A Bayesian Characterization of Ensemble Kalman Updates

Frederic J. N. Jorgensen* and Youssef M. Marzouk†

Abstract. The update in the Ensemble Kalman Filter (EnKF), called the Ensemble Kalman Update (EnKU), is widely used for Bayesian inference in inverse problems and data assimilation. At every filtering time step, it approximates the solution to a likelihood-free Bayesian inversion problem from an ensemble of particles $(X_i, Y_i) \sim \pi$ sampled from a joint measure π and an observation $y_\star \in \mathbb{R}^m$. The approximated empirical posterior measure $\hat{\pi}_{X|Y=y_\star}$ is constructed by transporting the particles (X_i, Y_i) through an affine map $L^{\mathrm{EnKU}}_{y_\star}(x, y)$ that is given by the Kalman gain. While the EnKU is exact for Gaussian joints π in the mean-field, exactness alone does not uniquely determine the EnKU. In fact, there are infinitely many affine maps $L_{y_\star}$ that push Gaussian π to the posterior $\pi_{X|Y=y_\star}$. This raises a natural question: which affine map should be used to estimate the posterior? In this paper, we offer a novel characterization of the EnKU among all these affine maps. We start by characterizing the set $\mathrm{E}^{\mathrm{EnKU}}$ of laws for which the EnKU yields exact conditioning, showing that it is much larger than just Gaussian distributions. Next, we show that except for a small class of highly symmetric distributions within $\mathrm{E}^{\mathrm{EnKU}}$ (including Gaussians), the EnKU is the *unique* exact affine conditioning map. Finally, we ask what the largest possible set of measures F is that any measure-dependent affine transport could be exact for. After characterizing F, we prove that the set of measures $\mathrm{E}^{\mathrm{EnKU}}$ for which the EnKU achieves exact conditioning is almost maximal in the sense that $\mathrm{F} = \mathrm{E}^{\mathrm{EnKU}} \cup \mathrm{S}_{\mathrm{nl-dec}}$ with a small symmetry class $\mathrm{S}_{\mathrm{nl-dec}}$. Thus, among affine transports, the EnKU is near-optimal for exact distributional conditioning beyond the Gaussian setting. Further, it is the unique affine update achieving exact conditioning for any measure in F except for a subclass of strongly symmetric distributions.

Key words. Ensemble Kalman filter; stochastic filtering; measure transport; Bayesian inverse problems; uncertainty quantification; mean-field limit; non-Gaussian setting; exact conditioning; data assimilation.

AMS subject classifications. 65C35, 62F15, 93E11

1. Introduction.

Given a probability measure $\pi \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^m)$ with finite second moments, we are considering the problem of likelihood-free Bayesian inversion, namely approximating the posterior $\pi_{X|Y=y_\star}$ given samples $(X, Y)_{i=1}^N \sim \pi$ from the joint. This is a problem commonly arising in the context of inverse problems and data assimilation.

1.1. The Ensemble Kalman Update.

One of the most widely used practical algorithms to solve this problem is the Ensemble Kalman Update (EnKU) as used in the Ensemble Kalman Filter (EnKF) [11, 12]. This method computes an empirical approximation to the posterior by applying the affine map

$$L^{\mathrm{EnKU}}_{y_\star}(x, y) \;=\; x + \hat{K}\,(y_\star - y), \qquad \hat{K} \;=\; \hat{\Sigma}_{XY}\,\hat{\Sigma}^{\dagger}_{YY}, \tag{1.1}$$

to every sample in $(X, Y)_{i=1}^N$, with $\hat{\Sigma}_{XY}$ the empirical cross-covariance between X and Y, $\hat{\Sigma}_{YY}$ the empirical auto-covariance of Y, and † the pseudo-inverse. The resulting empirical distribution of particles $\hat{\pi}_{X|Y=y_\star}$ is an estimate of $\pi_{X|Y=y_\star}$. In the data assimilation literature, $\hat{K}$ is often also referred to as the *Kalman gain*. It is well known that when π is jointly

*Department of Mathematics, Massachusetts Institute of Technology (fjorgen@mit.edu).
†Department of Aeronautics and Astronautics, Massachusetts Institute of Technology (ymarz@mit.edu).

Gaussian with non-singular Σ_{YY} and we have "infinitely many samples" (meaning we can replace empirical covariances $\hat{\Sigma}$ by population covariances Σ), this update is *exact*: $L^{\mathrm{EnKU}}_{y_\star}$ pushes the joint law π to the true posterior, i.e.

$$\left(L^{\mathrm{EnKU}}_{y_\star}\right)_\sharp \pi = \pi_{X|Y=y_\star}$$

π_Y-a.s. in $y_\star \in \mathbb{R}^m$ [7, 11, 12, 33]. Beyond Gaussians, practitioners still deploy the same affine recipe because it avoids likelihood evaluations, and only relies on computing empirical covariances [8, 11, 12]. Further, the ensemble implementation of the Kalman gain inherits the same algebraic structure while remaining computationally frugal: most of the analysis computations are carried out in ensemble space, so that the cost scales favorably with the usually large state dimension n [11, 12]. This makes the method well suited to the common setting where the state dimension n is very large compared to the ensemble size $N \ll n$ [4, 24, 33]. Moreover, a large body of work establishes stability and robustness of the filtering distribution, particularly when paired with covariance inflation and localization, which act as regularizers that suppress sampling error and spurious long–range correlations [3, 17, 19].

1.2. Ambiguity of the Ensemble Kalman Update. The EnKU is often derived by showing its exactness for the case where π is Gaussian. However, exactness does not single out the EnKU. Indeed, as we will explain more later, there are infinitely many affine maps $L_{y_\star} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ with $(L_{y_\star})_\sharp \pi_{XY} = \pi_{X|Y=y_\star}$. Why, then, the particular choice of $K = \Sigma_{XY}\Sigma_{YY}^\dagger$, and in what sense is the EnKF update preferable outside the Gaussian setting? There is literature showing that the Kalman gain is variance–minimizing among linear unbiased estimators for the posterior *mean* [7, 14]. In this paper, we give a new characterization of the EnKF update in which we characterize its properties among affine maps in terms of the predicted posterior *distribution*. We analyze the likelihood-free Bayesian inversion problem and investigate the question of when the EnKF update performs exact Bayesian inversion beyond Gaussian-linear settings. Our analysis will be single-step and focused on the likelihood-free Bayesian inversion setting, thus ignoring many other crucial aspects of filtering such as localization, covariance inflation, small ensemble sizes, and long-term stability [3, 10, 19, 24, 33, 42, 43].

1.3. Formalizing the Problem. The EnKU approximately solves the likelihood-free Bayesian inversion problem by transporting the empirical measure $\hat{\pi} = \frac{1}{N}\sum_{i=1}^{N} \delta_{(X_i,Y_i)}$ to the approximated posterior

$$\left(L^{\mathrm{EnKU}}_{\hat{\pi},y_\star}\right)_\sharp \hat{\pi} = \hat{\pi}_{X|Y=y_\star}$$

with $L^{\mathrm{EnKU}}_{\hat{\pi},y_\star}$ as in Equation 1.1. We include $\hat{\pi}$ in the subscript to make the dependency on the samples through the sample covariances explicit. Put differently, the EnKU takes a pair $(\hat{\pi}, y_\star)$ and returns an affine map $L^{\mathrm{EnKU}}_{\hat{\pi},y_\star}$. As such, it belongs to a broader class of transports that we term *affine conditioning maps*.

Definition 1.1. *An* affine conditioning map *is a mapping*

$$L : \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^m) \times \mathbb{R}^m \longrightarrow \{\textit{affine maps } \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n\}, \qquad (\pi, y_\star) \longmapsto L_{\pi,y_\star},$$

such that each $L_{\pi,y_\star}$ *admits the affine representation*

$$L_{\pi,y_\star}(x, y) = A(\pi, y_\star)\, x + B(\pi, y_\star)\, y + c(\pi, y_\star),$$

where $A(\pi, y_\star) \in \mathbb{R}^{n\times n}$, $B(\pi, y_\star) \in \mathbb{R}^{n\times m}$, *and* $c(\pi, y_\star) \in \mathbb{R}^n$.

It is clear that the EnKU is an affine conditioning map L^{EnKU} as defined through Equation 1.1. Note that A, B, and c in Definition 1.1 are allowed to depend on all of $\hat{\pi}$ and in particular on any of its moments. We will often write $L_{\hat{\pi},y}$ for L to make this dependency explicit. Despite this generality, in the course of this paper we will see that the EnKU is uniquely distinguished among *all* affine conditioning maps and that its predicted posterior $\hat{\pi}_{X|Y=y_\star}$ is very often accurate when *any* other affine conditioning map produces an incorrect prediction. To formalize all these claims, we will carry out the analysis of this paper in the *mean–field* setting. This simply means that we replace all empirical measures $\hat{\pi}$ with the true population measures $\pi \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^m)$. An example of this is replacing the sample covariances $\hat{\Sigma}$ in the Kalman gain with the population quantities Σ. Mean-field derivations are standard for transport-based methods: the maps are often derived in the continuum and then implemented with finite ensembles [7, 33]. Philosophically, this corresponds to assuming that we are in an asymptotic regime where N is large and sample quantities are close to population quantities. Central limit theorems connect the mean-field theory to the empirical reality through bounds of the form $\mathbb{E}\|\hat{\Sigma} - \Sigma\|_{\mathrm{F}} \propto N^{-1/2}$ for i.i.d. and certain non-i.i.d. settings [13, 16, 22, 25].
Returning to our problem, we are interested in the task of affine-transport based conditioning in the mean-field: given a measure $\pi \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^m)$ we want to find an affine conditioning map such that the distributional equation

$$(L_{\pi,y_\star})_\sharp \pi = \pi_{X|Y=y_\star} \tag{1.2}$$

holds. We formalize this in the following definition.

Definition 1.2. *Let* $\pi \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^m)$, $y_\star \in \mathbb{R}^m$, *and fix a version of the Markov kernel* $y \mapsto \pi_{X|Y=y}$. *We say that an affine map* $\ell(x, y) := Ax + By + c$ *with fixed* $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, c \in \mathbb{R}^n$ *is an exact affine map at* $y_\star$ *for* π *if*

$$\ell_\sharp \pi = \pi_{X|Y=y_\star}.$$

We say that an affine conditioning map L *is an exact affine conditioning map at* $y_\star$ *for* π *if* $\ell := L_{\pi,y_\star}$ *is an exact affine map at* $y_\star$ *for* π. *Further, if* π_Y*-a.s. in* $y_\star \in \mathbb{R}^m$ *it holds that* L *is an exact affine conditioning map at* $y_\star$ *for* π, *then we say that* L *is an exact affine conditioning map for* π. *This is abbreviated by "*L *is exact for* π*" or just "*L *is exact" if* π *is clear from the context.*

Crucially, note that exact affine conditioning at $y_\star$ for π requires a choice of the Markov kernel $\pi_{X|Y=y}$ and we will only invoke this definition if such a choice was made beforehand. Exactness of $L_{\pi,y}$ for π on the other hand is independent of the choice of Markov kernel $\pi_{X|Y=y}$. In the mean-field, the Ensemble Kalman Update is often motivated from a perspective of exact affine conditioning. Say that $\pi \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^m)$ is Gaussian with mean μ and covariance Σ. Then, defining the π-dependent Kalman gain $K = \Sigma_{XY}\Sigma_{YY}^\dagger$ and defining $L^{\mathrm{EnKU}}_{\pi,y_\star}$ through $A = I$, $B = -K$, $c = Ky_\star$ defines the Ensemble Kalman Update. A simple covariance calculation shows that $L^{\mathrm{EnKU}}_{\pi,y_\star}$ is indeed an exact affine conditioning map, no matter what Gaussian π is. However, there are infinitely many other affine conditioning maps $L_{\pi,y_\star}(x, y) = A(\pi, y_\star)x + B(\pi, y_\star)y + c(\pi, y_\star)$ that are exact for Gaussians. For example, for every choice of

B (assuming $\mathrm{Cov}(X + BY)$ has full rank for all $y_\star$) there are A and c such that $L_{\pi,y_\star}(x, y) = A(\pi, y_\star)x + B(\pi, y_\star)y + c(\pi, y_\star)$ is exact. This is a simple consequence of the fact that $X + BY$ is Gaussian and there is an affine transport map between any two non-singular Gaussians. So the resulting natural question is: why do we pick the EnKU out of all these possible choices? In this exposition, we characterize the EnKU beyond Gaussian settings from the perspective of exact conditioning as we defined previously. In order to better understand what distinguishes the EnKU among affine conditioning maps (and what does not), we study the exact conditioning set of the EnKU. Define the *exact set* of an affine conditioning map L:

$$\mathrm{E}(L) := \{\pi \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^m) \mid L \text{ is an exact affine conditioning map for } \pi\}. \tag{1.3}$$

We answer the following two questions in Section 2:

1. What is the set of measures $\pi \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^m)$ such that the EnKU update L^{EnKU} is an exact conditioning map for π? Or, more formally, what is the exact set
$$\mathrm{E}^{\mathrm{EnKU}} := \mathrm{E}(L^{\mathrm{EnKU}}) \tag{1.4}$$
of the EnKU?
2. Given a measure $\pi \in \mathrm{E}^{\mathrm{EnKU}}$ for which the EnKU is exact and an observation $y_\star \in \mathbb{R}^m$, is the EnKU update $L^{\mathrm{EnKU}}_{\pi,y_\star}$ the only affine map achieving exact affine conditioning? Or can there be other maps?

