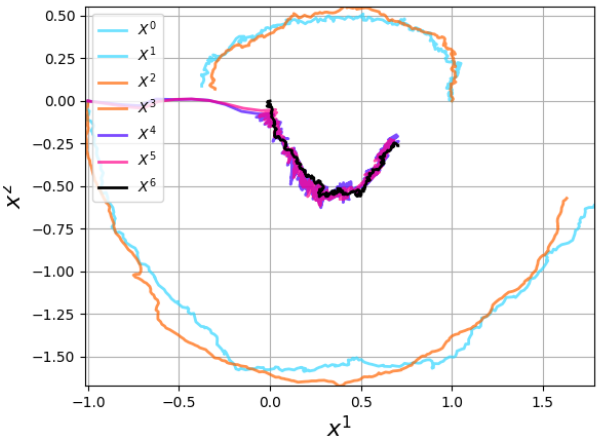
Then the admissible control set is defined as $\mathcal{U} \equiv (\Theta \times G)^{N_o}$. An example of a control target showing a slice $\Theta \times G$ of the admissible control set with $M_\theta = 12$, $M_G = 6$ and $\overline{G} = 0.5$ is shown in Figure (2a); $\overline{G}$ represents the largest control gain that can be applied. In implementation, we consider choosing the values of our control strategy from one observation to the next. The control function is then defined as a constant forcing from one observation to the next as follows: if $\Gamma_k > 0$, we define $b_k = 0\ \forall k$ in Equation (0.2) and otherwise $b_k(u_k) = [u_{k,2}\cos(u_{k,1}), u_{k,2}\sin(u_{k,1})]^T$. Let us consider a simple feedback control law for this section by choosing the control at each time k to be:

$$u_k \equiv \underset{u \in \Theta \times G}{\operatorname{argmin}} \|(m_k - t_{k|k})/\Delta t - b(u)\|_2$$

where m_k is the estimate of the vortices center at time k, $t_{k|k}$ is the posterior estimate of the position of the tracer. Figure (2b) shows the implementation of this control strategy, with an admissible control slice shown in Figure (2a), for the same problem as considered in Figure (1a). The control strategy prevents divergence of the filter and improves estimation beyond that shown for the case of no control and initial condition of (0, 0), see Figure (1b). If one does not start with perfect initial conditions, then such a strategy would not be as successful, as it is dependent on having fair knowledge of the mean of the vortices. In the next section, we introduce information theoretic concepts that could be used to construct useful cost functionals, with the aim that choosing a control strategy $u \in \mathcal{U}$ to maximize (minimize) a given cost functional should result in a control strategy the improves the data assimilation process, both in terms of robustness (i.e. low probability of filter divergence) and accuracy.



(a) A depiction of the set of admissible controls from one observation k to the next $k + 1$. The control is parameterized by an angle and gain (θ_k, γ_k). The color bar shown here is irrelevant, but is used in later figures to show the cost function associated with choosing a given control.



(b) Tracer with initial position $(-1, 0)$ and using a feedback control law that drives it towards the mean of the prior at each observation.

Figure 2: Trajectory paths: vortex signals X^0, X^1, vortex estimates X^2, X^3, tracer signal X^4, tracer estimate X^5, and vortex signal mean X^6. The vortex with $\Gamma_1 = 1$ has initial state (-1, 0). The second vortex with $\Gamma_2 = 2$ has initial state (0, 1).

Information Theoretic Cost Functionals

In this section, we first introduce several entropy concepts that will be used in the construction of information theoretic cost functionals, then we describe potential cost functional candidates and lastly demonstrate through numerical simulation their effectiveness for the two vortex, single tracer problems consider previously. We refer the reader to [7] for more details regarding the definitions given below. We start with the most important definition, that is Shannon entropy.

Definition 0.1 (Shannon Entropy)**.** Shannon entropy, an absolute entropy, quantifies the information content of a random variable, or how much memory (for example, in bits, if using $\log_2$) is required to store the information required to describe the random variable. It can be interpreted as how much uncertainty there is about the random variable. For a continuous random variable, its entropy $H(X)$ as a measure with respect to its density is:

$$H(X) \equiv -\int_{\mathcal{X}} p(x) \log p(x) \mu(dx)$$

From here on, if we refer to the entropy of a random variable, one may assume that we are referring to Shannon entropy.

