Continuum Covariance Propagation for Understanding Variance Loss in Advective Systems

Shay Gilpin†, Tomoko Matsuo‡, and Stephen E. Cohn§

Abstract. Motivated by the spurious variance loss encountered during covariance propagation in atmospheric and other large-scale data assimilation systems, we consider the problem for state dynamics governed by the continuity and related hyperbolic partial differential equations. This loss of variance is often attributed to reduced-rank representations of the covariance matrix, as in ensemble methods for example, or else to the use of dissipative numerical methods. Through a combination of analytical work and numerical experiments, we demonstrate that significant variance loss, as well as gain, typically occurs during covariance propagation, even at full rank. The cause of this unusual behavior is a discontinuous change in the continuum covariance dynamics as correlation lengths become small, for instance in the vicinity of sharp gradients in the velocity field. This discontinuity in the covariance dynamics arises from hyperbolicity: the diagonal of the kernel of the covariance operator is a characteristic surface for advective dynamics. Our numerical experiments demonstrate that standard numerical methods for evolving the state are not adequate for propagating the covariance, because they do not capture the discontinuity in the continuum covariance dynamics as correlations lengths tend to zero. Our analytical and numerical results demonstrate in the context of mass conservation that this leads to significant, spurious variance loss in regions of mass convergence and gain in regions of mass divergence. The results suggest that developing local covariance propagation methods designed specifically to capture covariance evolution near the diagonal may prove a useful alternative to current methods of covariance propagation.

Key words. covariance propagation, variance loss, data assimilation, advective systems

AMS subject classifications. 35L65, 47B38, 65M22, 86A10, 65M75, 65M32, 65M80

1. Introduction. At the heart of modern data assimilation is covariance propagation. Data assimilation techniques evolve the estimation error covariance along with the model state, either explicitly as in the Kalman filter, implicitly as in variational methods, or using a reduced-rank approximation as in ensemble schemes [19, 30, 7, 10]. To provide context for the problem addressed in this paper, we start with a stochastic model state $N$-vector $\mathbf{q}$, which is propagated discretely in data assimilation schemes from time $t_{k-1}$ to $t_k$ as

$$\mathbf{q}_k = M_{k,k-1} \mathbf{q}_{k-1},$$

Submitted to the editors August 25, 2021.

Funding: This work was funded by the National Science Foundation under grant No. DGE-1650115 and AGS-1848544, and by the National Aeronautics and Space Administration (NASA) Modeling, Analysis and Prediction (MAP) program under Core funding to the Global Modeling and Assimilation Office (GMAO) at Goddard Space Flight Center (GSFC).

†Department of Applied Mathematics, University of Colorado Boulder, Boulder, CO, Ann and H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO, (shay.gilpin@colorado.edu).
‡Ann and H.J. Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO, Department of Applied Mathematics, University of Colorado Boulder, Boulder, CO, (tomoko.matsuo@colorado.edu).
§Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, (stephen.e.cohn@nasa.gov)

This manuscript is for review purposes only.
where $M_{k,k-1}$ is the deterministic $N \times N$ propagation matrix representing the model dynamics. For simplicity, we consider here the linear case with no forcing, random or otherwise. From the model state we can define the $N \times N$ symmetric positive semi-definite covariance matrix at time $t_k$,

$$P_k = \mathbb{E}[(q_k - \bar{q}_k)(q_k - \bar{q}_k)^T],$$

where $\mathbb{E}[\cdot]$ is the expectation operator, $\bar{q}_k = \mathbb{E}[q_k]$ is the mean state, and superscript $T$ denotes transpose. The basic equation of discrete covariance propagation behind modern data assimilation schemes then follows directly from the discrete state propagation (1.1),

$$P_k = M_{k,k-1}P_{k-1}M_{k,k-1}^T,$$

where $P_{k-1}$ and $P_k$ are the covariance matrices at times $t_{k-1}$ and $t_k$, respectively [19, 18, chapter 6]. We omit a process noise term in (1.3), which will be discussed later.

Motivated by atmospheric data assimilation schemes used for global numerical weather prediction [9, 20], we consider covariance propagation associated with hyperbolic partial differential equations (PDEs). Let $x \in S^2_r$, where $S^2_r$ is the surface of the sphere of radius $r$ and take time $t \geq 0$. To fix ideas, we focus on the continuity equation, which describes the continuum evolution of the stochastic model state $q = q(x,t)$ as follows:

$$q_t + v \cdot \nabla q + (\nabla \cdot v)q = 0,$$

$$q(x,t_0) = q_0(x).$$

(1.4)

The subscript $t$ denotes the time derivative unless noted otherwise. The two-dimensional velocity field $v = v(x,t)$ is taken to be deterministic and continuously differentiable while the initial state $q_0$ is stochastic with mean $\bar{q}_0$. Equation (1.4) is the statement of mass conservation for a passively advected tracer in a thin layer of atmosphere between two isentropic surfaces, for instance [25, 14, sections 2.5–2.6]. For the model state in (1.4), we can define the covariance between two points $x_1, x_2 \in S^2_r$ as

$$P(x_1, x_2, t) = \mathbb{E}[(q(x_1,t) - \bar{q}(x_1,t))(q(x_2,t) - \bar{q}(x_2,t))].$$

(1.5)

The corresponding covariance evolution equation for (1.4) on $S^2_r \times S^2_r$ and for $t \geq 0$ becomes

$$P_t + v_1 \cdot \nabla_1 P + v_2 \cdot \nabla_2 P + (\nabla_1 \cdot v_1 + \nabla_2 \cdot v_2)P = 0,$$

$$P(x_1, x_2, t_0) = P_0(x_1, x_2).$$

(1.6)

Here $\nabla_i$ denotes the gradient with respect to $x_i$, and $v_i = v(x_i,t)$, $i = 1, 2$. Both the model state and covariance equations given in (1.4) and (1.6) are hyperbolic PDEs. For the covariance equation, the hyperplane $x_1 = x_2$ is everywhere characteristic, so that solutions on that hyperplane are independent of initial conditions away from that hyperplane [4, p. 3130]. As a result, we show in section 2 that there is a discontinuous change in solutions to (1.6) along the $x_1, x_2$-hyperplane in the limit as correlation lengths approach zero, for example in the vicinity of sharp gradients in the velocity field which can arise naturally, as seen for instance in [23].
First, suppose the initial covariance $P_0$ is continuous on $S^2_1 \times S^2_2$, and denote it by $P^d_0(x_1, x_2)$. Thus, the stochastic initial state $q_0$ is spatially correlated, with variance $\sigma_0^2(x) = P^d_0(x, x)$. It follows that the solution of the covariance evolution equation (1.6) along the $x_1, x_2$-hyperplane corresponds to the variance $\sigma^2(x, t) = P(x, x, t)$ and satisfies

$$\begin{align*}
\frac{d}{dt} \sigma^2 + \nabla \cdot \sigma^2 + 2(\nabla \cdot v) \sigma^2 &= 0, \\
\sigma^2(x, t_0) &= \sigma_0^2(x) = P^d_0(x, x),
\end{align*}$$

(1.7)

where $\sigma^2 = \sigma^2(x, t)$ for $x \in S^2_2$ and $t \geq 0$, which can be derived either from (1.6) or directly from (1.4).

Now, suppose the initial state $q_0$ is spatially uncorrelated, and denote its covariance function by $P^c_0(x_1)\delta(x_1, x_2)$ where $\delta$ is the Dirac delta and $P^c_0$ is continuous on $S^1_1$. As we show in subsection 2.2, covariances that are initially restricted to the $x_1, x_2$-hyperplane remain so for all time, as this hyperplane is a characteristic surface. For $t \geq 0$ the solution to (1.6) is then given by $P(x_1, x_2, t) = P^c(x_1, t)\delta(x_1, x_2)$ where $P^c$ satisfies

$$\begin{align*}
\frac{d}{dt} P^c + \nabla \cdot P^c + (\nabla \cdot v) P^c &= 0, \\
P^c(x, t_0) &= P^c_0(x).
\end{align*}$$

(1.8)

Thus, near zero correlation length, the behavior of solutions of the covariance evolution equation along the $x_1, x_2$-hyperplane changes abruptly from that of the variance equation (1.7) to that of (1.8), in the case of a nonzero divergent velocity field, $\nabla \cdot v \neq 0$.