The first question is answered in Proposition 2.1. In Theorem 2.4 we answer the second question: excluding strongly symmetric distributions, given $\pi \in \mathrm{E}^{\mathrm{EnKU}}$, $y_\star \in \mathbb{R}^m$ the EnKU update $L^{\mathrm{EnKU}}_{\pi,y_\star}$ is the *only* affine map that is exact for π at $y_\star$. Conversely, when choosing an affine conditioning map L from the infinitely many possibilities, to reduce bias we may choose L for which the set $\mathrm{E}(L)$ is maximally large. In Section 3, to study this question, we define *weakly $y_\star$-dependent affine conditioning maps* (short "weakly $y_\star$") as affine conditioning maps L of the form

$$L_{\pi,y_\star}(x, y) = A(\pi)x + B(\pi)y + c(\pi, y_\star),$$

generalizing commonly used affine conditioning maps like the EnKU or square-root updates [11, 12, 29, 41]. We investigate the size of the largest possible exact set $\mathrm{E}(L)$ that any weakly $y_\star$-dependent L might have, which turns out to take the form

$$\mathrm{F} := \bigcup_{L \text{ weakly } y_\star} \mathrm{E}(L).$$

In Theorem 3.3 we show that the EnKU is exact on all of F except for pathological counterexamples, thereby almost achieving the smallest possible bias any weakly $y_\star$-dependent affine conditioning map can have. More formally, we show that there is a small symmetry class $\mathrm{S}_{\mathrm{nl-dec}} \subseteq \mathrm{F}$ such that

$$\mathrm{F} = \mathrm{E}^{\mathrm{EnKU}} \cup \mathrm{S}_{\mathrm{nl-dec}},$$

showing that the EnKU is the optimal weakly $y_\star$-dependent affine transport up to the set $\mathrm{S}_{\mathrm{nl-dec}}$.

1.4. Notation. For $d \in \mathbb{N}$, we always consider $\mathbb{R}^d$ with inner product $\langle \cdot, \cdot \rangle$ and Euclidean norm $\| \cdot \|_2$; I_d or simply I is the $d \times d$ identity. For a matrix A, $A^\top$ is the transpose, $A^\dagger$ the Moore–Penrose pseudoinverse, and $\sqrt{A}$ denotes the principal symmetric square root when $A \succeq 0$. For an endomorphism/square matrix A, we refer to the spectrum through $\sigma(A)$. $\mathrm{GL}(n)$ is the general linear group. The Frobenius norm is $\| \cdot \|_{\mathrm{F}}$. Let $R_\theta \in \mathbb{R}^{2\times 2}$ denote the 2D rotation matrix $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, which rotates vectors in $\mathbb{R}^2$ counterclockwise by angle θ. $\mathcal{P}_2(\mathbb{R}^d)$ is the set of Borel probability measures on $\mathbb{R}^d$ with finite second moment. For a random vector X, its law is $\mathrm{Law}(X) \in \mathcal{P}_2(\mathbb{R}^d)$, expectation $\mathbb{E}(X)$, covariance $\mathrm{Cov}(X) = \mathbb{E}((X - \mathbb{E}X)(X - \mathbb{E}X)^\top)$, and centered version $\overline{X} := X - \mathbb{E}X$. For a joint law $\pi \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^m)$, π_X, π_Y denote the marginals of the $\mathbb{R}^n$ and $\mathbb{R}^m$ parts. $\pi_{X|Y=y}$ is a (fixed) version of the conditional law (a Markov kernel). Given a joint law $\pi \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^m)$ with marginals $X \sim \pi_X$, $Y \sim \pi_Y$, $\Sigma_{XY} := \mathrm{Cov}(X, Y) \in \mathbb{R}^{n\times m}$ is the cross–covariance under π, and $\Sigma_{YY} := \mathrm{Cov}(Y) \in \mathbb{R}^{m\times m}$ is the auto–covariance of Y under π. We say that $X_1 \overset{d}{=} X_2$ for random vectors X_1, X_2 if they have the same law. Independent copies are denoted by superscripts, e.g. $X^{(k)}$. The Dirac mass at x is δ_x. For a measurable map T, the pushforward is $T_\sharp \mu$. W_2 is the 2-Wasserstein distance on $\mathcal{P}_2(\mathbb{R}^d)$. For a subset $W \subseteq T$, W^{c} is the complement.

2. Characterizing the EnKF Update.

The EnKU takes the familiar form

$$L^{\mathrm{EnKU}}_{\pi, y_\star}(x, y) \;=\; x + K\,(y_\star - y), \qquad K(\pi) \;=\; \Sigma_{XY}\,\Sigma_{YY}^\dagger.$$

It is well known that the EnKU is exact for Gaussian distributions [23,33] . In this section, we will go beyond Gaussian distributions by identifying the set of measures $\mathrm{E}^{\mathrm{EnKU}} = \mathrm{E}(L^{\mathrm{EnKU}})$ (defined in Equation 1.3) on which the EnKU is exact and understanding the structure of filters that are exact for some element of $\mathrm{E}^{\mathrm{EnKU}}$. The EnKU does so by taking every x-sample and correcting it linearly with its corresponding increment $K(y_\star - y)$. This reveals the underlying structure of the EnKU: more so than operating on a Gaussian assumption, it operates on the assumption that there is a joint linear relationship between X and Y. Informally, if we can approximately expand

$$X \approx Z + MY + O(Y^2)$$

for Z independent of Y, a matrix $M \in \mathbb{R}^{n\times m}$, and $O(Y^2)$ suppressed, then the EnKU will yield accurate results. The following proposition formalizes this idea, completely characterizing all laws in $\mathrm{E}^{\mathrm{EnKU}}$.

Proposition 2.1. *Let $\mathcal{L}$ be the class of linear maps from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^n$. Then the following equation fully characterizes the exact set:*

$$\mathrm{E}^{\mathrm{EnKU}} = \Big\{ \pi \in \mathcal{P}_2\left(\mathbb{R}^n \times \mathbb{R}^m\right) \mid \exists \pi_{X|Y}, \nu \in \mathcal{P}_2\left(\mathbb{R}^n\right), O \in \mathcal{L} \quad s.t. \quad \pi_{X|Y=y} = O(\cdot, y)_\sharp \nu \;\forall y \in \mathbb{R}^m \Big\}. \tag{2.1}$$

A proof can be found in the appendix.

Remark 2.2. In settings where there are non-linear features ϕ such that $X \overset{d}{=} \phi(Z, Y)$, natural extensions of the EnKU like the conditional mean filter ($\phi(Z, Y) = Z + f(Y)$) [18,26] or the stochastic map filter (ϕ has triangular structure) [40] exist and have been studied.

The question we answer in the remainder of this section is whether there can be other affine transports

$$L_{\pi,y_\star}(x,y) = A(\pi,y_\star)x + B(\pi,y_\star)y + c(\pi,y_\star)$$

besides the EnKU that achieve exactness for $\pi \in \mathrm{E}^{\mathrm{EnKU}}$. In order to gain some intuition, we go back to the set of Gaussian π which is clearly contained in $\mathrm{E}^{\mathrm{EnKU}}$ as can be seen by considering the law of its posterior. As we explained in the introduction, there are many other affine conditional maps implementing exact Bayesian updates for Gaussian distributions π. The fundamental reason for this degree of freedom in the choice of L is that Gaussian distributions have strong symmetries. The law of a Gaussian vector $G \in \mathbb{R}^d$ is a *stable distribution*, meaning that the sum of two independent Gaussians is, again, Gaussian [30, 36, 47].[1] As a consequence of that, they are *self-decomposable*, meaning that for every $\lambda \in (0,1)$, G is λ-*decomposable* [27, 32, 37], meaning there exists another independent G_λ (that is actually also Gaussian) such that

$$G \overset{d}{=} \lambda G + G_\lambda.$$

Another strong symmetry non-singular Gaussian vectors G possess is a rescale-then-rotate symmetry: there is a matrix $C \in \mathbb{R}^{d\times d}$ (e.g. the inverse of any square root of the covariance matrix) such that $C\overline{G}$ is distributionally symmetric under any rotation. In the following theory we will demonstrate that it is due to these symmetries that there are many possible choices of exact affine conditioning maps for Gaussians. A third symmetry leading to many possible choices of conditioning maps is the case in which $Z \sim \nu$ corresponding to some $\pi \in \mathrm{E}^{\mathrm{EnKU}}$ has constant components, meaning that $v^\top Z$ is a.s. constant for some $v \neq 0$ (or equivalently Z has a singular covariance matrix).
Generalizing these three symmetries, namely singular covariance matrices, λ-decomposability of the joint, and the rescale-then-rotate symmetry to non-Gaussian joints, leads to the final EnKU characterization result presented in Theorem 2.4.

Definition 2.3. *We define the sets* $\mathrm{S}_{\mathrm{cov}}, \mathrm{S}_{\mathrm{dec}}, \mathrm{S}_{\mathrm{cyc}} \subseteq \mathrm{E}^{\mathrm{EnKU}}$. *Consider any* $\pi \in \mathrm{E}^{\mathrm{EnKU}}$, *meaning that there exist* $\nu \in \mathcal{P}_2(\mathbb{R}^n)$ *and a linear map* M *such that for* $Y \sim \pi_Y$ *and* $Z \sim \nu$ *independently,* $(Z + MY, Y) \sim \pi$. $\pi \in \mathrm{S}_{\mathrm{cov}}$ *if and only if* ν *has singular covariance.* $\pi \in \mathrm{S}_{\mathrm{dec}}$ *if and only if there exist complex vectors* $v \in \mathbb{C}^n\backslash\{0\}$, $w \in \mathbb{C}^m$, *and constants* $\lambda \in \mathbb{C}$, $|\lambda| < 1$, $b \in \mathbb{C}$ *such that*

$$v^\top \overline{Z} \overset{d}{=} \sum_{k=0}^{\infty} \lambda^k w^\top \overline{Y}^{(k)} + b$$

for i.i.d. copies $\overline{Y}^{(k)}$ *of* $\overline{Y}$. $\pi \in \mathrm{S}_{\mathrm{cyc}}$ *if and only if there exist real vectors* $v_1, v_2 \in \mathbb{R}^n\backslash\{0\}$ *such that* $Z_{\mathrm{cyc}} = (v_1^\top Z, v_2^\top Z)^\top$ *satisfies cyclic symmetry of some order* $k \in \mathbb{N}_{\geq 2}$, *meaning that*

$$\overline{Z}_{\mathrm{cyc}} \overset{d}{=} R_{2\pi/k}\overline{Z}_{\mathrm{cyc}}$$

for R_θ *the 2D rotation by angle* $\theta = \frac{2\pi}{k}$.

[1] Gaussians actually are the only stable random variables with finite second moment [13].

Within $\mathrm{E}^{\mathrm{EnKU}}$, each of the symmetry classes above carves out a highly non-generic and (topologically speaking) small subset of laws. If $\pi \in \mathrm{S}_{\mathrm{cov}}$, then ν has singular covariance, so Z lives a.s. in a proper linear subspace of $\mathbb{R}^n$. Let $\pi \in \mathrm{S}_{\mathrm{dec}}$, then one linear functional of $\overline{Z}$ is a geometrically weighted infinite linear combination of a *single* functional of $\overline{Y}$. The identity forces λ–decomposability of $v^\top \overline{Z} - b$ which is a special non-generic property [27, 32, 37]. A simple way of seeing that λ-decomposability for a random variable U with characteristic function ϕ_U is easily violated is noting that the defining equation $\phi_U(t) = \phi_U(\lambda t)\phi_{U_\lambda}(t)$ is unsatisfiable for many characteristic functions with zeroes (e.g. uniform distribution, atoms, etc.). Further, $\pi \in \mathrm{S}_{\mathrm{dec}}$ forces the decomposition variable to be a projection $w^\top \overline{Y}$, imposing a strong self-similar convolution equation on the joint. If $\pi \in \mathrm{S}_{\mathrm{cyc}}$, there are $v_1, v_2 \neq 0$ so that the 2-D projection $Z_{\mathrm{cyc}} = (v_1^\top Z, v_2^\top Z)$ is invariant under the finite rotation group $\{R_{2\pi m/k}\}_{m=0}^{k-1}$, imposing strong symmetry constraints. While by $\pi \in \mathrm{E}^{\mathrm{EnKU}}$ the EnKF is exact on each of these symmetry classes, the proof of the following theorem reveals that there are many other affine conditioning maps that are also exact. However, the following result shows that as soon as our distribution violates one of these symmetries, the space of possible affine filters contracts sharply. Before presenting this theorem, we uniquely fix the choice of Markov kernel for given $\pi \in \mathrm{E}^{\mathrm{EnKU}}$: let $K = \Sigma_{XY}\Sigma_{YY}^\dagger$ for the covariance matrix Σ of π and define $Z = X - KY$. Whenever we write down the Markov kernel $\pi_{X|Y=y_\star}$, we refer to the choice with law given by $Z + Ky_\star$. The "$\subseteq$" part in the proof of Proposition 2.1 demonstrates that this is indeed a valid Markov kernel for π. We present our main result for this section.