In the case of a random variable X with normal distribution $\mathcal{N}(\nu, \Sigma)$, the entropy of X is $H(X) = \log((2\pi e)^d |\Sigma|)/2$, where $|\cdot|$ will denote the determinant and d is the dimension of $\Sigma \in \mathbb{R}^{d\times d}$. Joint and conditional entropy between two random variables are defined similarly,

$$H(X,Y) \equiv -\int_{\mathcal{X}\times\mathcal{Y}} p(x,y) \log p(x,y) \mu(dx \otimes dy) \quad \text{and} \quad H(X|Y) \equiv -\int_{\mathcal{X}\times\mathcal{Y}} p(x,y) \log p(x|y) \mu(dx \otimes dy)$$

Definition 0.2 (Kullback-Leibler Divergence)**.** Kullback-Leibler divergence, abbreviated as D_{KL}, is a relative entropy that quantifies the 'distance' between two densities; quotes needed, since it does not satisfy all requirements for being a metric. Given densities p and q, their KL divergence is defined as:

$$D_{KL}(p||q) \equiv \int_{\mathcal{X}} p(x) \log\left(p(x)/q(x)\right) \mu(dx)$$

If p is the actual density for a random variable X, then $D_{KL}(p||q)$ can be interpreted as the loss of information due to using q instead of p.

Definition 0.3 (Mutual Information)**.** Mutual information between two random variables is the relative entropy between the joint density and the product of the marginal densities.

$$I(X;Y) \equiv \int_{\mathcal{X}\times\mathcal{Y}} p(x,y) \log\left(p(x,y)/p(x)p(y)\right) \mu(dx \otimes dy)$$

It can be shown that $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$ by symmetry; a most useful property that we exploit below. Hence, mutual information represents the reduction of entropy (uncertainty) in one random variable due to the knowledge of another random variable.

Cost Functionals

In this subsection, we will define several cost functionals $J(u_k|y_{0:k-1})$, that will allow us to determine the 'best' control $u_k \in \Theta \times G$, to be applied during the integration from time $k-1$ to time k; when the next observation will be made. The 'best' control will be the one that maximizes $J(u_k|y_{0:k-1})$; that is,

$$u_k^o \equiv \underset{u_k\in\Theta\times G}{\operatorname{argmax}} J(u_k|y_{0:k-1}) \tag{0.6}$$

Therefore the cost functional should guide the tracer to locations that 'best' improve the data assimilation process. A natural quantification of 'best' over one observation step becomes a complex question if we are trying to balance estimate improvement over a single observation step versus long term quality of estimate, and also taking into consideration the physical limitations of our tracer. Regardless, an appropriate cost functional serves the purpose of embodying this idea.

KL Maximization

The first cost functional that we consider is the maximization of the KL divergence of the posterior and prior densities, $p(x_k|y_{0:k};u_k)$ and $p(x_k|y_{0:k-1})$:

$$J(u_k|y_{0:k-1}) = \int_{\mathcal{Y}} D_{KL}(p(x_k|y_{0:k};u_k) \,||\, p(x_k|y_{0:k-1};u_k)) p(y_k|y_{0:k-1};u_k) dy_k \tag{0.7}$$

The KL divergence is dependent on Y_k, which is only available after we have selected a control strategy u_k. Therefore, in the general setting, we must take the expectation of the KL divergence over all possible Y_k, given the prior density

and control selection u_k. The motivation for maximizing expected KL divergence, is that the optimal admissible control solution u_k^0 will force a posterior that is distant from the prior and therefore is likely to produce new information. If the posterior density is statistically close to the prior, then very little new information was gained during the previous observation. Initial work on optimal sensor path control based on an information theoretic cost functional of this type was initiated in [13, 14] and furthered in [18], although others have pursued similar studies [8, 12, 16, 2]. Several applications of Bayes' Theorem relates the KL divergence to conditional entropy of the signal given observation:

$$\begin{aligned}
J(u_k|y_{0:k-1}) &= \int_{\mathcal{Y}} D_{KL}(p(x_k|y_{0:k};u_k)\,||\,p(x_k|y_{0:k-1};u_k))p(y_k|y_{0:k-1};u_k)dy_k \\
&= \int_{\mathcal{Y}}\left[\int_{\mathcal{X}} p(x_k|y_{0:k};u_k)\ln\frac{p(x_k|y_{0:k};u_k)}{p(x_k|y_{0:k-1};u_k)}dx_k\right]p(y_k|y_{0:k-1};u_k)dy_k \\
&= \int_{\mathcal{Y}}\left[\int_{\mathcal{X}} \frac{p(x_k,y_k|y_{0:k-1};u_k)}{p(y_k|y_{0:k-1};u_k)}\ln\frac{p(x_k|y_{0:k};u_k)}{p(x_k|y_{0:k-1};u_k)}dx_k\right]p(y_k|y_{0:k-1};u_k)dy_k \\
&= \int_{\mathcal{Y}}\int_{\mathcal{X}} p(x_k,y_k|y_{0:k-1};u_k)\ln p(x_k|y_{0:k};u_k)dx_kdy_k - \int_{\mathcal{Y}}\int_{\mathcal{X}} p(x_k,y_k|y_{0:k-1};u_k)\ln p(x_k|y_{0:k-1};u_k)dx_kdy_k \\
&= \int_{\mathcal{Y}}\int_{\mathcal{X}} p(x_k,y_k|y_{0:k-1};u_k)\ln p(x_k|y_{0:k};u_k)dx_kdy_k - \int_{\mathcal{X}} p(x_k|y_{0:k-1};u_k)\ln p(x_k|y_{0:k-1};u_k)dx_k \\
&= H\big|_{y_{0:k-1}}(X_k) - H\big|_{y_{0:k-1}}(X_k|Y_k;u_k)
\end{aligned}$$

Then using the symmetry property of mutual information given in Definition (0.3), we also get:

$$J(u_k|y_{0:k-1}) = H\big|_{y_{0:k-1}}(Y_k;u_k) - H\big|_{y_{0:k-1}}(Y_k|X_k;u_k) \tag{0.8}$$

Typically one must settle for brute force evaluation (e.g. Monte Carlo methods) of the entropy terms just defined to evaluate $J(u_k|y_{0:k-1})$, but in the case of an update scheme that assumes a linear sensor function and Gaussian statistics for the densities, we can then exploit the comment made in Definition (0.1). Specifically, the terms comprising the relations to J become:

$$\begin{aligned}
H\big|_{y_{0:k-1}}(X_k) &= \frac{1}{2}\ln(2\pi e)^d\big|C_{k|k-1}\big| \\
H\big|_{y_{0:k-1}}(X_k|Y_k;u_k) &= \frac{1}{2}\ln(2\pi e)^d\big|P_{k|k}\big| \\
H\big|_{y_{0:k-1}}(Y_k|X_k;u_k) &= \frac{1}{2}\ln(2\pi e)^d\big|R_k\big| \\
H\big|_{y_{0:k-1}}(Y_k;u_k) &= \frac{1}{2}\ln(2\pi e)^d\big|S_k\big|
\end{aligned}$$

The definition of S_k was given in Equation (0.5) and we further note that here we have specialized these definitions to match their use with the EnKF. Of importance is the fact that Equation (0.8) is written only in terms of the prior and therefore in the application of the EnKF with this control functional, we will not require taking the expectation over all possible observations; this component is already contributing via $H(Y_k;u_k)$. In the case of application with an SIS PF, it is best to use Equation (0.7) directly, since in this case the prior and posterior distributions have the same support; a linear combination of weighted Dirac distributions. In general, one must still account for virtual observations with the SIS PF. But for problems with small sensor noise, few observation realizations will be required.

Prior Error Minimization

A second cost functional of interest, is minimization of the prior error:

$$J(u_k|y_{0:k-1}) = -H\big|_{y_{0:k-1}}(X_k) = -\frac{1}{2}\ln(2\pi e)^d\big|C_{k|k-1}\big| \tag{0.9}$$

We still write the cost functional as a maximization problem for consistency with Equation (0.6).

Weighted Marginal, KL Maximization

Lastly, let $\alpha_k \in [0,1]$ and consider the following cost functional,

$$\begin{aligned}
J(u_k|y_{0:k-1}) &= \alpha_k H^1\big|_{y_{0:k-1}}(Y_k;u_k) - (1-\alpha_k)H^2\big|_{y_{0:k-1}}(Y_k;u_k) \\
&= \frac{\alpha_k}{2}\ln(2\pi e)^d\big|S_k^1\big| + \frac{(1-\alpha_k)}{2}\ln(2\pi e)^d\big|S_k^2\big|
\end{aligned} \tag{0.10}$$

where S_k^1 and S_k^2 are the marginal innovation covariances. The choice of α_k may be chose at each time k to emphasis that information gain should come more from one source than another. Intuitively, if at time $k-1$, the second vortex contribute more to the information gain, then an observation at k may be better suited for filter stability over greediness of apparent information gain from the second vortex. In this example, $\alpha \to 1$ momentarily. The reasoning for this consideration will be made clear by the numerical examples that we next consider.