The characteristic behavior of the $x_1, x_2$-hyperplane in the continuum does not translate into discrete space for typical discretizations (1.1) of (1.4). Diagonal elements of $P^c_k$ in (1.3) depend on off-diagonal elements of $P^c_{k-1}$, and a diagonal initial covariance in (1.3) does not remain diagonal for all time. In this paper, we study the behavior of solutions of the continuum covariance evolution equation (1.6) near the $x_1, x_2$-hyperplane and contrast this with the behavior of discretizations (1.3) near the diagonal, using a combination of analytical and numerical methods. We conduct numerical experiments using a one-dimensional version of (1.4) and study the covariance and variance propagation as a function of correlation length scales of the initial covariance. We find in some cases that the variance propagated numerically according to (1.3) bears little resemblance to that of the continuum variance dynamics (1.7). In particular, variances propagated numerically according to (1.3) tend to be better approximations of (1.8) than of (1.7) for short initial correlation lengths, quite independently of any numerical dissipation or dispersion effects. This property manifests itself as a large, spurious loss of variance in regions of mass convergence, and a spurious variance gain in regions of mass divergence, as our analytical and numerical results illustrate. This behavior is the result of the discontinuous change in the continuum covariance dynamics in the limit as the correlation length tends to zero.

Spurious loss of variance is well known in the data assimilation literature. How covariances propagated through data assimilation schemes tend to underestimate the exact error covariances has long been noted [24, pp. 23–24], but variance loss has been discussed primarily in the context of ensemble schemes [24, 22, 10], where spurious variance loss can be attributed to the use of reduced-rank covariance representations [11]. Several methods have been developed.
to address variance loss to prevent filter divergence, such as covariance inflation [1, 26, 24, chapter 9.2] and a scale-selective generalization of covariance inflation [5, section 2.4.4]. Loss of variance is sometimes addressed through an artificial model error or process noise term added to the discrete covariance propagation in (1.3) [18, sections 8.8–8.9]. Accurately estimating an appropriate model error/process noise term is difficult because spurious variance loss can be due to several different known and unknown sources, though it has been shown that adding a model error term can help rectify the negative impact of variance loss, for instance, by increasing ensemble spread in the case of the ensemble schemes [26, 16]. Stochastic parameterization of subgrid scale physics also helps to increase ensemble spread to prevent filter divergence [2, p. 567]. For the purpose of illuminating a root cause of variance loss, we will consider only the unforced covariance dynamics and omit a model error term, artificial or not, in this work. Specifically, the focus of this paper is on the spurious loss (and gain) of variance associated with the peculiar, discontinuous limiting behavior of solutions of the continuum covariance evolution equation (1.6). This is a full-rank effect. We believe ours is the first work to identify and study this effect.

The layout of this paper is as follows. In section 2, we consider a slightly more generalized form of the continuity equation to study the continuum covariance propagation. In subsection 2.1 we establish the generalized problem, defining the necessary operators and associated PDEs used for the analysis. This is followed by subsection 2.2, which derives the generalized version of (1.8) and discusses the discontinuous change in the continuum dynamics as initial correlation lengths approach zero. The analysis sections are followed by numerical experiments, where we illustrate spurious loss and gain of variance in full-rank covariance propagation through a simple one-dimensional example. Subsection 3.1 details the experimental setup of the one-dimensional problem, describing the numerical propagation methods, discretization schemes, and initial covariances. Subsection 3.2 summarizes the results from these numerical experiments. Section 4 contains concluding remarks, followed by Appendices A and B which contain additional derivations.

2. Analysis. To gain insight into the spurious loss of variance associated with full-rank numerical covariance propagation, we first study covariance propagation in the continuum. We will consider the state and covariance equations as PDEs with solutions in the Hilbert space $L^2$, define the associated linear operators, and use tools from functional analysis to study these equations and operators. This continuum analysis is crucial for interpreting the results of the numerical experiments given later in section 3.

2.1. Preliminaries. Let $\Omega = S^2$ and take $x \in \Omega$ and $t \geq 0$. We will consider the generalized advection equation for the model state $q = q(x, t)$,

\begin{equation}
q_t + v \cdot \nabla q + bq = 0,
\end{equation}

\begin{equation}
q(x, t_0) = q_0(x).
\end{equation}

Here $b = b(x, t)$ is a scalar, noting that setting $b = \nabla \cdot v$ yields the continuity equation (1.4). From (2.1) we have

\begin{equation}
\frac{d}{dt} \int_{\Omega} q^2 dx + \int_{\Omega} (2b - \nabla \cdot v)q^2 dx = 0,
\end{equation}

This manuscript is for review purposes only.
and we assume that $2b - \nabla \cdot v \in L^\infty(\Omega)$ and $q_0 \in L^2(\Omega)$ so that (2.1) has a unique solution $q \in L^2(\Omega)$ for all time. We write this solution as

$$q(x, t) = (\mathcal{M}_tq_0)(x),$$

(2.3)

where $\mathcal{M}_t : L^2(\Omega) \mapsto L^2(\Omega)$ is the solution operator of (2.1),

$$\langle \mathcal{M}_tf \rangle(x) = \int_\Omega M(x, t; \xi)f(\xi)d\xi.$$

(2.4)

The subscript $t$ on all operators denoted using the calligraphy font style, as in $\mathcal{M}_t$, for example, indicates the operator evaluated at time $t$, not the time derivative. The kernel of the operator $\mathcal{M}_t$, $M = M(x, t; \xi)$, is the fundamental solution of (2.1),

$$M_t + v \cdot \nabla M + bM = 0,$$

(2.5)

$$M(x, t_0; \xi) = \delta(x, \xi),$$

where the initial condition is the Dirac delta. Here, we simply view the Dirac delta as the kernel of the identity operator $I : L^2(\Omega) \mapsto L^2(\Omega)$,

$$\langle If \rangle(x) = f(x) = \int_\Omega \delta(x, \xi)f(\xi)d\xi.$$

(2.6)

Equation (2.3) is analogous to the discrete state propagation computed in data assimilation schemes. We can propagate our discrete state $q$ in (1.1) from time $t_0$ to $t_k$,

$$q_k = \underbrace{M_{k,k-1}M_{k-1,k-2}...M_{2,1}M_{1,0}}_{M_{k,0}} q_0,$$

(2.7)

which is the discrete version of (2.3) evaluated at time $t = t_k$; the operator $\mathcal{M}_{tk}$ is the continuum version of the propagation matrix $M_{k,0}$.

With the model state now defined, we can derive the corresponding covariance evolution equation for $P = P(x_1, x_2, t)$ with $x_1, x_2 \in \Omega$ and $t \geq 0$ using the definition given in (1.5),

$$P_t + v_1 \cdot \nabla_1 P + v_2 \cdot \nabla_2 P + (b_1 + b_2)P = 0,$$

(2.8)

$$P(x_1, x_2, t_0) = P_0(x_1, x_2),$$

where again, $\nabla_i$ refers to the gradient with respect to $x_i$, and $v_i = v(x_i, t)$, $b_i = b(x_i, t)$ for $i = 1, 2$.

The solution of the covariance evolution equation (2.8) can be expressed using the fundamental solution operator $\mathcal{M}_t$ and its adjoint $\mathcal{M}_t^*$. The adjoint fundamental solution operator is defined using the inner product over the Hilbert space $L^2(\Omega)$,

$$\langle \mathcal{M}_tf, g \rangle_2 = \langle f, \mathcal{M}_tg \rangle_2 \quad \forall f, g \in L^2(\Omega).$$

(2.9)

The adjoint operator, $\mathcal{M}_t^* : L^2(\Omega) \mapsto L^2(\Omega)$, can be expressed as an integral operator whose kernel $M^*$ is the solution to the adjoint final value problem associated with (2.1),

$$\langle \mathcal{M}_t^* f \rangle(\xi) = \int_\Omega M^*(\xi; x, t)f(x)d\xi.$$

(2.10)
With respect to the method of characteristics, the fundamental solution operator $\mathcal{M}_t$ propagates the solution forward along the characteristics determined by departure points, and the adjoint fundamental solution operator $\mathcal{M}_t^*$ propagates the solution backwards along the characteristics determined by arrival points. This yields the symmetry property [6, p. 729] that at any fixed time $t$ the kernels satisfy

$$M(x, t; \xi) = M^*(\xi; x, t).$$

Using (2.11), we can express the covariance in terms of the kernels of the fundamental solution and adjoint fundamental solution operators,

$$P(x_1, x_2, t) = \int_\Omega \int_\Omega M(x_1, t; \xi_1) P_0(\xi_1, \xi_2) M^*(\xi_2; x_2, t) d\xi_2 d\xi_1,$$

or simply

$$P_t = \mathcal{M}_t P_0 \mathcal{M}_t^*, \tag{2.13}$$

where $\mathcal{P}_0: L^2(\Omega) \rightarrow L^2(\Omega)$ is the operator whose kernel is $P_0$,

$$(\mathcal{P}_0 f)(x_1) = \int_\Omega P_0(x_1, x_2) f(x_2) dx_2, \tag{2.14}$$

and $\mathcal{P}_t: L^2(\Omega) \rightarrow L^2(\Omega)$ is the resulting operator at time $t$,

$$(\mathcal{P}_t f)(x_1) = \int_\Omega P(x_1, x_2, t) f(x_2) dx_2. \tag{2.15}$$

Thus, the covariance evolution equation (2.8) is interpreted as the evolution equation for the kernel of the covariance operator $\mathcal{P}_t$. As with the state propagation, (2.13) evaluated at time $t_k$ is the continuum version of the discrete covariance propagation,

$$P_k = M_{k,0} P_0 M_{k,0}^T, \tag{2.16}$$

following from (1.3) and (2.7).