Theorem 2.4. *Consider $\pi \in \mathrm{E}^{\mathrm{EnKU}}$. Pick some $y_\star \in \mathbb{R}^m$ and assume that $\ell(x, y) = Ax + By + c$ is an exact affine map for π at $y_\star$.*

1. *If $\pi \notin \mathrm{S}_{\mathrm{cov}}$, then $\rho(A) \leq 1$ and A is diagonal in the generalized complex eigenspace of all eigenvalues with magnitude 1.*
2. *If $\pi \notin \mathrm{S}_{\mathrm{dec}}$, then the spectrum of A has no complex eigenvalues with magnitude smaller than 1 and*
$$BP_Y = -A\Sigma_{XY}\Sigma_{YY}^\dagger P_Y$$
where $P_Y = Cov(Y)Cov(Y)^\dagger$ is the orthogonal projector onto the column space of $Cov(Y)$.
3. *If $\pi \notin \mathrm{S}_{\mathrm{cyc}}$, then A has no complex eigenvalues with $|\lambda| = 1$ and $\lambda \neq 1$.*

A proof is included in the appendix. The following corollary is also shown in the appendix and says that if a distribution violates all three of these symmetries, the only possible exact affine update is the EnKU. To rule out spurious constant offsets in the constant c, we assume Σ_{YY} is invertible. This is natural: singular directions of Y carry no information and can be projected out a priori.

Corollary 2.5. *Consider $\pi \in \mathrm{E}^{\mathrm{EnKU}}$ with non-singular covariance Σ_{YY}. Pick some $y_\star \in \mathbb{R}^m$ and assume that $\ell(x, y) = Ax + By + c$ is an exact affine transport for π at $y_\star$. If $\pi \notin \mathrm{S}_{\mathrm{cov}}$, $\pi \notin \mathrm{S}_{\mathrm{dec}}$, and $\pi \notin \mathrm{S}_{\mathrm{cyc}}$, then ℓ is the EnKU:*
$$\ell(x, y) = L^{\mathrm{EnKU}}_{\pi, y_\star}(x, y).$$

This is a unique characterization result of the EnKU. As the set of symmetry-free distributions is the largest part of $\mathrm{E}^{\mathrm{EnKU}}$, this is instructive for defaulting to the EnKU to avoid bias within

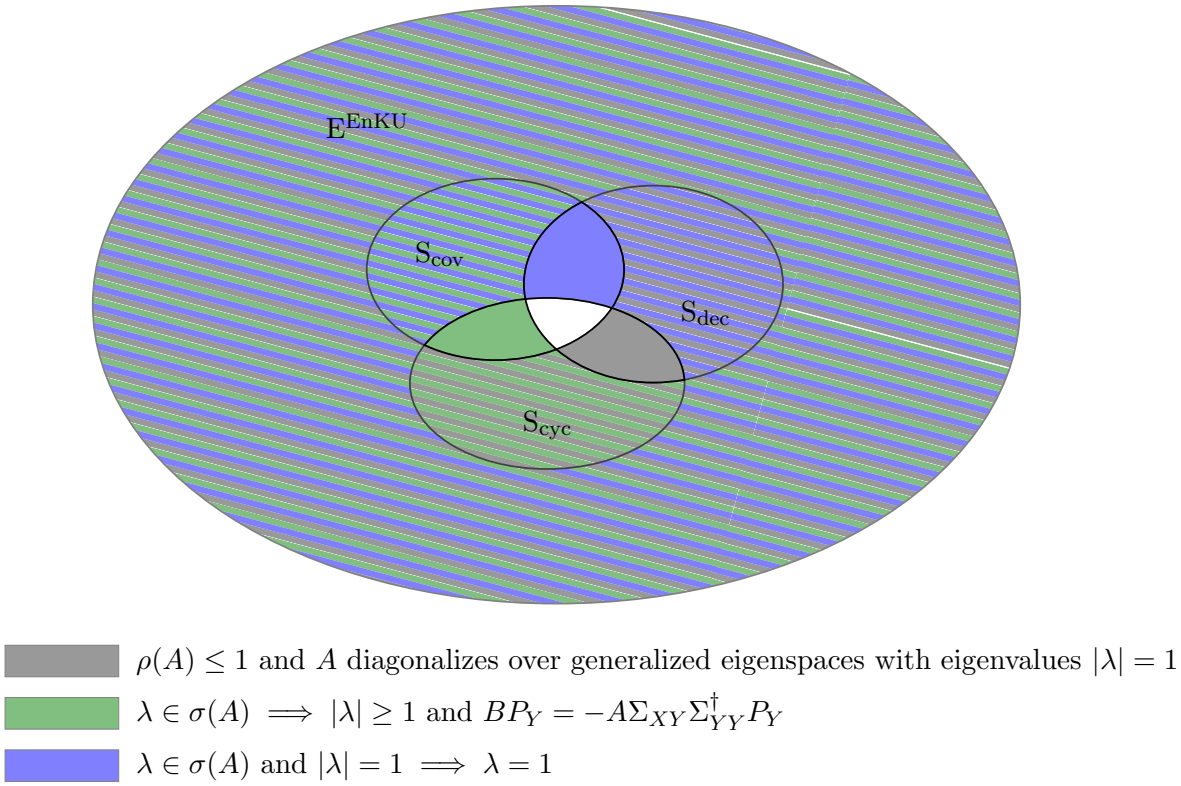

Figure 1. *Theorem 2.4 shows for any given* $\pi \in \mathrm{E}^{\mathrm{EnKU}}$ *that for any symmetry* $\mathrm{S}_{\mathrm{cov}}, \mathrm{S}_{\mathrm{dec}}, \mathrm{S}_{\mathrm{cyc}}$ *it violates, strong structural constraints are imposed on any affine conditioning map* $Ax + By + c$*. By Corollary 2.5, if it violates all these symmetries,* $Ax + By + c$ *must be the EnKU. This corresponds to the region outside* $\mathrm{S}_{\mathrm{cov}}$*,* $\mathrm{S}_{\mathrm{dec}}$*, and* $\mathrm{S}_{\mathrm{cyc}}$ *in the diagram.*

$\mathrm{E}^{\mathrm{EnKU}}$. Moreover, even if some of these symmetries hold, one would still need to identify them in order to construct an exact conditioning map—a requirement that seems inefficient in sample-constrained settings.

3. Beyond the Ensemble Kalman Update. The previous section established that, apart from a small symmetry class $\mathrm{S}_{\mathrm{cov}} \cup \mathrm{S}_{\mathrm{dec}} \cup \mathrm{S}_{\mathrm{cyc}}$, the Ensemble Kalman Update (EnKU) is the unique affine conditioning map that is exact for any element $\pi \in \mathrm{E}^{\mathrm{EnKU}}$ and observation $y_\star \in \mathbb{R}^m$. This observation raises a natural question: perhaps the restriction to $\mathrm{E}^{\mathrm{EnKU}}$ is too limiting. If one were to consider different affine conditioning maps $L_{\pi,y_\star}$, could the associated exactness set $\mathrm{E}(L_{\pi,y_\star})$ be strictly larger than $\mathrm{E}^{\mathrm{EnKU}}$? In other words, is it possible to design an update rule that is exact for a much broader class of distributions, thereby outperforming the EnKU in terms of bias reduction?

3.1. Maximal exactness of weakly $y_\star$-dependent affine conditioning maps . To investigate this possibility, we extend our analysis to the family of *weakly* $y_\star$*-dependent affine*

conditioning maps (short "weakly $y_\star$") as introduced in Equation 3.1, taking the form

$$L_{\pi,y_\star}(x,y) = A(\pi)x + B(\pi)y + c(\pi, y_\star). \tag{3.1}$$

Our restriction to this class is motivated by practice: these maps are general enough to cover most update rules practically used in high-dimensional ensemble-based data assimilation [11, 12, 29, 41]. In particular, they encompass commonly used deterministic alternatives such as square-root updates. Therefore, weakly $y_\star$-dependent affine conditioning maps provide a natural framework in which to ask whether moving beyond the EnKU can substantially enlarge the domain of exactness. Defining

$$\mathrm{F} := \bigcup_{L \text{ weakly } y_\star} \mathrm{E}(L),$$

F is the maximal exact set achievable by any single weakly $y_\star$-dependent affine update. The central result of this section is that the hoped-for enlargement beyond $\mathrm{E}^{\mathrm{EnKU}}$ is small: we show that

$$\mathrm{F} = \mathrm{E}^{\mathrm{EnKU}} \cup \mathrm{S}_{\mathrm{nl-dec}},$$

where $\mathrm{S}_{\mathrm{nl-dec}}$ is a narrow symmetry class. Thus, while alternative updates exist, they do not yield fundamentally larger exactness domains. Up to this residual symmetry class, the EnKU is already optimal among weakly $y_\star$-dependent affine conditioning maps. We give a simple necessary characterization criterion for elements of F.

Proposition 3.1. *Let $\pi \in \mathrm{F}$. Then there exists a Markov kernel $\pi_{X|Y=y}$, a measurable $d : \mathbb{R}^m \to \mathbb{R}^n$, and $\nu \in \mathcal{P}_2(\mathbb{R}^n)$ such that*

$$\pi_{X|Y=y_\star} = T_{d(y_\star)}\nu \text{ for all } y_\star \in \mathbb{R}^m$$

where we define the translation operator on measures $T_h : \mathcal{P}_2(\mathbb{R}^n) \to \mathcal{P}_2(\mathbb{R}^n)$ through $T_h\mu := (x \mapsto x+h)_\sharp\mu$ for every $h \in \mathbb{R}^n$. In particular, d is π_Y-a.s. unique up to an additive constant.

A proof can be found in the appendix. Before stating our main result of this section, we introduce the class $\mathrm{S}_{\mathrm{nl-dec}} \subseteq \mathrm{F}$.

Definition 3.2. *We define $\mathrm{S}_{\mathrm{nl-dec}} \subseteq \mathrm{F}$. Let $\pi \in \mathrm{F}$ and let (ν, d) witness Proposition 3.1. Set $Z \sim \nu$ and $Y \sim \pi_Y$ independently. Then we say $\pi \in \mathrm{S}_{\mathrm{nl-dec}}$ if and only if there exist complex vectors $v \in \mathbb{C}^n\backslash\{0\}$, $w \in \mathbb{C}^m, u \in \mathbb{C}^n$, and constants $\lambda \in \mathbb{C}$, $|\lambda| < 1$, $b \in \mathbb{C}$ such that*

$$v^\top \overline{Z} \overset{d}{=} \sum_{k=0}^{\infty} \lambda^k (w^\top Y^{(k)} + u^\top d(Y^{(k)})) + b$$

for i.i.d. copies $\{Y^{(k)}\}_{k\geq 0}$ of Y.

This class is "small" in the same sense as our earlier symmetry classes—it is defined by invariance/identities (e.g., a generalized λ–decomposition tying a one–dimensional nonlinear feature of the Y–marginal to a linear functional of Z).

Theorem 3.3. *The set of all $\pi \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^m)$ that have a weakly $y_\star$-dependent exact affine update decomposes as*

$$\mathrm{F} = \mathrm{E}^{\mathrm{EnKU}} \cup \mathrm{S}_{\mathrm{nl-dec}}.$$

The theorem is proved in the appendix and shows that weak $y_\star$–dependence leaves essentially no room to beat the EnKU: the maximal exact set collapses to $\mathrm{E}^{\mathrm{EnKU}}$ up to the narrow class $\mathrm{S}_{\mathrm{nl-dec}}$. Practically, unless one can exploit this special nonlinear decomposability, any weakly $y_\star$–dependent affine rule can do no better than the exactness domain of the EnKU.

3.2. Observation-dependent gain. The maximality result above hinges on the restriction that $A(\pi)$ and $B(\pi)$ are independent of $y_\star$. If we drop this and allow *fully $y_\star$–dependent* affine maps $L_{y_\star}(x, y) = A(y_\star)x + B(y_\star)y + c(y_\star)$, the situation changes: one can engineer many non-Gaussian π with exact affine transports that lie strictly beyond $\mathrm{E}^{\mathrm{EnKU}}$. We present the following example.

Example 3.1. *Consider any measure $r \in \mathcal{P}(\mathbb{R})$ and measurable function $f : \mathbb{R} \to \mathbb{R}$. Define π by pushing forward through $\phi : \mathbb{R}^2 \to \mathbb{R}^2, \phi(z, y) = (f(y)z, y)$:*

$$\pi_{XY} = \phi_\sharp(r \otimes r).$$

Clearly π is not in F *for general f and has the exact affine conditioning map $L_{y_\star}(x, y) = f(y_\star)y$.*

Another example is as follows.