Numerical Results

In this section, we provide numerical results demonstrating the typical behavior of the EnKF estimates when filtering is conducted in conjunction with an optimal control problem specified by either Equation (0.7), (0.9) or (0.10). Consider the same problem setup as that introduced in the no control and regulator control cases. Solving the aforementioned problem with $T = 10$ and using Equation (0.7) as the cost function results in a typical solution path as shown in Figure (3). The associated cost is also shown in Figure (3) and the costs associated with admissible control strategies at time $t_k = 0.0$ and $t_k = 5.8$ are shown in Figure 4a) and (4b) respectively.

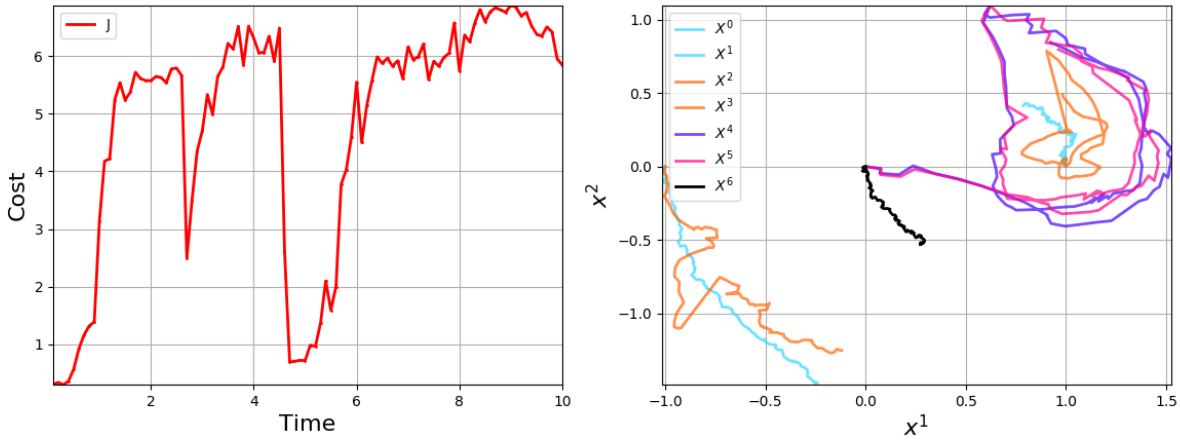
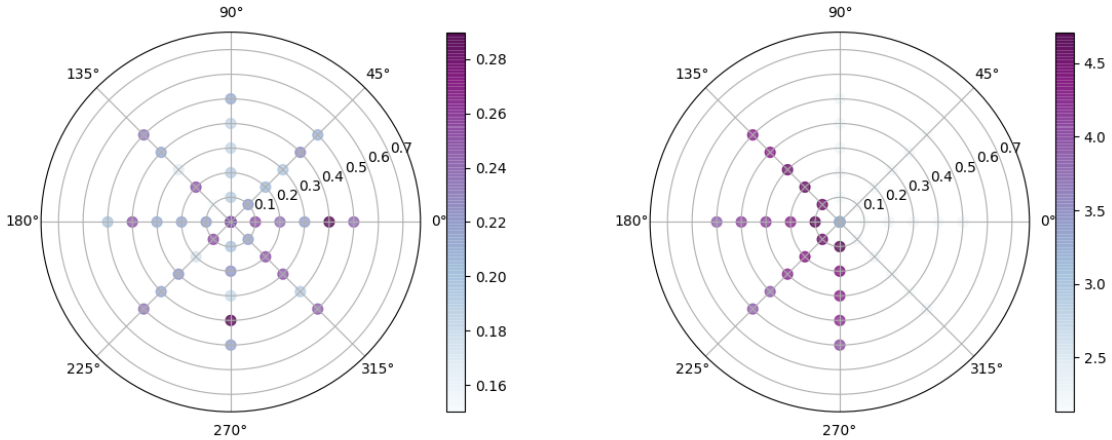


Figure 3: Trajectory paths: vortex signals X^0, X^1, vortex estimates X^2, X^3, tracer signal X^4, tracer estimate X^5, and vortex signal mean X^6 for the case of KL maximization, Equation (0.7).



(a) The admissible controls, u_k, for $t_k = 0.0$ of the solution path shown in Figure (3).

(b) The admissible controls, u_k, for $t_k = 5.8$ of the solution path shown in Figure (3).