We next define the (left) polar decomposition of the fundamental solution operator $\mathcal{M}_t$, which will bring to light important properties of the covariance evolution that will be discussed in subsection 2.2. The polar decomposition is a canonical form for all bounded linear operators on Hilbert spaces [28, p. 197]. It is the unique decomposition $\mathcal{M}_t = \mathcal{D}_t \mathcal{U}_t$ where $\mathcal{D}_t = (\mathcal{M}_t \mathcal{M}_t^*)^{1/2}$ and $\mathcal{U}_t$ is a partial isometry.

To derive the polar decomposition, we decompose the fundamental solution into

$$M(x, t; \xi) = d(x, t) u(x, t; \xi) \tag{2.17}$$

where $d = d(x, t)$ and $u = u(x, t; \xi)$ satisfy the following PDEs,

$$d_t + v \cdot \nabla d + \left( b - \frac{1}{2} \nabla \cdot v \right) d = 0, \tag{2.18}$$

$$d(x, t_0) = 1,$$
\[ u_t + v \cdot \nabla u + \frac{1}{2}(\nabla \cdot v)u = 0, \]
\[(2.19)\]
\[ u(x, t_0; \xi) = \delta(x, \xi). \]

The solution \( u \) of (2.19) is quadratically conservative,
\[(2.20)\]
\[ \frac{d}{dt} \int_{\Omega} u^2(x, t; \xi)dx = 0, \]
and therefore defines a bounded linear operator \( U_t : L^2(\Omega) \rightarrow L^2(\Omega) \) whose kernel is the solution to (2.19),
\[(2.21)\]
\[ (U_t f)(x) = \int_{\Omega} u(x, t; \xi)f(\xi)d\xi. \]

The operator \( U_t \) is an invertible isometry and therefore unitary.

We show in Appendix A that the solution of (2.8) with the initial condition \( P_0(x_1, x_2) = \delta(x_1, x_2) \) is \( P(x_1, x_2, t) = d^2(x_1, t)\delta(x_1, x_2) \), where \( d \) is the solution to (2.18). In other words, according to (2.13), \( \mathcal{M}_t \mathcal{M}_t^* \) is in fact a multiplication operator. A multiplication operator \( \mathcal{K} : L^2(\Omega) \rightarrow L^2(\Omega) \) is defined as
\[(2.22)\]
\[ (\mathcal{K} f)(x) = k(x)f(x) = \int_{\Omega} k(\xi)\delta(x, \xi)f(\xi)d\xi, \]
where \( k(x) \in L^\infty(\Omega) \) is the multiplication function. The operator \( \mathcal{M}_t \mathcal{M}_t^* : L^2(\Omega) \rightarrow L^2(\Omega) \),
\[(2.23)\]
\[ (\mathcal{M}_t \mathcal{M}_t^* f)(x) = d^2(x, t)f(x), \]
is a multiplication operator, where the multiplication function satisfies the following differential equation,
\[(2.24)\]
\[ d^2_t + v \cdot \nabla d^2 + (2b - \nabla \cdot v) d^2 = 0, \]
\[ d^2(x, t_0) = 1. \]

As the operator \( \mathcal{M}_t \mathcal{M}_t^* \) is non-negative, its square root exists, and we define the operator \( \mathcal{D}_t : L^2(\Omega) \rightarrow L^2(\Omega) \) as this square root,
\[(2.25)\]
\[ (\mathcal{D}_t f)(x) = ((\mathcal{M}_t \mathcal{M}_t^*)^{1/2} f)(x) = d(x, t)f(x), \]
with the multiplication function for \( \mathcal{D}_t \) being the solution to (2.18). Note that because the operator \( \mathcal{D}_t \) is a multiplication operator with a real-valued multiplication function, it is self-adjoint.

The decomposition (2.17) of the kernel of the fundamental solution operator gives us the (left) polar decomposition of \( \mathcal{M}_t \),
\[(2.26)\]
\[ \mathcal{M}_t = \mathcal{D}_t \mathcal{U}_t, \]
with \( \mathcal{D}_t \) and \( \mathcal{U}_t \) defined above. In the next section, we use this polar decomposition to study the continuum covariance propagation.

This manuscript is for review purposes only.
2.2. Continuum covariance propagation. The solution of the continuum covariance evolution equation (2.8) depends on the initial covariance \( P_0(x_1, x_2) \), and we show in this section how the behavior of the solution changes as the initial correlation length tends to zero. To do so, we will interpret the covariance function \( P(x_1, x_2, t) \) as the kernel of the operator \( \mathcal{P}_t \) defined in (2.15), whose evolution can be written in terms of the fundamental solution operator, its adjoint, and the initial covariance, as given in (2.13).

We consider two cases for the initial covariance \( P_0(x_1, x_2) \). First, assume the initial covariance is continuous on \( \Omega \times \Omega \), and denote it as \( P^c_0(x_1, x_2) \). Since \( P^c_0 \) is continuous, the solution to the covariance equation (2.8) is a strong solution and has a bounded \( L^2 \)-norm. Therefore the corresponding covariance operator, which we will denote as \( \mathcal{P}^d_t \), is a self-adjoint Hilbert-Schmidt operator [28, pp. 210–211]. Hilbert-Schmidt operators are a subclass of the compact operators, and it follows from spectral theory that self-adjoint Hilbert-Schmidt operators only contain eigenvalues in their spectrum [17, pp. 230–232] with the possible exception of zero in the continuous spectrum. The set of eigenvalues is often referred to as the discrete spectrum, hence we can refer to the covariance operator \( \mathcal{P}^c_0 \) as the discrete spectrum covariance operator.

Now, consider the case where the initial state \( q_0 \) is spatially uncorrelated, whose covariance we represent by \( P^e_0(x_1) \delta(x_1, x_2) \), and assume that the function \( P^e_0 \) is continuous on \( \Omega \). The Dirac delta in the initial covariance reduces the initial covariance operator (2.14) to a multiplication operator; denote this operator as \( \mathcal{P}^e_0 \). We can see how this impacts the corresponding covariance operator, which we denote as \( \mathcal{P}^c_t \), by applying the polar decomposition (2.26) to (2.13) with \( \mathcal{P}_0 = \mathcal{P}^e_0 \),

\[
\mathcal{P}_t^c = \mathcal{D}_t \mathcal{U}_t \mathcal{P}^e_0 \mathcal{U}^*_t \mathcal{D}_t = \mathcal{D}_t \tilde{\mathcal{P}}_t \mathcal{D}_t, \tag{2.27}
\]

where

\[
\tilde{\mathcal{P}}_t = \mathcal{U}_t \mathcal{P}^e_0 \mathcal{U}^*_t. \tag{2.28}
\]

From the definition of \( \mathcal{U}_t \) in (2.21), it follows that the operator \( \tilde{\mathcal{P}}_t : L^2(\Omega) \rightarrow L^2(\Omega) \) has kernel \( \tilde{P} = \tilde{P}(x_1, x_2, t) \) given by

\[
\tilde{P}_t + v_1 \cdot \nabla_1 \tilde{P} + v_2 \cdot \nabla_2 \tilde{P} + \frac{1}{2}(\nabla_1 \cdot v_1 + \nabla_2 \cdot v_2)\tilde{P} = 0, \tag{2.29}
\]

whose solution is shown in Appendix A to be

\[
\tilde{P}(x_1, x_2, t) = \tilde{P}^c(x_1, t)\delta(x_1, x_2), \tag{2.30}
\]

where

\[
\tilde{P}^c_t + v \cdot \nabla \tilde{P}^c = 0, \tag{2.31}
\]

\[
\tilde{P}^c(x, t_0) = P^e_0(x). \tag{2.32}
\]

Thus, \( \tilde{\mathcal{P}}_t \) is a multiplication operator, and it follows from (2.18) and (2.31) that \( \mathcal{P}^c_t \) in (2.27) is also a multiplication operator,
CONTINUUM COVARIANCE PROPAGATION FOR UNDERSTANDING VARIANCE LOSS IN ADVECTIVE SYSTEMS

where $P^c = d^2 \tilde{P}$ is the solution to the continuous spectrum equation

\begin{equation}
P^c_t + v \cdot \nabla P^c + (2b - \nabla \cdot v)P^c = 0,
\end{equation}

\begin{equation}
P^c(x,t_0) = P^c_0(x).
\end{equation}

Since multiplication operators contain only a continuous spectrum [17, pp. 219, 240], we will refer to $P^c$ as the continuous spectrum covariance operator and to $P^c_0$ as the continuous spectrum solution.