Example 3.2. *Consider the hypercube $C = [0, 1]^2$ and any orthogonal $R \in O(2)$. Let $(X, Y) \sim \mathrm{Unif}(RC)$ be uniformly distributed. For any $y^\star$ in the support of Y there are $a(y^\star)$, $b(y^\star)$ such that*

$$X|Y = y^\star \sim \mathit{Unif}([a(y^\star), b(y^\star)]).$$

So, an exact affine conditioning map is for example

$$L_{y^\star}(x, y) = (b(y^\star) - a(y^\star))e_1^\top R^\top (x, y) + a(y^\star).$$

This perspective aligns with recent "learned ensemble filters" [5, 28, 34], where the analysis maps are chosen as $L_{\pi,y_\star}(x, y) = x + B(\pi, y_\star)y + c(\pi, y_\star)$ with the gain terms B and c parameterized by a neural network in an observation–dependent manner. In that sense, our negative result in Theorem 3.3 for weakly $y_\star$–dependent maps helps understand why learned methods pursue $y_\star$–dependent updates: without such dependence, there is essentially no headroom beyond EnKU, whereas allowing dependencies of $B(y_\star)$ on $y_\star$ could potentially realize exact updates for broader constructions. An interesting direction is to understand the enlargement of the exactness class when A and B are allowed to depend on $y_\star$, compared to F.

4. Numerical Experiments. We empirically illustrate our main claim: in the mean–field limit and within affine conditioning maps, the EnKU is the only method that remains exact beyond highly symmetric laws (such as Gaussians).

4.1. Examples. To expose the finite–sample implications, we simulate several affine updates while increasing the ensemble size N. We pick three joint laws $\pi \in \mathrm{E}^{\mathrm{EnKU}}$ in dimension $n = m = 2$ with witnessing $\nu \in \mathcal{P}_2(\mathbb{R}^n)$ and linear map $O(x, y) = x + y$ as defined in Equation 2.1 . Thus π is fully defined by the marginal choices for $Z \sim \nu$ and $Y \sim \pi_Y$ (listed below), while preserving the linear coupling that places each example in $\mathrm{E}^{\mathrm{EnKU}}$. We test the following three examples for the joint (X, Y).

- Experiment 1: Gaussian. As a sanity check, we consider the standard linear-Gaussian that most ensemble filters are derived from, namely

$$Z \sim \mathcal{N}(\mu_Z, \Sigma_Z), \qquad Y \sim \mathcal{N}(\mu_Y, \Sigma_Y).$$

  As mentioned in the introduction, infinitely many affine transports result in exact conditioning for Gaussians in the mean-field. We use

$$\mu_Z = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Sigma_Z = \begin{pmatrix} 10 & -2.5 \\ -2.5 & 1 \end{pmatrix}$$
$$\mu_Y = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Sigma_Y = \begin{pmatrix} 1 & 1.5 \\ 1.5 & 5 \end{pmatrix}.$$

- Experiment 2: Gaussian mixtures. This is an example of a measure that is in the set $\mathrm{E}^{\mathrm{EnKU}}$ but strongly multimodal and non-Gaussian:

$$Z \sim \sum_{k=1}^{6} w_k^{(Z)} \mathcal{N}(\mu_k^{(Z)}, \Sigma_k^{(Z)}), \qquad Y \sim \sum_{\ell=1}^{6} w_\ell^{(Y)} \mathcal{N}(\mu_\ell^{(Y)}, \Sigma_\ell^{(Y)}).$$

  The parameters w_ℓ, μ_ℓ, and Σ_ℓ are randomly and independently drawn from

$$\begin{aligned} w^{(Z)}, w^{(Y)} &\sim \mathrm{Dir}(\mathbf{1}_6) && \\ \mu_k^{(Z)}, \mu_k^{(Y)} &\sim \mathcal{N}(0, 36) && \text{for all } k = 1, \dots, 6 \\ \Sigma_k^{(Z)}, \Sigma_k^{(Y)} &\sim \mathcal{C} && \text{for all } k = 1, \dots, 6 \end{aligned}$$

  with Dir defined as the Dirichlet distribution, $\mathbf{1}_6$ the vector of 6 ones, and $\mathcal{C}$ defined as the law of the matrix M in

$$F \in \mathbb{R}^{2\times 2}, (F)_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1), \quad s \in \mathbb{R}^2, s_i \overset{i.i.d.}{\sim} \mathrm{Unif}(0.2, 1.5)$$
$$M = F\,\mathrm{diag}(s)\,F^\top + 10^{-6} \cdot I_2.$$

- Experiment 3: Ring density. We consider another example for a strongly non-Gaussian distribution that is in $\mathrm{E}^{\mathrm{EnKU}}$. Consider $K = 3$ rings and $M = 6$ angular modes. Spread out the radii $\ell_r, r = 1, ..., K$ uniformly between $\ell_1 = 1.4$ and $\ell_K = 4.0$. Consider an independently uniformly distributed ring mode $r \sim \mathrm{Unif}(\{1, \dots, K\})$ and angular mode $j \sim \mathrm{Unif}(\{1, \dots, M\})$ with centers $\mu_j = \frac{2\pi j}{M}$. Conditioning on (r, j), let

$$\theta \mid j \sim \mathrm{vM}(\mu_j, \kappa), \quad \kappa = 25, \quad \rho \mid r \sim \mathcal{N}(\ell_r, \sigma^2)$$

with vM the von Mises distribution and $\sigma = 0.2$. This defines Z through polar parametrization

$$Z = \begin{pmatrix} \rho \cos\theta \\ \rho \sin\theta \end{pmatrix}.$$

For Y, we consider a Gaussian mixture with 6 components sampled in the same manner as in Experiment 2.

We condition on the fixed observation $y_\star = (0.4, -0.2)^\top$ and compare several affine conditioning maps as the ensemble size N increases, reporting the W_2-distance between the analysis ensemble and the true posterior. Experiment 1 is a first simple test case and any affine method that matches second moments is exact in the mean-field. Therefore, we expect parametric error decay in N for any such method. Experiments 2 and 3 go beyond simply moment matching and feature highly non-Gaussian distributions that are contained in $\mathrm{E}^{\mathrm{EnKU}}$. Since the EnKU is exact for these distributions in the mean field, we expect its error to decrease with N to 0 at a parametric rate. Alternative affine maps, on the other hand, that are not mean–field exact for the given joint should plateau at a nonzero bias floor once the mean-field governs the error behavior.

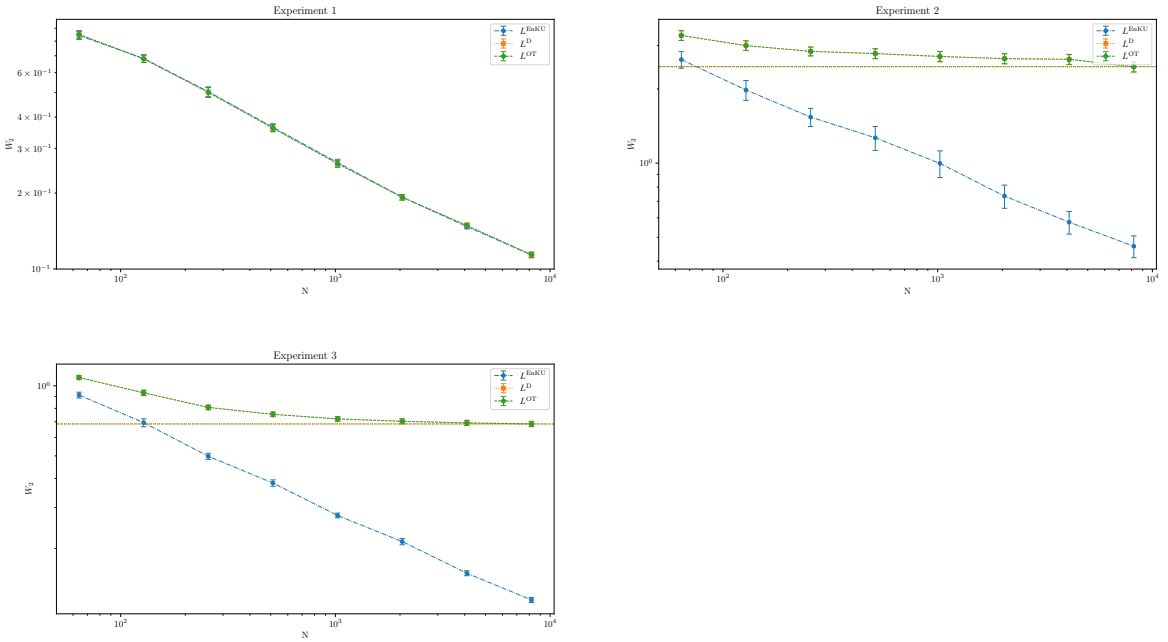

Figure 2. *Convergence of affine updates with ensemble size. Log–log W_2 error versus ensemble size N for the three data-generating models. Experiment 1 (Gaussian): all Gaussian–exact affine maps exhibit decreasing error with N (no bias floor). Experiments 2–3 (non-Gaussian): EnKU continues to improve with N, whereas the alternative affine maps plateau at a nonzero bias floor (dashed horizontal guides), indicating mean–field bias under non-Gaussian structure. Error bars show mean ± standard error over Monte Carlo replicates.*

4.2. Methods Compared. In finite samples, we consider the likelihood-free Bayesian inversion task:

$$\text{given i.i.d. } \{(X_i, Y_i)\}_{i=1}^N \sim \pi \text{ compute } \{Z_i\}_{i=1}^N \text{ such that } \frac{1}{N}\sum_{i=1}^N \delta_{Z_i} \approx \pi_{X|Y=y_\star}.$$

We will compare the EnKU with Kalman gain $\hat{K} = \hat{\Sigma}_{XY}\hat{\Sigma}_{YY}^\dagger$ estimated from the sample covariances $\hat{\Sigma}$ to two other affine updates used in likelihood-free Bayesian inversion. First, we will compare to the non-stochastic (meaning independent of y) square-root choice

$$L_{y_\star}^{\mathrm{D}}(x, y) = \sqrt{\hat{\Sigma}_{X|Y}}\hat{\Sigma}_X^{\dagger/2}(x - \hat{m}_X) + \hat{K}(y_\star - \hat{m}_Y) + \hat{m}_X$$

that is for example introduced in [7]. $\hat{m}_Y$ ($\hat{m}_X$) is the sample mean of Y_i (X_i) and $\hat{\Sigma}_{X|Y} := \hat{\Sigma}_X - \hat{\Sigma}_{XY}\hat{\Sigma}_Y^\dagger\hat{\Sigma}_{YX}$. All square-roots in the equation above are principal choices and we define

$$M^{\dagger/2} := \sqrt{M^\dagger}$$

for positive semidefinite square matrices M. Second, we compare to another non-stochastic affine transport given by the optimal transport solution

$$L_{y_\star}^{\mathrm{OT}}(x, y) = \hat{\Sigma}_X^{\dagger/2}\left(\sqrt{\hat{\Sigma}_X}\hat{\Sigma}_{X|Y}\sqrt{\hat{\Sigma}_X}\right)^{1/2}\hat{\Sigma}_X^{\dagger/2}(x - \hat{m}_X) + \hat{K}(y_\star - \hat{m}_Y) + \hat{m}_X.$$

The choices $L_{y_\star}^{\mathrm{D}}$ and $L_{y_\star}^{\mathrm{OT}}$ implement particular versions of Ensemble Square Root Filters (more specifically, Ensemble Adjustment Kalman Filters) [6, 20, 41, 46]. This can be seen by a straightforward calculation of the mean and covariance. A fuller derivation and connections to the EAKF and Ensemble Transform Kalman Filter (ETKF) are explained in the appendix section A.1. In particular, each of these affine maps is exact for Gaussian laws under a mean-field approximation.

4.3. Results. We run affine ensemble algorithms at increasing ensemble sizes, investigating the W_2 -error of their predicted posterior compared to the true posterior. For each ensemble size N we estimate the empirical W_2 between the predicted analysis ensembles $\{x_i\}_{i=1}^N$ and i.i.d. samples from the ground–truth posterior $\{x_i^{\mathrm{true}}\}_{i=1}^{6N}$ using POT's `ot.emd2` algorithm. We plot W_2 vs. N (log–log) with mean ± standard error over 30 experiment repetitions with independent randomness in Figure 2. The results match the mean–field predictions. In Experiment 1 (Gaussian), all Gaussian–exact affine maps show error decreasing with N and no bias floor. For measures in $\mathrm{E}^{\mathrm{EnKU}}$ that are non-Gaussian (Experiments 2–3), the EnKU continues to improve as N grows, while the alternative affine maps stabilize at a nonzero error, revealing a mean–field bias floor. The posterior density plots in Figure 3 corroborate this: EnKU reproduces the multimodal and ring-like posterior structure, whereas the other affine updates smear or collapse features, consistent with their moment-matching but distributionally biased behavior.