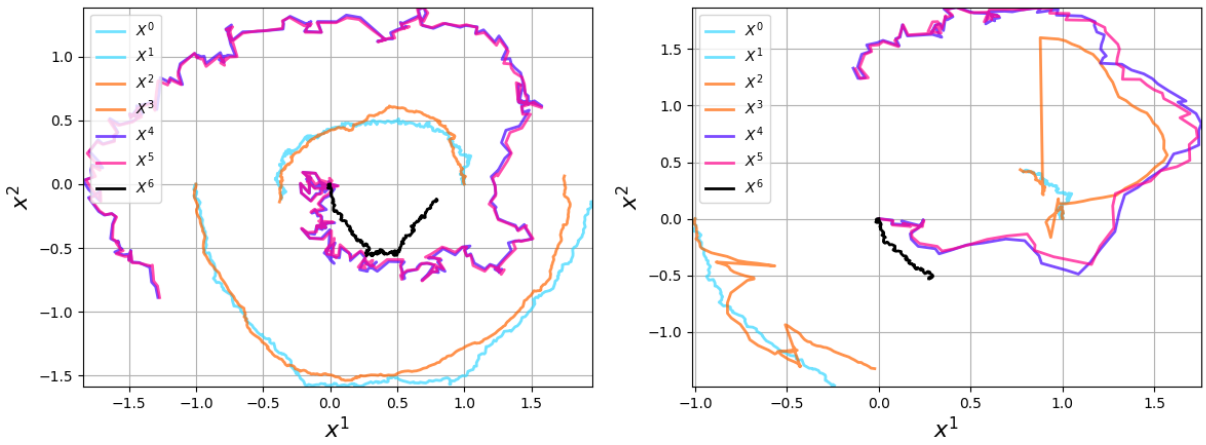
Figure 4

To understand the solution behavior displayed in Figure (3), one may observe that maximization of expected KL divergence encourages the maximization of the determinate of the innovation covariance. With a constant observation covariance, this implies increasing the determinate of $\nabla h C (\nabla h)^T$ evaluated at the prior. As the determinant can be written as a product of the eigenvalues of this linear mapping, it indicates that a control solution maximizing KL divergence will most likely result in a large divergence of the ensemble. A solution that tends towards the strongest vortex, as shown in Figure (3) now makes based on our previous comments regarding the singular vector field near a vortex.

The evaluation of the control solutions at different instances of time also warrant attention. The control solution evaluation at $t_k = 0.0$ shown in Figure (4a), makes clear the fact that without large enough control authority, or virtual simulation of more than one observation step, the optimization of a given cost functional may not be meaningful; it will be strongly dependent on initial conditions. For instance, the control solutions that are strongest in Figure (4a) lie on the $\gamma = 0.4$ circle. A clear preference for a control tending towards the second vortex becomes much less clear if one were to restrict $\gamma < 0.4$ (i.e. not searching this control space). Lastly, the control target after some time becomes mostly uni-direction, as shown in the Figure (4b). This is a good indication that the tracer has found its way to a 'basin' (*if one were minimizing*) associated with the cost function.

It would appear from our numerical results that maximization of the expectation of KL divergence is simply too greedy for the nonlinear, potentially singular, dynamics of the problem consider here. Based on the solution of the regulator control example previously explored, one might heuristically consider a cost functional that minimizes the prior error covariance, since control solutions should avoid driving the tracer towards regions with strong vector fields - like those near the vortices. An example of one realization is shown in Figure (5a). The result here should be compared with Figure (1b). The behavior of the solution path is initial desirable, as it appears to retard the natural motion to drift away from the virtual path describing the mean of the vortices. But it should be mentioned that just minimizing the prior error covariance is also not ideal for the given problem, as solution paths that tend far from the vortices also have favorable objective values. Therefore, the control solutions are sometimes encouraged to push the tracers away from valuable information for too long.

Our last numerical investigation is shown in Figure (5b) and shows a typical path solution that results for minimizing Equation (0.10) with $\alpha_k = 0.75 \, \forall k$, chosen to encourage the tracer to not over-emphasis the stronger vortex in information collection. As mentioned earlier, α_k could be selected at each time k. Here we consider the case of fixed α for numerical simplicity. The behavior of the solution path would appear to be more favorable than simply maximizing the expectation of the KL divergence, but in practice, estimation error tends to be similar.



(a) Solution path associated with minimizing Equation (0.9) for $T = 30$.

(b) Solution path associated with minimizing Equation (0.10) for $T = 10$ and $\alpha = 0.75$.