Thus, we have shown that the solution of the covariance evolution equation (2.8) for initial condition $P^c_0(x_1)\delta(x_1,x_2)$ is $P(x_1,x_2,t) = P^c(x_1,t)\delta(x_1,x_2)$; white noise evolved under the state dynamics (2.1) remains white. Further, we can see explicitly that the diagonal dynamics of the covariance are governed by the continuous spectrum equation (2.33) for spatially uncorrelated initial states, rather than by the variance equation

\begin{equation}
\sigma^2_t + v \cdot \nabla \sigma^2 + 2b\sigma^2 = 0,
\end{equation}

\begin{equation}
\sigma^2(x,t_0) = \sigma^2_0(x) = P^d_0(x,x),
\end{equation}

which follows directly from (2.8) when the initial covariance is continuous on $\Omega \times \Omega$. Only in cases where the velocity field is divergence-free are the continuous spectrum equation and variance equation identical. The diagonal dynamics of the covariance are governed by the variance equation (2.34) for all continuous initial covariances, independently of nonzero initial correlation lengths, but at zero correlation length the diagonal dynamics change abruptly to those of the continuous spectrum equation (2.33).

Covariances associated with spatially uncorrelated initial states, which correspond to covariances with zero initial correlation length scales, are limiting cases in our analysis, as well as in practice. Atmospheric wind fields, for example, have sharp vertical correlation structures relative to long horizontal correlations that need to be represented in covariances for numerical weather prediction models [27], and wind shear leads to tracer correlations that shrink in the direction perpendicular to the flow such that they are no longer spatially resolved [23]. Through careful analysis and use of the polar decomposition, we are able to derive the discontinuous change in the diagonal dynamics as the correlation length approaches zero, which is not readily apparent when first considering the model state and covariance equations or when only considering these equations in discrete space.

3. Numerical experiments. The purpose of the numerical experiments is to examine the evolution of the variance during discrete covariance propagation and its relation to the discontinuous change in diagonal dynamics as the initial correlation length approaches zero we have demonstrated in the continuum. To clearly illustrate the problems associated with full-rank covariance propagation, a simple example is used so that our numerical results can be compared to a known exact solution. We conduct these experiments for various types of initial covariances specified with different correlation kernel functions with varying correlation length scales. For each type, cases with spatially stationary and nonstationary variance are considered.

3.1. Experimental setup. For these experiments, we will consider the one-dimensional version of the continuity equation (1.4) over the unit circle $S^1$. We take the advection speed
to be independent of time and spatially-varying,

\begin{equation}
(3.1) \quad v(x) = \sin(x) + 2.
\end{equation}

The solution to the one-dimensional continuity equation with this advection speed can be obtained explicitly using the method of characteristics and is used as reference for our experiments; see Appendix B for further discussion.

The spatial domain $S_1$ is discretized on a uniform grid, $x_i = i\Delta x$, $i = 0, 1, \ldots, N - 1$ where $N = 200$ and $\Delta x = \frac{2\pi}{N}$. The time discretization is given by $t_k = k\Delta t$ where the time step $\Delta t$ is determined from the Courant-Friedrichs-Lewy number $\lambda$,

\begin{equation}
(3.2) \quad \lambda = \max_{x \in [0, 2\pi]} v(x) \frac{\Delta t}{\Delta x} = 3 \frac{\Delta t}{\Delta x} \leq 1.
\end{equation}

For these experiments, we take $\lambda = 1$. We ran experiments with several other values of $\lambda < 1$ (not shown) and found it did not have a significant impact on the results.

### 3.1.1. Numerical covariance propagation.

We use two methods of propagation to illustrate the impact of numerical schemes on the discrete covariance propagation (1.3) and to leverage insights from continuum covariance analysis in the form of the polar decomposition. The covariance matrix is propagated discretely using either of two methods:

1. Traditional propagation: the covariance is propagated as in (1.3), where the matrix $M_{k,k-1}$ is the finite difference discretization of the fundamental solution equation (2.5).

2. Polar decomposition propagation: the polar decomposition of the fundamental solution operator (2.26) is discretized and used in place of the matrix $M_{k,k-1}$ in (1.3).

In the polar decomposition propagation, (2.26) is discretized as follows. The operator $D_t$ of (2.25) is a self-adjoint multiplication operator, therefore when $t = t_k$ its corresponding discretization is the diagonal matrix $D_{k,0}$ whose diagonal elements are $(D_{k,0})_{ii} = d(x_i, t_k)$. Here $d(x_i, t_k)$ is the solution to (2.18) evaluated on the discrete spatial grid up to time $t_k$, which we generate using the exact solution to (2.18), as discussed in Appendix B. The discretization of the operator $U_t$ of (2.21) is done via finite differences to generate the propagation matrix $U_{k,k-1}$ corresponding to (2.19). The matrix $U_{k,k-1}$, which propagates the solution from time $t_{k-1}$ to $t_k$, is independent of time by virtue of (3.1), therefore we will denote $U_{k,k-1} = U$ for simplicity. To compute the covariance matrix at time $t_k$, $P_k$, we do not construct the matrix $M_{k,k-1}$ explicitly using the polar decomposition, but instead compute the covariance as follows:

\begin{equation}
(3.3) \quad P_k = D_{k,0}U_{k,k-1}U_{k-1,k-2} \ldots U_{1,0}P_0U_{1,0}^T \ldots U_{k,k-1}^T D_{k,0}
\end{equation}

The Lax-Wendroff [21] and Crank-Nicolson [8] finite difference schemes are used to generate the matrices $U_{k,k-1}$ and $M_{k,k-1}$ corresponding to their respective PDEs (2.19) and (2.5). Like $U_{k,k-1}$, $M_{k,k-1}$ is independent of time, therefore through the rest of the paper we will denote $M_{k,k-1}$ simply as $M$. We choose these two simple finite difference schemes to illustrate their contrasting behaviors, particularly when generating the matrix $U$. According

This manuscript is for review purposes only.
to the continuum analysis, the operator $\mathcal{U}_t$ is unitary. The Crank-Nicolson discretization preserves the unitary property of $\mathcal{U}_t$ because the scheme is quadratically conservative, but the Lax-Wendroff scheme does not. In the numerical results, subsection 3.2, propagated covariances are labeled by the finite difference scheme used to construct the matrices $M$ and $U$ (Lax-Wendroff or Crank-Nicolson) followed by the method of propagation (traditional or polar decomposition) as described in the beginning of this section.

![Figure 3.1. Examples of the correlation functions used to generate the initial covariances for numerical propagation. The GC correlation functions (left) are functions of compact support length parameter $c$, where the FOAR correlation functions (right) are functions of length scale $L$.](image)

### 3.1.2. Initial Covariances

We generate four types of initial covariances from the following two different correlation kernel functions: the Gaspari and Cohn (GC) 5th order piecewise rational function [12, Eq. 4.10], and the first order autoregressive function (FOAR) [13, Eq. (23)]. Using each correlation function, we construct spatially stationary and nonstationary initial covariance matrices. The stationary initial covariances have an initial variance of one, while the spatially nonstationary initial covariances take the initial variance with spatially-varying standard deviation given as

$$
\sigma_0(x) = \frac{\sin(3x)}{3} + 1.
$$

Nonstationarity in the initial covariances is introduced only through the variance.

The GC correlation function $C_0(r(x_i, x_j), 1/2, c)$ is a compactly supported approximation
to a Gaussian function, supported on the interval \(0 \leq r(x_i, x_j) \leq 2c\), where

\[
(3.5) \quad r(x_i, x_j) = 2 \sin(|x_i - x_j|/2)
\]

is the chordal distance between \(x_i\) and \(x_j\) on \(S^1\). On the spatial grid for these experiments, values of \(c = 1, 0.25, \) and \(0.05\) correspond to 100, 16, and 3 grid lengths (\(\Delta x\)), respectively, from the peak of the correlation function to where it becomes zero, and 33, 8, and just 1 grid length, respectively, from the peak value of 1 to values less than 0.2; see Figure 3.1 for these examples.

The FOAR correlation function given by

\[
(3.6) \quad F_L(r(x_i, x_j)) = \exp(-r(x_i, x_j)/L)
\]

is continuous but non-differentiable at the origin because of its cusp-like behavior (see Figure 3.1). As with the GC correlation function, the chordal distance (3.5) is used to reflect periodicity of the domain. The FOAR correlation functions are nonzero on the full spatial domain. On the spatial grid, \(L = 0.5, 0.25, \) and \(0.03\) correspond to 26, 12, and just 1 grid length, respectively, from the peak of the correlation to where it becomes less than 0.2.