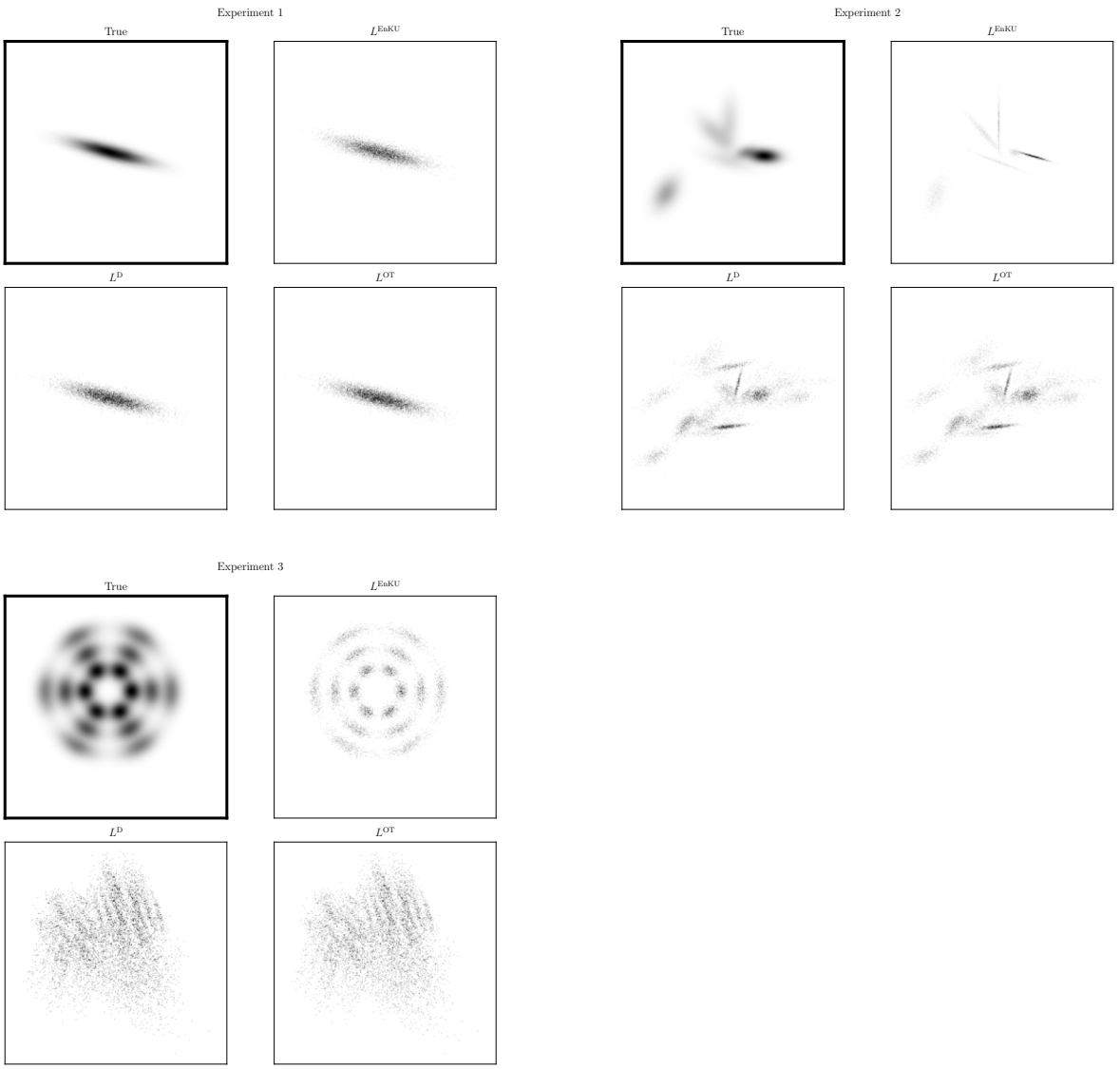

Figure 3. *Posterior structure recovered by each method (largest N). For each experiment, we show the true posterior (left/top panels) alongside analysis ensembles produced by EnKU, the deterministic map L^{D}, and the OT map L^{OT}. In the Gaussian case (Exp. 1), all methods match the target shape. In the non-Gaussian cases (Exp. 2–3), the EnKU best preserves multimodality and ring structure, while L^{D} and L^{OT} blur or collapse features—visual evidence of the bias floor quantified in the W_2 plots.*

5. Discussion. Our maximality result for weakly $y_\star$–dependent affine maps shows that there is essentially no headroom beyond the EnKF Update (EnKU): the largest possible exactness set F collapses to $\mathrm{E}^{\mathrm{EnKU}}$ up to the narrow symmetry class $\mathrm{S}_{\mathrm{nl-dec}}$ (Theorem 3.3). Further, we showed that within $\mathrm{E}^{\mathrm{EnKU}}$, the EnKU is the unique affine exact conditioning map up to small symmetry classes $\mathrm{S}_{\mathrm{cov}}, \mathrm{S}_{\mathrm{dec}}$, and $\mathrm{S}_{\mathrm{cyc}}$ (Theorem 2.4). Many questions remain open. Importantly, our analysis is mean-field and does not model many practical effects—localization, covariance inflation, finite-N sampling error, model error/mis-specification, and adaptive tun-

ing—which are known to strongly impact performance. Further, in practical data assimilation and inverse problem questions, the true joint rarely lies in $\mathrm{E}^{\mathrm{EnKU}}$ and deviates even further from Gaussianity. Regardless, affine filters are applied in these settings. Therefore, another lens to study the question of choosing affine filters is the aspect of bias–variance tradeoff. Affine filters are usually used in high dimensions where the dimension is large compared to the ensemble size, which is non-i.i.d. after one filtering step. For these two reasons, accepting bias in the estimator to reduce variance is inevitable. Quantifying this tradeoff in nonlinear settings remains an important direction. A related open question is treating the corresponding multi-step behavior of the EnKU (e.g., EKI), its exactness, and how nonlinear effects re-enter through evolving covariances [21, 38, 39].

Appendix A. Appendix. In this appendix, we provide the proofs of the theorems stated in the paper and clarify how the affine transports used in the numerical experiments relate to square-root filters.

A.1. Connection to Ensemble Square Root Filters. Ensemble Square Root Filters (ESRFs) are deterministic variants of the Ensemble Kalman Filter that update the ensemble without requiring perturbed observations, typically improving stability and accuracy [6,20,41,46]. They are usually derived in settings where we have access to i.i.d. samples $\{X_i\}_{i=1}^N$ ("forecast") and we have the dependency $Y = HX + \xi$ with linear H, independent mean-zero ξ, and $\mathrm{Cov}(\xi) = \Gamma$ finite. Defining the forecast matrix $\hat{X}_f := (X_1 \dots X_N) \in \mathbb{R}^{n\times N}$ with n the state dimension and the forecast covariance $\hat{C}_f := \frac{1}{N-1}\hat{X}_f \left(I_N - \frac{1}{N}\mathbf{1}\mathbf{1}^\top\right) \hat{X}_f^\top$ where $\mathbf{1}$ is the vector with all entries 1. The main idea in ESRFs is to find an affine map

$$s : \mathbb{R}^{n\times N} \to \mathbb{R}^{n\times N}$$

such that with $\hat{X}_a := s(\hat{X}_f)$ we have the following Gaussian-consistent moment conditions:

$$\begin{aligned}\hat{m}_a &= \hat{m}_f + K\left(y_\star - H\hat{m}_f\right)\\ \hat{C}_a &= C_a\end{aligned}$$

where

$$\hat{m}_f := \frac{1}{N}\hat{X}_f\mathbf{1}, \quad \hat{m}_a = \frac{1}{N}\hat{X}_a\mathbf{1},$$
$$\hat{C}_a := \frac{1}{N-1}\hat{X}_a\left(I_N - \frac{1}{N}\mathbf{1}\mathbf{1}^\top\right)\hat{X}_a^\top, \quad C_a = \hat{C}_f - \hat{C}_f H^\top (H\hat{C}_f H^\top + \Gamma)^{-1} H\hat{C}_f.$$

The prediction for the posterior $\pi_{X|Y=y_\star}$ in an ESRF is then

$$\hat{\pi}_{X|Y=y_\star} = \frac{1}{N}\sum_{i=1}^N \delta_{\hat{X}_a^i}$$

for $\hat{X}_a^i$ the columns of $\hat{X}_a$. There are multiple versions of ESRFs as the choice of s is not unique. The most important versions of the ESRF are the Ensemble Transform Kalman Filter (ETKF) [6,41] and the Ensemble Adjustment Kalman Filter (EAKF):

1. The ETKF is defined by requiring s to operate on the anomaly matrix $\hat{X}_f^{(c)} = \hat{X}_f\left(I_N - \frac{1}{N}\mathbf{1}\mathbf{1}^\top\right)$ in ensemble space

$$s(\hat{X}_f) = \hat{X}_f^{(c)}\hat{T} + \hat{b}\mathbf{1}^\top$$

where $\hat{b} \in \mathbb{R}^n$ is a bias term uniquely determined by the first-order condition [6,20] . $\hat{T} \in \mathbb{R}^{N\times N}$ is therefore a matrix satisfying the second-moment condition

$$\hat{T}\hat{T}^\top = I_N - \left(\hat{X}_f^{(c)}\right)^\top H^\top \left(H\hat{C}_f H^\top + \Gamma\right)^{-1} H\hat{X}_f^{(c)}.$$

The unique principal square root of the right-hand side has been shown to be particularly stable. It is usually chosen for $\hat{T}$ [20, 29, 35, 44]. This is unsurprising since it is the choice that is the "least transformative", i.e. $\sqrt{M} = \arg\min_{\tilde{M}\tilde{M}^\top = M} \left\| \tilde{M} - I \right\|_F$ for $\sqrt{\cdot}$ the principal square root, M positive semidefinite, and $\|\cdot\|_F$ the Frobenius-norm. Therefore, we let

$$\hat{T} = \sqrt{I_N - \left(\hat{X}_f^{(c)}\right)^\top H^\top \left(H\hat{C}_f H^\top + \Gamma\right)^{-1} H\hat{X}_f^{(c)}}$$

be the principal square root.

2. The EAKF, on the other hand, acts on the rows of the anomaly matrix [2, 41], meaning that

$$s(\hat{X}_f) = \hat{A}\hat{X}_f^{(c)} + \hat{b}\mathbf{1}^\top.$$

$\hat{A} \in \mathbb{R}^{n\times n}$ is therefore a matrix satisfying

$$\hat{A}\hat{C}_f\hat{A}^\top = \hat{C}_a.$$

The symmetric solution for this equation is given by

$$\hat{A}^{(1)} = \hat{C}_f^{\dagger/2}\left(\sqrt{\hat{C}_f}\hat{C}_a\sqrt{\hat{C}_f}\right)^{1/2}\hat{C}_f^{\dagger/2}$$

with all square roots as the principal choice. Another possible choice is

$$\hat{A}^{(2)} = \sqrt{\hat{C}_a}\hat{C}_f^{\dagger/2}.$$

In practice, the following choice of square root is used more frequently instead [2, 15, 41]: let

$$\hat{A}^{(3)} = \hat{X}_f^{(c)}\, C\, (I + D)^{-1/2}\, G^\dagger F^T,$$

where $\hat{X}_f^{(c)} = FGU^T$ is the SVD and $(\hat{X}_f^{(c)})^\top H^\top \Gamma^{-1} H\hat{X}_f^{(c)} = CDC^T$ is the eigenvalue decomposition with the eigenvectors in the null space arranged as the final columns of C.

As we do not make the linear assumption $Y = HX + \xi$ in our paper, we need to translate the expressions for $\hat{T}$ and $\hat{A}$ to this more general setting. Doing this for the EAKF is immediate. We simply replace the estimated analysis covariance with its population counterpart:

$$C_a' \;=\; \hat{C}_f \;-\; \frac{1}{N-1}\hat{X}_f^{(c)}\left(\hat{Y}_f^{(c)}\right)^\top \left(\hat{Y}_f^{(c)}\left(\hat{Y}_f^{(c)}\right)^\top\right)^\dagger \hat{Y}_f^{(c)}\left(\hat{X}_f^{(c)}\right)^\top.$$

This shows directly that $L_{y_\star}^{\mathrm{OT}}$ and $L_{y_\star}^{\mathrm{D}}$ implement the EAKF updates $\hat{A}^{(1)}$ and $\hat{A}^{(2)}$. For the ETKF, the idea is similar. Starting with $\hat{T}$, we note that the expression containing $H\hat{X}_f^{(c)}$ or Γ involves prior knowledge of the covariance structure, namely the assumption $Y = HX + \xi$. The generalization of $\hat{T}$ to non-linear settings is therefore

$$\hat{T}' = \sqrt{I_N - \left(\hat{Y}_f^{(c)}\right)^\top \left(\hat{Y}_f^{(c)}\left(\hat{Y}_f^{(c)}\right)^\top\right)^\dagger \hat{Y}_f^{(c)}}$$

with $\hat{Y}_f^{(c)} \in \mathbb{R}^{m\times N}$ the centered ensemble matrix of the observations Y_i. As the second term in $\hat{T}'$ is a projection onto the row space of $\hat{Y}_f^{(c)}$, $\hat{T}'$ is the principal square root of a projector onto the orthogonal complement of the row space of $\hat{Y}_f^{(c)}$. Orthogonal projectors are their own principal square roots and therefore

$$\hat{T}' = I_N - \left(\hat{Y}_f^{(c)}\right)^\top \left(\hat{Y}_f^{(c)} \left(\hat{Y}_f^{(c)}\right)^\top\right)^\dagger \hat{Y}_f^{(c)}.$$

However, now note that

$$\begin{aligned}\hat{X}_f^{(c)}\hat{T}' &= \hat{X}_f^{(c)} - \hat{X}_f^{(c)} \left(\hat{Y}_f^{(c)}\right)^\top \left(\hat{Y}_f^{(c)} \left(\hat{Y}_f^{(c)}\right)^\top\right)^\dagger \hat{Y}_f^{(c)} \\ &= \hat{X}_f^{(c)} - \hat{\Sigma}_{XY}\hat{\Sigma}_{YY}^\dagger \hat{Y}_f^{(c)}.\end{aligned}$$

This shows that the generalized ETKF and the EnKU perform the same update. In that sense, everything we say above about the EnKU, applies to the ETKF as they are the same outside the linear-Gaussian setting.