Figure 5: Trajectory paths: vortex signals X^0, X^1, vortex estimates X^2, X^3, tracer signal X^4, tracer estimate X^5, and vortex signal mean X^6. The vortex with $\Gamma_1 = 1$ has initial state (-1, 0). The second vortex with $\Gamma_2 = 2$ has initial state (0, 1).

Conclusions

In this paper, we have constructed a framework for using information theoretic cost functionals to perform dynamic data-driven adaptive observations in a nonlinear stochastic environment. Our application and approach have been novel with regard to the type of problem consider here; it is not of the typical sensor selection variety. The cost functionals that were suggested and numerical explored have led to better insight into the behavior of solution paths and appropriate modifications that should be further investigated. The results of maximizing information theoretic quantities do show some favorable behavior, but it is obvious that a naive approach to complex and sensitive problems as consider here can result in poorer performance if not appropriate applied.

Acknowledgements: The authors acknowledges the support of the AFOSR under grant number FA9550-17-1-0001

References

[1] O. AGULLO AND A. VERGA, *Effect of viscosity in the dynamics of two point vortices: Exact results*, Phys. Rev. E, 63 (2001).

[2] E. H. AOKI, A. BAGCHI, P. MANDAL, AND Y. BOERS, *A theoretical look at information-driven sensor management criteria*, in Proceedings of the 14th International Conference on Information Fusion (FUSION), Chicago, Illinois, 2011, pp. 1180–1187.

[3] M. S. ARULAMPALAM, S. MASKELL, N. GORDON, AND T. CLAPP, *A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking*, IEEE Trans. Signal Process., 50 (2002), pp. 174–188.

[4] A. BAIN AND D. CRISAN, *Fundamentals of Stochastic Filtering*, Springer, Aug. 2013.

[5] A. BARREIRO, S. LIU, N. S. NAMACHCHIVAYA, P. W. SAUERS, AND R. B. SOWERS, *Data Assimilation in the Detection of Vortices*, Device Applications of Nonlinear Dynamics, (2009), pp. 44–60.

[6] A. J. CHORIN, *Numerical study of slightly viscous flow*, Journal of Fluid Mechanics, 57 (1973), pp. 785–796.

[7] T. M. COVER AND J. A. THOMAS, *ELEMENTS OF INFORMATION THEORY*, Wiley-Interscience, 2nd ed., Jan. 2006.

[8] E. ERTIN, J. W. FISHER, AND L. C. POTTER, *Maximum mutual information principle for dynamic sensor query problems*, Information Processing in Sensor Networks; Lecture Notes in Computer Science, 2634 (2003), pp. 405–416.

[9] G. EVENSEN, *The ensemble kalman filter: theoretical formulation and practical implementation*, Ocean Dynamics, 53 (2003), pp. 343–367.

[10] I. KARATZAS AND S. E. SHREVE, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics, Springer, second ed., Jan. 1998.

[11] A. J. KRENER, *Eulerian and Lagrangian observability of point vortex flow*, Tellus A, 60 (2008), pp. 1089–1102.

[12] C. KREUCHER, J. WEGRZYN, M. BEAUVAIS, AND R. CONTI, *Multiplatform information-based sensor management: an inverted uav demonstration*, In Defense and Security Symposium, International Society for Optics and Photonics: 65780Y-65780Y, (2006).

[13] M. MAMLOUK, *Information based sensor control in a two-vortex flowfield*, May 2013.

[14] M. MAMLOUK, H. C. YEONG, AND N. SRI NAMACHCHIVAYA, *Sensor Control in a Two-vortex Flowfield*, in ENOC 2014, University of Illinois Urbana-Champaign, Apr. 2014, pp. 1–6.

[15] C. MARCHIORO AND M. PULVIRENTI, *Hydrodynamics in two dimensions and vortex theory*, Communications in Mathematical Physics, 84 (1982), pp. 483–503.

[16] A. RYAN AND J. K. HEDRICK, *Particle filter based information-theoretic active sensing*, Robotics and Autonomous Systems, 58 (2010), pp. 574–584.

[17] P. J. VAN LEEUWEN, *Nonlinear data assimilation in geosciences: an extremely efficient particle filter*, Quart. J. Royal Meteor. Soc., 136 (2010), pp. 1991–1999.

[18] H. YEONG, R. BEESON, N. S. NAMACHCHIVAYA, N. PERKOWSKI, AND P. W. SAUER, *Dynamic data-driven adaptive observations in data assimilation for multi-scale systems*, InfoSymbiotics/DDDAS Conference, (2016).