**Figure 3.2.** Top: Solutions to the one-dimensional variance equation (1.7) and continuous spectrum equation (1.8) at various times with advection speed (3.1), for unit initial condition (top row) and spatially-varying initial condition (bottom row) taken to be the square of (3.4). Time \(T = 3.979\) corresponds to slightly after a full period. Exact solutions for each case are given in black dashed (variance) and black dot-dashed lines (continuous spectrum). Green curves denote the solution to the variance equation computed with Crank-Nicolson (abbreviated CN); blue curves are solutions to the continuous spectrum equation computed with Crank-Nicolson. Regions highlighted in grey correspond to regions where \(\sigma^2 > P^c\); unhighlighted regions correspond to regions where \(\sigma^2 < P^c\).

**3.2. Experimental results.** From the continuum analysis, we expect the diagonal of the covariance matrix \(P_k\) to behave according to the variance equation for covariances with non-
zero initial correlation lengths and according to the continuous spectrum equation for covariances with zero initial correlation length. To establish a baseline, Figure 3.2 illustrates the solutions to the variance equation (1D version of (1.7)) and continuous spectrum equation (1D version of (1.8)) for unit initial condition and for the spatially-varying initial condition given by the square of (3.4). Considering the exact solutions first (black), we see that the dynamics of the variance solution and continuous spectrum solution are quite different due to the spatially-varying advection speed (3.1). We also see that solving either the variance equation or continuous spectrum equation directly using the Crank-Nicolson scheme produces solutions that are nearly indistinguishable from the exact solutions. Solutions to the variance and continuous spectrum equations computed using Lax-Wendroff differ very slightly from Crank-Nicolson and are not shown.

Figure 3.2 shows that propagating the diagonal of the covariance matrix $P_k$ numerically, independently of the rest of the matrix, using the known diagonal dynamics of either the variance equation or the continuous spectrum equation, produces minimal discretization error. Figure 3.3 illustrates further that, at least for covariance matrices with relatively long initial correlation lengths, the numerically propagated full covariance itself also contains only minor discretization errors typically expected from finite difference approximations. Figure 3.3 shows results for the GC initial correlation supported on the full spatial domain ($c = 1$). The dissipative behavior of both Lax-Wendroff schemes are clear in the normalized spectra (left panels), whereas both the traditional Crank-Nicolson propagation and polar decomposition propagation using the Crank-Nicolson $U$ (hereafter referred to as Crank-Nicolson polar decomposition) capture the normalized spectra quite well. The small amount of variance loss and gain seen in both the constant initial variance and spatially-varying initial variance cases (middle panels of Figure 3.3) are consistent with dissipation and phase errors expected from finite differences. The polar decomposition methods (dashed) reduce the errors in the variance only slightly compared to the traditional methods (solid). The correlations at row 150 (right panels), which corresponds to where the variance reaches a maximum as the advection speed is at a minimum, are nearly identical to the exact correlations, suggesting minimal errors in correlation propagation.

The accuracy of numerically propagated full covariances is much worse for short initial correlation lengths, increasingly so as they approach zero. We monitor the total amount of variance lost or gained over time through the trace of the covariance matrix, $Tr(P_k)$. Figure 3.4 shows the trace time series for both the GC and FOAR cases with spatially-varying initial variance as initial correlation lengths tend to zero (the constant initial variance case results are similar to Figure 3.4 and not shown). The GC and FOAR cases in Figure 3.4 exhibit similar behaviors in the trace over time, even though their initial correlation structures are quite different. As the initial correlation lengths tend to zero, the amount of variance lost during propagation increases strikingly in the Lax-Wendroff schemes. The polar decomposition propagation schemes (dashed) are an improvement over traditional propagation in some cases but are worse in others, and generally suffer similar amounts of variance loss. We also observe that as initial correlation lengths tend to zero, the numerical schemes gradually approach their own limiting behavior at $c = L = 0$, rather than a discontinuous change in dynamics as expected from the continuum analysis.

Had we not performed the continuum analysis of section 2, one might assume that the vari-
Figure 3.3. Propagated covariances at the final time $T$ (slightly after a full period) for GC initial correlations with $c = 1$ (supported on the full domain) for all four propagation methods. Top row: unit initial variance. Bottom row: spatially-varying initial variance. The left panels correspond to the normalized spectra (relative to largest eigenvalue), the middle panels show the variances (diagonals), and the right panels show the correlations at row 150 (location of maximum in variance in space and time). The exact normalized spectra are given in solid black, and the exact variances and correlations are given in black dashed. Crank-Nicolson (CN) and Lax-Wendroff (LW) M refer to traditional propagation using Crank-Nicolson or Lax-Wendroff (solid). CN and LW PD refer to propagation using the polar decomposition (dashed) with the matrix $U$ constructed via the Crank-Nicolson or Lax-Wendroff scheme, respectively.

Variance loss in Figure 3.4 is caused simply by dissipation. However, the Crank-Nicolson scheme is not dissipative and yet produces significant variance loss. In fact, we see for the traditional Crank-Nicolson propagation (solid light blue) that there are regions of both variance loss and gain. We also see that for short, nonzero initial correlation lengths ($c = 0.05$, $L = 0.03$ in particular), the numerical schemes better approximate the limiting case of $c, L > 0$ than the correct behavior for $c, L > 0$. This suggests that inaccurate variance propagation is particularly pronounced for short correlation lengths.

The behavior of the trace time series in Figure 3.4 indicates that covariance propagation
Figure 3.4. Trace time series for GC (top row) and FOAR (bottom row) for the spatially-varying initial variance taken to be the square of (3.4). Each panel corresponds to different values of $c$ and $L$, decreasing from left to right towards $c = L = 0$. Refer to Figure 3.3 for description of the curves. For the cases when $c = L = 0$ (rightmost panels), the exact solutions (solid black) are constant in time due to the fact that continuous spectrum solution $P^c$ satisfies the continuity equation (1.8), hence its integral over space is constant.

 itself can be a source of spurious loss and gain of variance, however it does not indicate where exactly this manifests itself. To gain a better understanding of the source of variance loss and gain, we examine various aspects of the propagated covariance matrix for different types of initial covariances as we did in Figure 3.3, but now for covariances with shorter initial correlation lengths.

Figures 3.5 and 3.6 are final time snapshots (in the same format as shown in Figure 3.3) of propagated covariances specified using the GC correlation function with $c = 0.25$ and FOAR correlation function with $L = 0.25$, respectively, which are initially well-resolved as described in subsection 3.1.2 and correspond to the mildest variance loss cases shown in leftmost panels of Figure 3.4. The normalized spectra in both of these cases are similar to the spectra seen in Figure 3.3, where the Lax-Wendroff schemes are severely dissipative and both Crank-Nicolson methods approximate the exact spectrum moderately well.

The variances extracted from the numerically propagated covariances in Figures 3.5 and 3.6 are strikingly different from the exact solution. Though covariances with these values of $c$ and $L$ are well-resolved initially, the resulting variance curves are smooth but wholly inaccurate. Across all four methods of propagation we see regions of both variance loss and variance gain, clearly illustrating that inaccurate variance propagation is the problem more so than dissipation. The variance propagation becomes worse when the initial covariance has a spatially-varying variance (bottom rows of Figures 3.5 and 3.6), which is a more realistic situation in practice. The errors we observe in the variances, interestingly, are not reflected

This manuscript is for review purposes only.
in the normalized spectra; considering the normalized spectra alone would not even hint at the problems occurring in the variance propagation. The correlations, as expected, show dispersion off the diagonal. The oscillatory behavior of both Crank-Nicolson schemes due to numerical dispersion is expected for this finite difference scheme [3, p. 46].

In the limiting case when the initial correlation lengths become zero, we see two contrasting behaviors in the numerically propagated covariances depending on the initial variance; see Figure 3.7. When the initial variance is constant (i.e. the initial covariance is the identity matrix), the Crank-Nicolson polar decomposition is the only scheme that correctly captures the behavior of the exact covariance (top row of Figure 3.7). This is expected from the continuum analysis coupled with the quadratically conservative property of the Crank-Nicolson scheme used to construct the matrix $U$. When the initial variance varies spatially (bottom row of Figure 3.7), the Crank-Nicolson polar decomposition propagation instead behaves more similarly to the Crank-Nicolson traditional propagation and both have regions of variance loss and variance gain. Carefully comparing the traditional and polar decomposition diagonals, the Crank-Nicolson polar decomposition propagation is a slight improvement over the Crank-
Figure 3.6. Same as Figure 3.5 for the FOAR correlation function with $L = 0.25$.