Remark A.1. The presented "generalizations" of the ESRF are not meant to be good filtering methods. In fact, they forfeit the main advantage of ESRFs, namely the deterministic (non-stochastic) update. The point of introducing them above lies instead in providing insight into the bias inherent in ESRF methods and clarifying their connection to the EnKU.

A.2. Proofs. We start by presenting a proof of Proposition 2.1

Proof of Proposition 2.1. $\subseteq$: Pick $\pi \in \mathrm{E}^{\mathrm{EnKU}}$, meaning that

$$\left(L^{\mathrm{EnKU}}_{\pi,y_\star}\right)_\sharp \pi = \pi_{X|Y=y_\star}$$

π_Y-a.s. in $y_\star \in \mathbb{R}^m$. Letting $(X,Y) \sim \pi$ and defining $Z = X - KY$, the equation above is equivalent to

$$\mathrm{Law}\left(Z + Ky_\star\right) = \pi_{X|Y=y_\star}.$$

Since Z does not depend on $y_\star$, this shows that for $\nu = \mathrm{Law}(Z)$ and $O(x,y) = x + Ky$ we have

$$O(\cdot, y_\star)_\sharp \nu = \pi_{X|Y=y_\star}$$

π_Y-a.s. in $y_\star$. $O(\cdot, y_\star)_\sharp\nu$ is a Markov kernel, concluding this direction.
$\supseteq$: Consider π and its corresponding O and ν as in the right-hand side of the equation we prove in this proposition. Write $O(x,y) = A_1x + A_2y$ for matrices $A_1 \in \mathbb{R}^{n\times n}$, $A_2 \in \mathbb{R}^{n\times m}$, and let $Z \sim \nu$. Then $\pi_{X|Y=y} = \mathrm{Law}(A_1Z + A_2y)$. Let $Y \sim \pi_Y$, independent of Z, so that $(X,Y) := (A_1Z + A_2Y, Y) \sim \pi$. By direct computation,

$$\mathrm{Cov}(\pi)_{XY} = \mathrm{Cov}(A_1Z + A_2Y, Y) = A_2\mathrm{Cov}(Y) = A_2\mathrm{Cov}(\pi)_{YY}.$$

Therefore, $L^{\mathrm{EnKU}}_{\pi,y^\star}(x,y) = x + A_2\mathrm{Cov}(\pi)_{YY}\left(\mathrm{Cov}(\pi)_{YY}\right)^\dagger (y^\star - y)$. Let $\tilde{Y} \sim \pi_Y$, independent of (Z,Y). Define the projection $P_Y := \mathrm{Cov}(\pi)_{YY}\left(\mathrm{Cov}(\pi)_{YY}\right)^\dagger$, the orthogonal projection

onto $\mathrm{Im}(\mathrm{Cov}(\pi_Y))$. Then $L^{\mathrm{EnKU}}_{\tilde{Y}}(A_1Z + A_2Y, Y) = A_1Z + A_2Y + A_2P_Y(\tilde{Y} - Y)$. Because $Y - \tilde{Y} \in \mathrm{Im}(\mathrm{Cov}(\pi_Y))$ a.s., this a.s. simplifies to $L^{\mathrm{EnKU}}_{\tilde{Y}}(A_1Z + A_2Y, Y) = A_1Z + A_2\tilde{Y}$. Thus,

$$\mathrm{Law}\left(L^{\mathrm{EnKU}}_{\tilde{Y}}(X, Y)\,\middle|\,\tilde{Y}\right) = \mathrm{Law}(A_1Z + A_2\tilde{Y}) = \pi_{X|Y=\tilde{Y}}$$

concluding the proof. ∎

Similarly, we can show Proposition 3.1.

Proof of Proposition 3.1. Let $\pi \in \mathrm{F}$. Then there exists a weakly $y_\star$-dependent affine map

$$L_{\pi,y_\star}(x, y) = A(\pi)x + B(\pi)y + c(\pi, y_\star)$$

and a Markov kernel such that there is a Borel set $Q \in \mathcal{B}(\mathbb{R}^m)$ with $\pi_Y(Q) = 1$ and

$$(L_{\pi,y_\star})_\sharp \pi = \pi_{X|Y=y_\star}$$

for all $y_\star \in Q$. Since A and B do not depend on $y_\star$, for any $y_0, y_\star \in Q$ we have

$$\pi_{X|Y=y_\star} = T_{c(\pi,y_\star)-c(\pi,y_0)}\,\nu$$

where we set $\nu := \pi_{X|Y=y_0}$. Now, we construct a measurable $d(\cdot)$ such that

$$d(y_\star) = c(\pi, y_\star) - c(\pi, y_0)$$

for all $y_\star \in Q$ and note that this concludes the proof. Define $d : \mathbb{R}^m \to \mathbb{R}^n$ through

$$d(y_\star) = \begin{cases} c(\pi, y_\star) - c(\pi, y_0), & y_\star \in Q, \\ 0, & y_\star \notin Q. \end{cases}$$

For any Borel set $W \in \mathcal{B}(\mathbb{R}^n)$ we have

$$d^{-1}(W) = (Q \cap d^{-1}(W)) \cup (Q^{\mathrm{c}} \text{ if } 0 \in W) = d^{-1}_{|Q}(W) \cup (Q^{\mathrm{c}} \text{ if } 0 \in W)$$

meaning that all we have to show is that the restriction $d_{|Q}$ is measurable. Consider the translation map

$$\Phi : \mathbb{R}^n \to \mathcal{P}_2(\mathbb{R}^n), \qquad \Phi(h) := T_h\nu$$

where $\mathcal{P}_2(\mathbb{R}^n)$ is endowed with the Wasserstein-topology. The map Φ is continuous and injective; by the Lusin–Souslin Theorem [9, Lemma 8.3.8] and since $\mathbb{R}^n$ and $\mathcal{P}_2(\mathbb{R}^n)$ are Polish spaces, the inverse on its image $\mathcal{O} = \Phi(\mathbb{R}^n)$, namely

$$\Psi : \mathcal{O} \to \mathbb{R}^n, \qquad \Psi(T_h\nu) = h,$$

is measurable with respect to the Borel algebra induced by the subspace topology of $\mathcal{O}$. By the first part of the proof in [1, Lemma 12.4.7], the map $y \mapsto \pi_{X|Y=y}$ is $\mathcal{B}(\mathbb{R}^m)$-to-Borel$(\mathcal{P}_2(\mathbb{R}^n))$ measurable; hence its restriction $Q \to \mathcal{P}_2(\mathbb{R}^n)$ is $(Q, \mathcal{B}(Q))$-measurable. We established that on Q we have $\pi_{X|Y=y} \in \mathcal{O}$ and thus we have that

$$d_{|Q}(y) = \Psi\big(\pi_{X|Y=y}\big), \qquad y \in Q.$$

and $d_{|Q}$ is measurable as a composition of measurable maps $y \mapsto \pi_{X|Y=y} \mapsto \Psi(\pi_{X|Y=y})$. $T_{d(y_\star)}\nu$ is a valid choice of Markov kernel by measurability of d. Further, d is π_Y-a.s. unique by π_Y-a.s. uniqueness of Markov kernels. ∎

The following theorem, while not explicitly stated in the paper, is the main theoretical basis for the remaining results presented in this paper.

Theorem A.2. *Let $A \in \mathbb{R}^{n\times n}$ and let U be an $\mathbb{R}^n$-valued random vector with $\mathbb{E}\|U\|^2 < \infty$. Assume $X \in \mathbb{R}^n$ is independent of U. Consider the fixed-point-in-law equation*

$$X \overset{d}{=} AX + U. \tag{A.1}$$

By the real Jordan decomposition, there exist A-invariant subspaces such that

$$\mathbb{R}^n \;=\; V_s \oplus V_r \oplus V_u$$

and for all restrictions $A_\bullet := A_{|V_\bullet}, \bullet \in \{u, s, r\}$

1. *all complex eigenvalues of A_s have magnitude less than 1*
2. *all complex eigenvalues of A_r have magnitude equal to 1*
3. *all complex eigenvalues of A_u have magnitude larger than 1.*

Further, decompose the complexification $V_r^{\mathbb{C}} \subseteq \mathbb{C}^n$

$$V_r^{\mathbb{C}} = V_r^{(1)} \oplus V_r^{(2)}$$

with $V_r^{(1)}$ the space of all eigenvectors of A_r with eigenvalues $|\lambda| = 1$. Denote by $P_\bullet$ the corresponding projections and write $X_\bullet := P_\bullet X$, $U_\bullet := P_\bullet U$. There exists a solution X with $\mathbb{E}\|X\|^2 < \infty$ to Equation A.1 if and only if U_u and U_r are a.s. constant vectors and $U_r \in Im(I - A_r)$ a.s. The blockwise solutions, if they exist, satisfy:

(a) *There is a* unique *solution in law in the stable component given by*

$$X_s \overset{d}{=} \sum_{k=0}^{\infty} A_s^k U_s^{(k)},$$

where $\{U_s^{(k)}\}_{k\geq 0}$ are i.i.d. copies of U_s, independent of each other; the series converges in L^2.

(b) *$X_r^{(2)}$ is a.s. constant.*

(c) *X_u is a.s. constant with the a.s. value*

$$X_u \;=\; (I - A_u)^{-1} U_u.$$

Before presenting a proof, we need to show a few lemmas.

Lemma A.3. *Consider a matrix $B \in \mathbb{R}^{d\times d}$ with $\rho(B) < 1$. Then there is a norm $\|\cdot\|$ on $\mathbb{R}^d$ such that the operator norm satisfies $\|B\| < 1$.*

Proof. The discrete Lyapunov equation

$$B^\top P B - P \;=\; -I$$

has a unique positive-definite solution $P \succ 0$ [31]. Define the (equivalent) norm $\|x\|_P := (x^\top P x)^{1/2}$. Then

$$\|Bx\|_P^2 = x^\top B^\top P B x = x^\top (P - I) x = \|x\|_P^2 - \|x\|_I^2 \leq \left(1 - \tfrac{1}{\lambda_{\max}(P)}\right)\|x\|_P^2.$$

Note that $\|x\|_P^2 = x^\top P x = x^\top B^\top P B x + \|x\|_I^2 \geq \|x\|_I^2$ implies that $\lambda_{\max}(P) \geq 1$. Hence $\|B\|_P \leq \sqrt{1 - 1/\lambda_{\max}(P)} =: q < 1$ as claimed. ∎

Lemma A.4. *Let $J \in \mathbb{R}^{n\times n}$ be a Jordan block for an eigenvalue $|\lambda| = 1$. Let $Q \succeq 0$. If a symmetric $P \succeq 0$ satisfies the discrete Lyapunov (Stein) equation*

$$P \;=\; JPJ^* + Q,$$

then necessarily $Q = 0$. Further, all entries of P but P_{11} must be zero.
If instead $|\lambda| > 1$, there can only be a solution if $P = Q = 0$.

Proof. Consider the case $|\lambda| = 1$ first. If $n = 1$ there is nothing to show so we can assume $n > 1$. Say $P \succeq 0$ is a matrix satisfying the Lyapunov equation. Write $J = \lambda I + N$. Translating the Lyapunov equation $P = (\lambda I + N)P(\bar{\lambda} I + N^T) + Q$ into components, writing p_{ij} and q_{ij} for the indices of P and Q yields

$$\lambda p_{i,j+1} + \bar{\lambda} p_{i+1,j} + p_{i+1,j+1} + q_{ij} = 0$$

for all i, j with $p_{ab} = 0$ if an index exceeds n. We proceed by induction over $n + 1 \geq m > 1$ with the hypothesis that $q_{m,m} = p_{m,m} = 0$. Our inductive base is $m = n + 1$ for which there is nothing to show. Let $m > 1$ and assume that $p_{m+1,m+1} = q_{m+1,m+1} = 0$. Then by Cauchy-Schwarz also $p_{ij} = q_{ij} = 0$ if either i or j is $m + 1$. The $(i, j) = (m, m)$ equation tells us that $q_{m,m} = 0$. The $(i, j) = (m, m - 1)$ equation says $\lambda p_{m,m} + q_{m,m-1} = 0$ and by Cauchy-Schwarz $q_{m,m-1} = 0$, showing that $p_{m,m} = q_{m,m} = 0$. This induction shows that $q_{ij} = p_{ij} = 0$ except for $i = j = 1$. Finally, the $(i, j) = (1, 1)$ equation is simply $q_{ij} = 0$, completing the proof of the first part.
Let $|\lambda| > 1$. Our strategy is to construct a unique solution to the unconstrained problem for P and show uniqueness for this solution. Consider the series

$$P \;=\; -\sum_{k=1}^{\infty} J^{-k} Q J^{-*k}.$$

Because $\rho(J^{-1}) < 1$, by Lemma A.3 the sequence $\|J^{-k}\|$ decays geometrically in some matrix norm, ensuring absolute convergence of the series. A direct computation shows that it solves the unconstrained equation

$$JPJ^* \;=\; -\sum_{k=1}^{\infty} J^{-(k-1)} Q J^{-*(k-1)} \;=\; -Q - \sum_{k=1}^{\infty} J^{-k} Q J^{-*k}.$$

Say $Q \neq 0$. Then P is negative semi-definite and non-zero contradicting positive semi-definiteness. Therefore $Q = P = 0$ for this solution. We conclude the proof by showing that this is the unique solution of the unconstrained problem. The unconstrained problem is a linear operator problem that can be vectorized

$$\Psi(X) = \mathrm{vec}(Q)$$

with $\Psi(X) = \mathrm{vec}(X) - (J^* \otimes J)\mathrm{vec}(X)$. The spectrum of $J^* \otimes J$ consists of the products $\{\bar{\lambda}_i \lambda_j\}$ where $\{\lambda_i\}$ are the eigenvalues of J. Since every $|\bar{\lambda}_i \lambda_j| > 1$, the operator Ψ has a trivial kernel and the solution is unique. ∎

Lemma A.5. *Let $r \in \mathbb{R}$. The set*

$$\{\, nr \bmod 1 : n \in \mathbb{Z} \,\}$$

is dense in $[0,1]$ if and only if r is irrational.

Proof. This is a well-known fact following from the Equidistribution Theorem [45]. We include a concise proof for completeness. If $r = p/q \in \mathbb{Q}$, then $nr \bmod 1$ takes at most q values, so the set is not dense. Conversely, assume r is irrational. Fix $m \in \mathbb{N}$. By the pigeonhole principle, among the $m+1$ distinct numbers

$$0, r, 2r, \ldots, mr \pmod 1$$

there exist distinct i, j with $0 \le i < j \le m$ such that

$$\big\|(j-i)r\big\| < \tfrac{1}{m},$$

where $\|x\| := \min_{k\in\mathbb{Z}} |x-k|$. Hence the step size $(j-i)r \bmod 1$ is within $1/m$ of 0. Therefore, integer multiples of $(j-i)r$ modulo 1 form a $1/m$-net of $[0,1]$. Since m was arbitrary, the set is dense in $[0,1]$. ∎

Lemma A.6. *Consider a random vector $X \in \mathbb{R}^2$ that is symmetric under a rotation R_θ of angle $\theta \in [0, 2\pi)$*

$$X \overset{d}{=} R_\theta X.$$

Then $\frac{\theta}{2\pi} \in \mathbb{Q}$ or X is invariant under all rotations.

Proof. Assume $\frac{\theta}{2\pi} \notin \mathbb{Q}$ and define the set $S = \left\{ \frac{\theta k}{2\pi} \bmod 1 \,|\, k \in \mathbb{N} \right\}$. S is dense in $[0,1)$ by Lemma A.5. Pick any point $s \in [0,1)$ and choose a sequence $s_k \in S$ such that $\lim_{k\to\infty} s_k = s$. Consider any $f : \mathbb{R}^2 \to \mathbb{R}$ that is bounded and continuous. By repeatedly applying invariance, we have $\mathbb{E}\left(f(R_{2\pi s_k} X)\right) = \mathbb{E}\left(f(X)\right)$ for any k. Therefore,

$$\mathbb{E}\left(f(R_{2\pi s}X)\right) = \mathbb{E}\left(\lim_{k\to\infty} f(R_{2\pi s_k}X)\right) = \lim_{k\to\infty} \mathbb{E}\left(f(R_{2\pi s_k}X)\right) = \lim_{k\to\infty} \mathbb{E}\left(f(X)\right) = \mathbb{E}\left(f(X)\right).$$

where the second equality is the Dominated Convergence Theorem. ∎

We are now in a position to prove Theorem A.2.

Proof of Theorem A.2. Say $\mathbb{E}\,\|X\|^2 < \infty$ is a solution of the fixed-point equation. We will proceed by showing that this implies that U_u and U_r are a.s. constant vectors, $U_r \in \mathrm{Im}(I - A_r)$ a.s., and X satisfies (a) – (c).

From $X \overset{d}{=} AX + U$ and subspace-invariance we can conclude the following equations:

$$X_\bullet \overset{d}{=} A_\bullet X_\bullet + U_\bullet.$$

We proceed by treating each block separately.

(a) Stable block V_s. We can choose a norm with $\|A_s\| < 1$ by Lemma A.3 since $\rho(A_s) < 1$. Define $\mathcal{T}(\mu) := (A_s)_{\#}\mu * \mathrm{Law}(U_s)$ on the metric space $(\mathcal{P}_2(V_s), W_2)$ with $*$ the convolution

of measures. Pushforward by A_s is W_2-Lipschitz with constant $\|A_s\| < 1$ and convolution is 1-Lipschitz, so $\mathcal{T}$ is a strict contraction; by Banach's fixed-point theorem, there is a unique fixed point μ_s. Now, let $m := \mathbb{E}U_s$ and write $U_s^{(k)} = \tilde{U}_s^{(k)} + m$ with i.i.d. copies $\{U_s^{(k)}\}_{k\geq -1}$ of U_s. Set

$$X_s^{\mathrm{det}} := \sum_{k=0}^{\infty} A_s^k m, \qquad X_s^{\mathrm{rnd}} := \sum_{k=0}^{\infty} A_s^k \tilde{U}_s^{(k)}.$$

Since $\|A_s\| < 1$, the Neumann series $\sum_{k\geq 0} A_s^k$ converges in operator norm and X_s^{det} is well-defined. For the random series,

$$\mathbb{E}\Big\| \sum_{k=N}^{M} A_s^k \tilde{U}_s^{(k)} \Big\|^2 = \sum_{k=N}^{M} \mathbb{E}\|A_s^k \tilde{U}_s\|^2 \leq \sum_{k=N}^{M} \|A_s^k\|^2 \, \mathbb{E}\|\tilde{U}_s^{(0)}\|^2,$$

where the cross terms vanish because the summands are independent and centered. By the geometric series $\sum_{k\geq 0} \|A_s^k\|^2 < \infty$, this shows Cauchy in L^2, hence X_s^{rnd} converges in L^2 by completeness. Defining

$$X_s \;:=\; X_s^{\mathrm{det}} + X_s^{\mathrm{rnd}},$$

by the continuous mapping theorem

$$A_s X_s + U_s^{(-1)} = \sum_{k\geq 1} A_s^k U_s^{(k-1)} + U_s^{(-1)} \;\overset{d}{=}\; \sum_{k\geq 0} A_s^k U_s^{(k)} \;=\; X_s.$$

Thus $\mathrm{Law}(X_s)$ is the unique fixed point on V_s.
(b) Rotational block V_r. Choose a complex basis $v_1, \ldots, v_{d_r}$ of the complexified space $(V_r)^{\mathbb{C}}$ with d_r its dimension and put A_r into its complex Jordan form $\mathrm{diag}(J_1, \ldots, J_{n_r}) \in \mathbb{C}^{d_r \times d_r}$ with Jordan blocks J_i. For every Jordan block, the distributional equation

$$X_r^i \overset{d}{=} J_i X_r^i + U_r^i$$

holds where X_r^i, U_r^i are the coordinates of X_r, U_r in the Jordan block J_i. Computing complex covariances yields

$$P = J_i P J_i^* + Q$$

for P and Q the complex covariance matrices of X_r^i and U_r^i. Apply the first part of Lemma A.4 to see from this that $Q = 0$ and that P_{11} is the only nonzero index of P. Note that P_{11} corresponds to the eigenvector in the Jordan chain of J_i. Applying this argument to every block shows that U_r is a.s. constant and the only potentially non-a.s.-constant part of X_r is the eigenvector component $X_r^{(1)}$. Note also by taking expectations that

$$(I - A_r)\mathbb{E}(X_r) = \mathbb{E}(U_r)$$

which means that since U_r is a.s. constant it must be a.s. in the image of $(I - A_r)$.
(c) Unstable block V_u ($\rho(A_u) > 1$). Using the same Jordan reduction argument as in (b), we arrive at the equation

$$P = JPJ^* + Q.$$

Apply the second part of Lemma A.4 to conclude that U_u and X_u are a.s. constant. The distributional equation becomes an a.s. equation and we have that a.s.

$$X_u = (I - A_u)^{-1} U_u.$$

Now, conversely, say that U_u and U_r are a.s. constant vectors with $U_r \in \mathrm{Im}(I - A_r)$ a.s. Construct the solution blockwise and make the blocks statistically independent so that blockwise satisfaction of the distributional equation is sufficient. In the stable and unstable blocks choose the solution as described in the theorem statement. Finally, choose X_r constant such that it solves the linear equation

$$(I - A_r)X_r = U_r$$

a.s. It is clear that this is a valid solution from our previous argument, completing the proof. ■

Using Theorem A.2, we can prove Theorem 2.4.

Proof of Theorem 2.4. $\pi \in \mathrm{E}^{\mathrm{EnKU}}$ means that there is a measure $\nu \in \mathcal{P}_2(\mathbb{R}^n)$ such that for $Z \sim \nu$ independent of $Y \sim \pi_Y$, $(X, Y) = (Z + MY, Y) \sim \pi$ for $M = \Sigma_{XY}\Sigma_{YY}^{\dagger}$. Fix $y_\star \in \mathbb{R}^m$ and an exact affine map $\ell(x, y) = Ax + By + c$. This means that

$$AX + BY + c \overset{d}{=} Z + My_\star$$

which can be rewritten as

$$A\overline{Z} + (AM + B)\overline{Y} + ((A - I)\mathbb{E}(Z) + (AM + B)\mathbb{E}(Y) + c - My_\star) \overset{d}{=} \overline{Z}.$$

Defining $U = (AM + B)\overline{Y} + ((A - I)\mathbb{E}(Z) + (AM + B)\mathbb{E}(Y) + c - My_\star)$, this is equivalent to the following fixed point equation with $Z \perp U$:

$$A\overline{Z} + U \overset{d}{=} \overline{Z}.$$

(1) $\pi \notin \mathrm{S}_{\mathrm{cov}}$. Assume ν has a non-singular covariance meaning it does not have a constant linear component. By Theorem A.2 (in the notation of the theorem), $Z_r^{(2)}$ and Z_u are a.s. constant. However, as we assumed that Z has non-singular covariance, this means that the sum of generalized eigenspaces of A with $|\lambda| > 1$ is empty and that A is diagonalizable over the generalized eigenspace of all eigenvalues with magnitude 1. In particular, $\rho(A) \leq 1$.
(2) $\pi \notin \mathrm{S}_{\mathrm{dec}}$. Let $\pi \notin \mathrm{S}_{\mathrm{dec}}$ and assume that V_s is non-trivial, meaning that there is at least one complex eigenvalue λ of A with magnitude $|\lambda| < 1$. There exists a left non-zero eigenvector $p \in \mathbb{C}^n$ such that

$$p^\top A = \lambda p^\top.$$

Plugging into the fixed point equation yields the 1D fixed point equation for $p^\top \overline{Z}$

$$p^\top \overline{Z} \overset{d}{=} \lambda p^\top \overline{Z} + p^\top U.$$

By point (a) of Theorem A.2 this implies that

$$p^\top \overline{Z} \overset{d}{=} \sum_{k=0}^{\infty} \lambda^k p^\top U^{(k)}$$

for i.i.d. copies $U^{(k)}$ of U. Writing $q^\top = p^\top(AM+B)$ and using $b \in \mathbb{C}$ as a centering variable that includes the constant term of U, we can rewrite this as

$$p^\top \overline{Z} \overset{d}{=} \sum_{k=0}^{\infty} \lambda^k q^\top \overline{Y}^{(k)} + b.$$

for i.i.d. copies $Y^{(k)}$ of Y. However, this means that $\pi \in \mathrm{S}_{\mathrm{dec}}$ and so V_s must have been trivial. This implies that U is constant by Theorem A.2 and therefore $(AM+B)P_Y = 0$ where P_Y is the orthogonal projector onto the column space of $\mathrm{Cov}(Y)$. This part of the statement is finalized by recognizing that $MP_Y = \Sigma_{XY}\Sigma_{YY}^\dagger P_Y$ as shown in the proof of Proposition 2.1.
(3) $\pi \notin \mathrm{S}_{\mathrm{cyc}}$. Finally, assume that $\pi \notin \mathrm{S}_{\mathrm{cyc}}$. By projection, we have that

$$\overline{Z}_r \overset{d}{=} A_r \overline{Z}_r + U_r.$$

By Theorem A.2, U_r is a.s. constant. Taking expectations shows that since $\overline{Z}_r$ is mean-zero, U_r is a.s. 0 so that we have

$$\overline{Z}_r \overset{d}{=} A_r \overline{Z}_r.$$

Assume that A_r has an eigenvalue $|\lambda| = 1$ with $\lambda \neq 1$ and derive a contradiction. Consider the case $\lambda = -1$. Then there is a nonzero real p such that $p^\top A = -p^\top$. This implies that

$$p^\top \overline{Z}_r \overset{d}{=} -p^\top \overline{Z}_r$$

contradicting $\pi \notin \mathrm{S}_{\mathrm{cyc}}$ for $Z_{\mathrm{cyc}} = (p^\top P_r Z, p^\top P_r Z)$ and angle $\theta = \pi$. So $\lambda \notin \mathbb{R}$ cannot be real. Write $\lambda = e^{i\theta}$ and let $p = p_1 + ip_2$ be a nonzero left eigenvector for λ. Note that neither p_1 nor p_2 can be zero as otherwise the equation $A_r p_i = e^{i\theta} p_i$ would hold for one of $i = 1, 2$. This is impossible because the left-hand side is purely real while the right-hand side is not. Taking real and imaginary parts of $A_r^\top p = e^{i\theta} p$ yields

$$A_r^\top p_1 = \cos\theta p_1 - \sin\theta p_2, \quad A_r^\top p_2 = \sin\theta p_1 + \cos\theta p_2.$$