Nicolson traditional propagation, but these differences are relatively minor compared to their absolute errors. The Lax-Wendroff schemes are substantially dissipative in all cases. We also observe that as the values of $c$ and $L$ decrease towards zero, the normalized spectra in Figures 3.3 and 3.5 to 3.7 decay more slowly and become relatively flat. This suggests that low-rank approximations would have difficulty capturing these covariances as correlation lengths shrink.

Comparing the diagonals at the final time across a series of initial correlation lengths in Figures 3.8 and 3.9 demonstrates the severity of the variance loss and gain caused by inaccurate variance propagation and provides a closer look at the approach to a limiting behavior seen in the trace time series. Without prior knowledge of the discontinuous change in diagonal dynamics as the initial correlation length tends to zero, one might surmise from Figures 3.8 and 3.9 that the observed diagonal behavior is caused by dissipation. Knowing the continuum behavior, however, makes it clear that we are observing inaccurate variance propagation associated with the discontinuous change in continuum dynamics. The Crank-Nicolson polar decomposition propagation for a constant initial variance is the only scheme...
Figure 3.7. Same as Figure 3.5 for covariances with zero initial correlation length \((c = L = 0)\).

that captures the correct diagonal behavior when the initial covariance is the identity, and gradually approaches this behavior as the initial correlation length decreases. For all other cases in Figures 3.8 and 3.9, the resulting variances are smooth solutions, but incorrect ones, and gradually approach their own limiting behavior at \(c = L = 0\) rather than changing abruptly as in the continuum case. The Lax-Wendroff schemes are severely dissipative, hardly resembling the correct diagonal dynamics, and are not shown in these and subsequent figures.

The errors in the diagonal propagation are not limited to the final time; errors start to accumulate early on in the propagation cycle. We show this for the GC case in Figure 3.10, where the FOAR results are similar but not shown. We can see clearly that rather than approximating the variance for \(c > 0\) (black dashed), the Crank-Nicolson schemes (blue) tend to approximate the diagonal for \(c = 0\) (brown), e.g. the continuous spectrum, which is not the correct diagonal behavior for this case. Along with the variances extracted from the propagated covariances in Figure 3.10, we include the diagonal propagated independently by solving the one-dimensional version of (1.7) using the Crank-Nicolson scheme, as was shown in Figure 3.2. Including the variance solution computed using the Crank-Nicolson scheme

This manuscript is for review purposes only.
Figure 3.8. Covariance diagonals extracted from Crank-Nicolson traditional and polar decomposition propagation methods at the final time $T$ for GC initial correlations as $c$ approaches zero. Top row: constant initial variance. Bottom row: spatially-varying initial variance. Left column: propagation via traditional Crank-Nicolson. Right column: propagation via Crank-Nicolson polar decomposition. Black curves are the exact diagonals, dashed for covariances with nonzero initial correlation lengths (variance solution) and dot-dashed for covariances with zero initial correlation length (continuous spectrum solution). These exact curves here labeled as Exact ($c > 0$) and Exact ($c = 0$) are the same exact curves labeled as Exact $\sigma^2(x,t)$ and Exact $P^c(x,t)$ in Figure 3.2. In the top right panel, the exact curve for $c = 0$ (black dot-dashed) and CN PD curve for $c = 0$ (magenta) identically overlap.

in Figure 3.10 emphasizes that propagating the covariance diagonal independently using the known diagonal dynamics significantly reduces the errors in the variance compared to the variance extracted from the propagated covariance matrix.

Grey regions in Figure 3.10 correspond to where the exact variance solution is larger than the exact continuous spectrum solution, and these regions tend to correspond to where the numerically propagated diagonal for the short, nonzero initial correlation length (solid light blue) exhibits variance loss. Further insight into this behavior can be gained from the continuum perspective as follows. Assuming $P^d_0(x,x) = P^c_0(x)$, as is the case in Figure 3.10, one can show, using the generalized variance equation (2.34) and continuous spectrum equation
(3.7) \[ \frac{\sigma^2(x, t)}{P^c(x, t)} = m(x, t), \]
where \( m = m(x, t) \) satisfies the continuity equation with unit initial condition,
\[ m_t + \mathbf{v} \cdot \nabla m + (\nabla \cdot \mathbf{v})m = 0, \]
\[ m(x, t_0) = 1. \]

If we consider (3.8) in the context of mass conservation, the ratio \( \sigma^2/P^c \) must be conserved. Therefore in regions of mass convergence (\( m > 1 \)) we have \( \sigma^2 > P^c \), and conversely in regions of mass divergence (\( m < 1 \)) we have \( \sigma^2 < P^c \). In fact, the function \( m \) of (3.8) on \( S^1 \) with advection speed (3.1) can be expressed explicitly using (B.7) taking one as the initial condition. Solving \( m = 1 \), or equivalently \( v(s(x, t)) - v(x) = 0 \) (following the notation of Appendix B), determines the boundaries between regions of mass convergence and divergence in \( S^1 \) at time \( t > 0 \).
Figure 3.10. Variance time series for GC case with $c = 0.05$. Top row: constant initial variance. Bottom row: spatially-varying initial variance. Crank-Nicolson traditional and polar decomposition propagation are shown (solid and dashed blue) as well as propagating the variance by solving the variance equation (one-dimensional version of (1.7)) independently using Crank-Nicolson (green). The exact solutions to the variance and continuous spectrum equations are shown in black-dashed and solid brown and denoted as Exact ($c = 0.05$) and Exact ($c = 0$), respectively. Regions where the exact variance is larger (smaller) than the exact continuous spectrum are highlighted in grey (white).

For short initial correlation lengths, if the numerical schemes are better approximating the continuous spectrum solution $P_c$, regions of mass convergence in the ratio $\sigma^2 / P_c$ should correspond to variance loss and regions of mass divergence in the ratio $\sigma^2 / P_c$ should correspond to variance gain. Regions where the variance (black-dashed) is larger than the continuous spectrum (brown), e.g. mass convergence, highlighted in grey in Figure 3.10, generally do coincide with regions where we see variance loss in the Crank-Nicolson traditional and polar decomposition schemes. Conversely, in the unshaded regions of Figure 3.10 where the variance is less than the continuous spectrum, e.g. mass divergence, the Crank-Nicolson traditional and polar decomposition schemes generally exhibit variance gain. Hence, we can exploit the ratio in (3.7) to quantify variance loss and gain in relation to mass convergence and divergence.

Since the function $m$ must satisfy mass conservation, there will always be regions of mass convergence and mass divergence in the spatial domain at a given time $t$. In the context of advective systems, this implies that we should see both loss and gain of variance over time as a consequence of the ratio in (3.7); whether there will be global variance loss or variance gain depends on the advection speed $\mathbf{v}$. In all the numerical results, Figures 3.3 to 3.10, we see both spurious variance loss and gain as a result of inaccurate variance propagation. Data assimilation literature tends to focus on loss of variance because of its negative impact on data
assimilation schemes [18, section 8.8], [24, section 9.2], [1, 15, 23]. However, we see here for a
general advective system that both loss and gain exist because of conservation, cumulatively
causing inaccurate variance propagation. Variance inflation, a tool often used in data assimil-
ation practice to combat variance loss by rescaling the variance, typically by a multiplicative
or additive factor, could be optimally adjusted in this context. Distinguishing between regions
of mass convergence and divergence in general isolates the regions of variance loss and gain,
thus the inflation factor can be tuned to only inflate in regions of mass convergence where
we expect variance loss. This would avoid using a single scale-factor that inflates the whole
variance function and prevent variance inflation in regions where we already see variance gain.

Since covariance information on and close to the diagonal may be sufficient information
for many applications, local covariance evolution, where the variance and correlation lengths
are propagated rather than the full matrix, may prove useful. [4] first discussed local co-
variance evolution through continuum analysis of hyperbolic and parabolic PDEs, similar to
the equations discussed here. The covariance equation in $2N$ spatial dimensions, where $N$
is the number of spatial dimensions of the state, is reduced to a system of auxiliary PDEs
in $N$ dimensions consisting of variance and correlation length equations, which approximate
the covariance locally. Our results, together with the results of [4], suggest further investiga-
tion into local covariance propagation, which may help reduce computational expense and the
spurious loss and gain of variance observed in full covariance propagation.

4. Conclusions and discussion. In this work, we study covariance propagation associated
with random state variables governed by a generalized advection equation. We do this in
an effort to understand the root causes of spurious loss of variance, which is observed in
data assimilation schemes wherein the covariance is explicitly or implicitly propagated. Using
a continuum analysis to guide the interpretation of our numerical results, new insights are
gained through the detailed study of both the continuum and discrete covariance evolution.
The main conclusions are as follows:

1. Continuum analysis of the state and covariance equations is necessary to establish a
fundamental understanding of the covariance evolution. In particular, the continuum
analysis uncovers the discontinuous behavior of the diagonal dynamics as the correla-
tion length approaches zero, for example in the vicinity of sharp gradients, which is
an insight crucial for understanding the spurious loss and gain of variance observed in
our numerical experiments.

2. Comparison of the numerical results with the continuum analysis shows that full-
rank covariance propagation via (1.3) typically results in considerable spurious loss
of variance. This is due to the peculiar discontinuous behavior of the dynamics far
more than to any numerical dissipation. Discrete propagation using (1.3) produces
inaccurate variances, resulting in both variance loss in regions of mass convergence
and variance gain in regions of mass divergence. When propagating initial covariances
with short, nonzero correlation lengths, the numerical schemes better approximate
the dynamics of the zero correlation length case than those of the nonzero correlation
length case.

3. Isolating the variance and propagating it independently ameliorates the variance loss
and gain observed during full-rank covariance propagation and yields accurate prop-
agation because it adheres to the continuum dynamics. This result suggests further investigation into alternate methods of covariance propagation.

Though the continuum analysis performed here may not be tractable in all situations, it serves as a foundation for understanding covariance evolution and interpreting the results of our numerical experiments. Central to the continuum analysis is the polar decomposition of the fundamental solution operator (2.26), which is general in that it is a canonical form for all bounded linear operators on Hilbert spaces. Thus, we expect that much of the work presented here can be extended to general hyperbolic systems of PDEs having a quadratic energy functional, e.g., [5].

It is important to recognize that the loss of variance observed in our numerical experiments is a result of the discontinuous change in continuum covariance dynamics discussed in subsection 2.2. Even when propagating covariance matrices using a fully Lagrangian scheme, as done in [23], propagated covariances still suffer from spurious loss of variance that is not due to the numerical scheme but rather to the discontinuous change in dynamics that we’ve identified in this work. Covariance propagation tends to be overlooked in the data assimilation literature as a potential source of variance loss, particularly when using the same numerical method that propagates the state. Data assimilation schemes that do not propagate the covariance explicitly may experience errors similar to what we observe here because the underlying cause of these errors is the covariance dynamics, not the numerical scheme. Studying full-rank covariance propagation as in (1.3) isolates the spurious loss of variance as an issue with covariance dynamics and implies that approximations to (1.3), such as in ensemble Kalman filters, can suffer variance loss in a similar manner. The errors caused during the covariance propagation may be a neglected source of the model error or “system error” observed in data assimilation schemes [16, p. 3285].

Our work brings to light a fundamental issue associated with current approaches to numerical covariance propagation and recommends investigation into alternate methods of covariance propagation. Figures 3.2 and 3.10, which display the independent variance propagation for example, suggest that local covariance evolution may be an adequate alternative. As discussed in [4], for applications in which information on and close to the diagonal is sufficient, evolving the variance and correlation lengths themselves may serve as an alternative to full covariance evolution.

Appendix A. Proofs of solutions with Dirac delta initial conditions. We begin by verifying that $P(x_1, x_2, t) = d^2(x_1, t)\delta(x_1, x_2)$, where $d^2$ satisfies (2.24), is the solution to the covariance evolution equation (2.8) for $P_0(x_1, x_2) = \delta(x_1, x_2)$.

Proof. The function $P(x_1, x_2, t) = d^2(x_1, t)\delta(x_1, x_2)$ is a weak (distribution) solution of (2.8) with $P_0(x_1, x_2) = \delta(x_1, x_2)$ if

$$\langle L^* \phi, P \rangle = 0$$

for all test functions $\phi \in C^1(\Omega \times \Omega \times [t_0, T])$ with period $t = T$, where $L^*$ is the adjoint differential operator

$$L^* = -\partial_t - \nu_1 \cdot \nabla_1 - \nu_2 \cdot \nabla_2 + b_1 + b_2 - \nabla_1^t \nu_1 - \nabla_2^t \nu_2$$

This manuscript is for review purposes only.
corresponding to the differential operator

\[(A.3) \quad \mathcal{L} = \partial_t + v_1 \cdot \nabla_1 + v_2 \cdot \nabla_2 + b_1 + b_2\]

of (2.8). Here we denote \((\cdot, \cdot)\) as the inner product over \(L^2(\Omega \times \Omega \times [t_0, T])\) and we will denote \((\cdot, \cdot)'\) as the inner product over \(L^2(\Omega \times [t_0, T])\). Substituting \(P = d^2(x_1, t)\delta(x_1, x_2)\) into the expression for \((\mathcal{L}' \phi, P)\), applying the Dirac delta, and expanding the result yields

\[(A.4) \quad (\mathcal{L}' \phi, P) = (\phi_t(x_1, x_1, t) - \nabla_1 \cdot (v_1 \phi(x_1, x_1, t)) + (2b_1 - \nabla_1 \cdot v_1) \phi(x_1, x_1, t), d^2(x_1, t))' ,\]

where we note that \(\nabla_2 \cdot v_2 |_{x_2 = x_1} = \nabla_1 \cdot v_1\) and \(\nabla_1 \phi(x_1, x_1, t) = \nabla_1 \phi(x_1, x_2, t)|_{x_2 = x_1} + \nabla_2 \phi(x_1, x_2, t)|_{x_2 = x_1}\). Using integration by parts to move derivatives off of the test function \(\phi\) onto \(d^2\) in (A.4) gives us

\[(A.5) \quad (\mathcal{L}' \phi, P) = (\phi(x_1, x_1, t), d_1^2 + v_1 \cdot \nabla_1 d^2 + (2b_1 - \nabla_1 \cdot v_1) d^2)' = 0 ,\]

since \(d^2\) satisfies (2.24). Thus, \(P(x_1, x_2, t) = d^2(x_1, t)\delta(x_1, x_2)\) is a weak solution to (2.8).

To show that \(\tilde{P} = \tilde{P}^c(x_1, t)\delta(x_1, x_2)\) is the solution to (2.29) for \(\tilde{P}_0 = \tilde{P}_0^c(x_1)\delta(x_1, x_2)\) follows an identical proof as given above, where we take

\[(A.6) \quad b_i \mapsto \frac{1}{2} \nabla_i \cdot v_1, \quad i = 1, 2 ,\]

\[(A.7) \quad d^2(x_1, t) \mapsto \tilde{P}^c(x_1, t) ,\]

where \(\tilde{P}^c\) satisfies (2.31).

**Appendix B. Solution to the state equation.** In this appendix we solve the one-dimensional version of (1.4) used in the numerical experiments with advection speed (3.1),

\[(B.1) \quad \begin{align*} q_t + vq_x + v'q &= 0, \quad x \in S^1, \\ q(x, t_0) &= q_0(x) ,\end{align*} \]

\[v = \sin(x) + 2.\]

To solve (B.1), we start by rewriting the equation in terms of the Lagrangian/total derivative

\[(B.2) \quad \frac{D}{Dt} \equiv \partial_t + v \partial_x.\]

By doing so, (B.1) becomes the ordinary differential equation,

\[(B.3) \quad \frac{Dq(x(t), t)}{Dt} = -v'(x(t))q(x(t), t), \quad q(x, t_0) = q_0(x).\]

Associated with (B.3) are the Lagrangian trajectories, or characteristic equations, that determine how the spatial variable \(x\) evolves over time,

\[(B.4) \quad \frac{dx(t)}{dt} = v(x(t)), \quad x(t_0) = s.\]
Here, we take $s$ as a general initial parameter and define our solutions to (B.4) as $x(t; s)$, following the notation of [29, chapter 3].

Since the state equation is reduced to an ordinary differential equation in (B.3), solutions are of the form

$$q(x, t) = q_0(x(t_0))e^\int_{t_0}^{t} -v'(x(\tau))d\tau.$$  \hspace{1cm} (B.5)

We can solve the integral in the exponential of (B.5) using a simple $u-$substitution of $u = x(t)$ along with (B.4),

$$\int_{t_0}^{t} -v'(x(\tau))d\tau = \int_{x(t_0)}^{x(t)} -v'(u)du = \ln \left( \frac{v(s)}{v(x(t))} \right).$$  \hspace{1cm} (B.6)

To rewrite (B.6) in terms of $(x, t)$, we first solve for $x(t; s)$ using the characteristic equation (B.4), then invert to determine $s = s(x, t)$, which can be done since the advection speed $v$ is continuously differentiable [29, chapter 3]. Therefore, substituting (B.6) and $x(t_0) = s(x, t)$ into (B.5), we have the explicit solution to (B.1),

$$q(x, t) = q_0(s(x, t)) \left( \frac{\sin(s(x, t)) + 2}{\sin(x) + 2} \right).$$  \hspace{1cm} (B.7)

$$s(x, t) = 2 \tan^{-1} \left( \frac{\sqrt{3}}{2} \tan \left( \tan^{-1} \left( \frac{2\tan(x/2) + 1}{\sqrt{3}} \right) - \frac{\sqrt{3}t}{2} \right) - \frac{1}{2} \right).$$  \hspace{1cm} (B.8)

The function $\tan^{-1}$ refers to the principle branch of inverse tangent, where $y = \tan^{-1}(x)$, $x \in \mathbb{R}$, $y \in (-\pi/2, \pi/2)$, defined by solving $x = \tan(y)$ [31, section 6.3.3].

Figure B.1 plots (B.7) for two different initial conditions, along with solving (B.1) using Lax-Wendroff and Crank-Nicolson schemes on a uniform grid of 200 grid points and unit Courant-Friedrichs-Lewy number.

The function $m$ of (3.8) on $S_1^1$ with advection speed (3.1) can be expressed explicitly using (B.7) taking one as the initial condition. The exact solution $d$ of (2.18) on $S_1^1$ for $b = v_x$, as in the numerical experiments, can also be computed explicitly from (B.7) (as $d^2$ in this case satisfies (B.1) with unit initial condition).

Acknowledgments. This material is based upon work supported by the National Science Foundation (NSF) Graduate Research Fellowship Program under Grant No. DGE-1650115. TM is supported by the NSF CAREER Program under Grant No. AGS-1848544. SEC is supported by the National Aeronautics and Space Administration (NASA) under Modeling, Analysis and Prediction (MAP) program Core funding to the Global Modeling and Assimilation Office (GMAO) at Goddard Space Flight Center (GSFC). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF or of NASA.

REFERENCES

This manuscript is for review purposes only.
Figure B.1. Solutions to (B.1) for $q_0(x) = 1$ (top row) and $q_0(x) = \sin(3x)/3 + 1$ (bottom row) at four different times. The exact solution (black dashed) is shown along with finite-difference solutions using Crank-Nicolson (blue) and Lax-Wendroff (orange). The finite difference solutions are nearly indistinguishable from the exact solution.

[1] J. L. Anderson and S. L. Anderson, A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Weather Rev., 127 (1999), pp. 2741–2758, https://doi.org/10.1175/1520-0493(1999)127⟨2741:AMCIOT⟩2.0.CO;2.

[2] J. Berner, U. Achatz, L. Batté, L. Bengtsson, A. d. l. Câmara, H. M. Christensen, M. Colin-geli, D. R. B. Coleman, D. Crommelin, S. I. Dolaptchiev, C. L. E. Franzke, P. Friederichs, P. Iskeller, H. Järvinen, S. Jurc̆ke, V. Kitsios, F. Lott, V. Lucarini, S. Mahajan, T. N. Palmer, C. Penland, M. Sakradzija, J.-S. von Storch, A. Weisheimer, M. Wegner, P. D. Williams, and J.-I. Yano, Stochastic parameterization: Towards a new view of weather and climate models, B. Am. Meteorol. Soc., 98 (2017), pp. 565–588, https://doi.org/10.1175/BAMS-D-15-00268.1.

[3] G. Beylin and J. Keiser, An adaptive pseudo-wavelet approach for solving nonlinear partial differential equations, in Multiscale Wavelet Methods for Partial Differential Equations, vol. 6 of Wavelet Analysis and Applications, Academic Press, 1997.

[4] S. E. Cohn, Dynamics of short-term univariate forecast error covariances, Mon. Weather Rev., 121 (1993), pp. 3123–3149, https://doi.org/10.1175/1520-0493(1993)121⟨3123:DOSTUF⟩2.0.CO;2.

[5] S. E. Cohn, The principle of energetic consistency in data assimilation, in Data Assimilation, Springer Berlin Heidelberg, 2010, ch. 7, pp. 137–216, https://doi.org/10.1007/978-3-540-74703-1_7.

[6] R. Courant and D. Hilbert, Methods of Mathematical Physics, vol. II, Interscience Publishers, 1962.

[7] P. Courtier and O. Talagrand, Variational assimilation of meteorological observations with the ad- joint vorticity equation. II: Numerical results, Q. J. Roy. Meteor. Soc., 113 (1987), pp. 1329–1347, https://doi.org/10.1002/qj.49711347813.

[8] J. Crank and P. Nicolson, A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type, Math. Proc. Cambridge, 43 (1947), pp. 50–67, https://doi.org/10.1017/S0305004100023197.

[9] R. Daley, Atmospheric Data Analysis, vol. 2, Cambridge University Press, 1993.

[10] G. Evenson, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99 (1994), pp. 10143–10162, https://doi.org/10.1029/94JC00572.
CONTINUUM COVARIANCE PROPAGATION FOR UNDERSTANDING VARIANCE LOSS IN ADVECTIVE SYSTEMS

[11] R. Furrer and T. Bengtsson, *Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants*, J. Multivariate Anal., 98 (2007), pp. 227–255, https://doi.org/10.1016/j.jmva.2006.08.003.

[12] G. Gaspari and S. E. Cohn, *Construction of correlation functions in two and three dimensions*, Q. J. Roy. Meteor. Soc., 125 (1999), pp. 723–757, https://doi.org/10.1002/qj.4971255417.

[13] G. Gaspari, S. E. Cohn, J. Guo, and S. Pawson, *Construction and application of covariance functions with variable length-fields*, Q. J. Roy. Meteor. Soc., 132 (2006), pp. 1815–1838, https://doi.org/10.1256/qj.05.08.

[14] J. Holton and G. J. Hakim, *An Introduction to Dynamic Meteorology*, Academic Press, 5th ed., 2013.

[15] P. L. Houtekamer and H. L. Mitchell, *An adaptive ensemble Kalman filter*, Mon. Weather Rev., 128 (2000), pp. 416–433, https://doi.org/10.1175/1520-0493(2000)128(0416:AAEKF)2.0.CO;2.

[16] P. L. Houtekamer and H. L. Mitchell, *Ensemble Kalman filtering*, Q. J. Roy. Meteor. Soc., 131 (2005), pp. 3269–3289, https://doi.org/10.1256/qj.05.135.

[17] J. K. Hunter and B. Nachtergaele, *Applied Analysis*, World Scientific Publishing Co., 2001.

[18] A. H. Jazwinski, *Stochastic Processes and Filter Theory*, Academic Press, 1970.

[19] R. Kalman, *A new approach to linear filtering and prediction problems*, Journal of Basic Engineering, 82 (1960), pp. 35–45, https://doi.org/10.1115/1.3662552.

[20] E. Kalnay, *Atmospheric modeling, data assimilation and predictability*, Cambridge University Press, 2003.

[21] P. Lax and B. Wendroff, *Systems of conservation laws*, Communications on Pure and Applied Mathematics, 13 (1960), pp. 217–237, https://doi.org/10.1002/cpa.3160130205.

[22] C. E. Leith, *Theoretical skill of Monte Carlo forecasts*, Mon. Weather Rev., 102 (1974), pp. 409–418, https://doi.org/10.1175/1520-0493(1974)102(0409:TSOMCF)2.0.CO;2.

[23] P. M. Lyster, S. E. Cohn, B. Zhang, L.-P. Chang, R. Menard, K. Olson, and R. Renka, *A Lagrangian trajectory filter for constituent data assimilation*, Q. J. Roy. Meteor. Soc., 130 (2004), pp. 2315–2334, https://doi.org/10.1256/qj.02.234.

[24] P. S. Maybeck, *Stochastic models, estimation and control*, vol. 2, Academic Press, 1982.

[25] R. Menard, S. E. Cohn, L.-P. Chang, and P. M. Lyster, *Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part I: Formulation*, Mon. Weather Rev., 128 (2000), pp. 2654–2671, https://doi.org/10.1175/1520-0493(2000)128(2654:AOSCTO)2.0.CO;2.

[26] H. L. Mitchell and P. L. Houtekamer, *An adaptive ensemble Kalman filter*, Mon. Weather Rev., 128 (2000), pp. 416–433, https://doi.org/10.1175/1520-0493(2000)128(0416:AAEKF)2.0.CO;2.

[27] N. A. Phillips, *The spatial statistics of random geostrophic modes and first-guess errors*, Tellus A, 38 (1986), pp. 314–332, https://doi.org/10.1111/j.1600-0870.1986.tb00418.x.

[28] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis*, Academic Press, 1972.

[29] M. Shearer and R. Levy, *Partial Differential Equations: An Introduction to Theory and Applications*, Princeton University Press, 2015.

[30] O. Talagrand and P. Courtier, *Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory*, Q. J. Roy. Meteor. Soc., 113 (1987), pp. 1311–1328, https://doi.org/10.1002/qj.49711347812.

[31] D. Zwillinger, ed., *CRC Standard Mathematical Tables and Formulae*, CRC Press, Inc., 30th ed., 1996.

This manuscript is for review purposes only.