This implies that

$$\begin{aligned}(p_1^\top A_r \overline{Z}_r, p_2^\top A_r \overline{Z}_r)^\top &\overset{d}{=} ((\cos\theta p_1 - \sin\theta p_2)^\top \overline{Z}_r, (\sin\theta p_1 + \cos\theta p_2)^\top \overline{Z}_r)^\top \\ &\overset{d}{=} R_\theta (p_1^\top \overline{Z}_r, p_2^\top \overline{Z}_r)^\top.\end{aligned}$$

By Lemma A.6, $\theta = 2\pi\frac{s}{t}$ for $s \leq t \in \mathbb{N}$. Choose $s \leq t$ such that $\gcd(s,t) = 1$. Then

$$\frac{\mathbb{N}s}{t} \bmod 1 = \frac{\mathbb{N}}{t} \bmod 1.$$

In particular, the relation above holds for some $\theta = \frac{2\pi}{t}$ with $t \in \mathbb{N}$. This is a contradiction to $\pi \notin \mathrm{S}_{\mathrm{cyc}}$ and we must have $\lambda = 1$. ∎

Corollary 2.5 follows from Theorem 2.4.

Proof of Corollary 2.5. Since $\pi \notin \mathrm{S}_{\mathrm{cov}}$, Theorem 2.4 implies that $\rho(A) \leq 1$ and A is diagonal in the generalized eigenspace of all eigenvalues with magnitude 1. Further, since $\pi \notin \mathrm{S}_{\mathrm{dec}}$, the spectrum of A has no eigenvalues with magnitude smaller than 1, so we can write $A = PDP^{-1}$ for $P, D \in \mathbb{C}^{n\times n}$ with D diagonal and all diagonal entries of complex magnitude one. As $\pi \notin \mathrm{S}_{\mathrm{cyc}}$, A has no eigenvalues with $|\lambda| = 1$ and $\lambda \neq 1$, so $A = I$. Additionally, by $\pi \notin \mathrm{S}_{\mathrm{dec}}$ and full covariance rank of Σ_{YY},

$$B = -K$$

where $K = \Sigma_{XY}\Sigma_{YY}^{\dagger}$. Finally, we derive the value of c. Let ν witness $\pi \in \mathrm{E}^{\mathrm{EnKU}}$, meaning that $(Z + MY, Y) \sim \pi$ for $Z \sim \nu$ independent of $Y \sim \pi_Y$ and $M = K$. Then, by exactness

$$Z + KY + BY + c \overset{d}{=} Z + Ky_\star.$$

Taking expectations on both sides shows

$$c = Ky_\star$$

and completes the proof. ∎

Finally, we can prove Theorem 3.3,

Proof of Theorem 3.3. Say $\pi \in \mathrm{F} \cap \mathrm{S}^{\mathrm{c}}_{\mathrm{nl-dec}}$. We show $\pi \in \mathrm{E}^{\mathrm{EnKU}}$. By Proposition 3.1 there are measurable $d : \mathbb{R}^m \to \mathbb{R}^n$ and $\nu \in \mathcal{P}_2(\mathbb{R}^n)$ such that

$$\pi_{X|Y=y_\star} = T_{d(y_\star)}\nu \text{ for all } y_\star \in \mathbb{R}^m.$$

Letting $(X, Y) = (Z + d(Y), Y)$ for $Z \sim \nu$ independent of $Y \sim \pi_Y$, this means that $(X, Y) \sim \pi$. Since π has an exact weakly $y_\star$-dependent affine conditioning map there are $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $c : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$AZ + Ad(Y) + BY + c(y_\star) \overset{d}{=} Z + d(y_\star)$$

π_Y-a.s. For any such $y_\star$ we can rewrite this as

$$A\overline{Z} + U \overset{d}{=} \overline{Z}$$

for $U = (A - I)\mathbb{E}(Z) + Ad(Y) - d(y_\star) + BY + c(y_\star)$. Theorem A.2 implies that U_u and U_r are a.s. constant vectors. Further, writing $A_s = P_s A P_s$, we have

$$\overline{Z}_s \overset{d}{=} \sum_{k=0}^{\infty} A_s^k U_s^{(k)}$$

for $U_s^{(k)}$ independent copies of U_s that are chosen through independent copies $Y_s^{(k)}$ of Y. Say V_s is nontrivial, then there is a nonzero eigenvector p of $A_s^\top$ with eigenvalue $|\lambda| < 1$. This implies that

$$p^\top \overline{Z}_s \overset{d}{=} \sum_{k=0}^{\infty} \lambda^k p^\top U_s^{(k)}.$$

for some $|\lambda| < 1$. We can expand

$$p^\top U_s^{(k)} = p^\top P_s B Y^{(k)} + p^\top P_s A d(Y^{(k)}) + p^\top P_s \left((A - I)\mathbb{E}(Z) - d(y_\star) + c(y_\star)\right)$$

and defining $b = \frac{1}{1-\lambda} p^\top P_s \left((A - I)\mathbb{E}(Z) - d(y_\star) + c(y_\star)\right)$, $q^\top = p^\top P_s B$, $w^\top = p^\top A_s$, $v^\top = p^\top P_s$ we can rewrite this as

$$p^\top \overline{Z} = \sum_{k=0}^{\infty} \lambda^k \left(q^\top Y^{(k)} + w^\top d(Y^{(k)}) \right) + b.$$

Since $p \neq 0$, this contradicts $\pi \in \mathrm{S}^{\mathrm{c}}_{\mathrm{nl-dec}}$ and thus we must have that U_s is empty. Therefore, U (and in particular $Ad(Y) + BY$) is a.s. constant. Further, $A \in \mathrm{GL}(n)$ as the stable block is trivial. Therefore, almost surely

$$d(Y) = -A^{-1}BY + f$$

for a constant $f \in \mathbb{R}^n$, meaning that d is a.s. affine. However, then $\pi \in \mathrm{E}^{\mathrm{EnKU}}$. ∎

REFERENCES

[1] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*, Springer, 2005.

[2] J. L. Anderson, *An ensemble adjustment kalman filter for data assimilation*, Monthly weather review, 129 (2001), pp. 2884–2903.

[3] J. L. Anderson, *An adaptive covariance inflation error correction algorithm for ensemble filters*, Tellus A: Dynamic meteorology and oceanography, 59 (2007), pp. 210–224.

[4] M. Asch, M. Bocquet, and M. Nodet, *Data assimilation: methods, algorithms, and applications*, SIAM, 2016.

[5] E. Bach, R. Baptista, E. Calvello, B. Chen, and A. Stuart, *Learning enhanced ensemble filters*, arXiv preprint arXiv:2504.17836, (2025).

[6] C. H. Bishop, B. J. Etherton, and S. J. Majumdar, *Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects*, Monthly weather review, 129 (2001), pp. 420–436.

[7] E. Calvello, S. Reich, and A. M. Stuart, *Ensemble kalman methods: A mean-field perspective*, Acta Numerica, 34 (2025), pp. 123–291.

[8] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen, *Data assimilation in the geosciences: An overview of methods, issues, and perspectives*, Wiley Interdisciplinary Reviews: Climate Change, 9 (2018), p. e535.

[9] D. L. Cohn, *Measure theory*, vol. 1, Springer, 2013.

[10] P. Del Moral and J. Tugaut, *On the stability and the uniform propagation of chaos properties of ensemble kalman–bucy filters*, The Annals of Applied Probability, 28 (2018), pp. 790–850.

[11] G. Evensen, *The ensemble kalman filter: Theoretical formulation and practical implementation*, Ocean dynamics, 53 (2003), pp. 343–367.

[12] G. Evensen, *The ensemble kalman filter for combined state and parameter estimation*, IEEE Control Systems Magazine, 29 (2009), pp. 83–104.

[13] W. Feller et al., *An introduction to probability theory and its applications*, vol. 963, Wiley New York, 1971.

[14] A. Gelb et al., *Applied optimal estimation*, MIT press, 1974.

[15] I. Grooms, *A note on the formulation of the ensemble adjustment kalman filter*, arXiv preprint arXiv:2006.02941, (2020).

[16] P. Hall and C. C. Heyde, *Martingale limit theory and its application*, Academic press, 2014.

[17] T. M. Hamill, J. S. Whitaker, and C. Snyder, *Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter*, Monthly Weather Review, 129 (2001), pp. 2776–2790.
[18] T.-V. Hoang, S. Krumscheid, H. G. Matthies, and R. Tempone, *Machine learning-based conditional mean filter: A generalization of the ensemble kalman filter for nonlinear data assimilation*, arXiv preprint arXiv:2106.07908, (2021).
[19] P. L. Houtekamer and H. L. Mitchell, *A sequential ensemble kalman filter for atmospheric data assimilation*, Monthly weather review, 129 (2001), pp. 123–137.
[20] B. R. Hunt, E. J. Kostelich, and I. Szunyogh, *Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter*, Physica D: Nonlinear Phenomena, 230 (2007), pp. 112–126.
[21] M. A. Iglesias, K. J. Law, and A. M. Stuart, *Ensemble kalman methods for inverse problems*, Inverse Problems, 29 (2013), p. 045001.
[22] O. Kallenberg, *Foundations of modern probability*, Springer, 1997.
[23] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Journal of Basic Engineering, (1960).
[24] K. Law, A. Stuart, and K. Zygalakis, *Data assimilation*, Cham, Switzerland: Springer, 214 (2015), p. 52.
[25] F. Le Gland, V. Monbet, and V.-D. Tran, *Large sample asymptotics for the ensemble Kalman filter*, PhD thesis, INRIA, 2009.
[26] J. Lei and P. Bickel, *A moment matching ensemble filter for nonlinear non-gaussian data assimilation*, Monthly Weather Review, 139 (2011), pp. 3964–3973.
[27] M. Loève, *Nouvelles classes de lois limites*, Bulletin de la Société Mathématique de France, 73 (1945), pp. 107–126.
[28] M. McCabe and J. Brown, *Learning to assimilate in chaotic dynamical systems*, Advances in neural information processing systems, 34 (2021), pp. 12237–12250.
[29] L. Nerger, T. Janjić, J. Schröter, and W. Hiller, *A unification of ensemble square root kalman filters*, Monthly Weather Review, 140 (2012), pp. 2335–2345.
[30] J. P. Nolan, *Multivariate stable distributions: Approximation, estimation, simulation and identification*, in A Practical Guide to Heavy Tails: Statistical Techniques and Applications, 1998, pp. 509–525.
[31] P. Parks, *Liapunov and the schur-cohn stability criterion*, IEEE Transactions on Automatic Control, 9 (1964), pp. 121–121.
[32] T. Rajba, *On multiple decomposability of probability measures on r*, Demonstratio Mathematica, 34 (2001), pp. 63–82.
[33] S. Reich and C. Cotter, *Probabilistic forecasting and Bayesian data assimilation*, Cambridge University Press, 2015.
[34] G. Revach, N. Shlezinger, X. Ni, A. L. Escoriza, R. J. Van Sloun, and Y. C. Eldar, *Kalmannet: Neural network aided kalman filtering for partially known dynamics*, IEEE Transactions on Signal Processing, 70 (2022), pp. 1532–1547.
[35] P. Sakov and P. R. Oke, *Implications of the form of the ensemble transformation in the ensemble square root filters*, Monthly Weather Review, 136 (2008), pp. 1042–1053.
[36] G. Samorodnitsky and M. S. Taqqu, *Stable non-Gaussian random processes: stochastic models with infinite variance*, vol. 1, CRC press, 1994.
[37] K.-I. Sato, *Lévy processes and infinitely divisible distributions*, vol. 68, Cambridge university press, 1999.
[38] C. Schillings and A. M. Stuart, *Analysis of the ensemble kalman filter for inverse problems*, SIAM Journal on Numerical Analysis, 55 (2017), pp. 1264–1290.
[39] C. Schillings and A. M. Stuart, *Convergence analysis of ensemble kalman inversion: the linear, noisy case*, Applicable Analysis, 97 (2018), pp. 107–123.
[40] A. Spantini, R. Baptista, and Y. Marzouk, *Coupling techniques for nonlinear ensemble filtering*, SIAM Review, 64 (2022), pp. 921–953.
[41] M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, *Ensemble square root filters*, Monthly weather review, 131 (2003), pp. 1485–1490.
[42] X. T. Tong, A. J. Majda, and D. Kelly, *Nonlinear stability of the ensemble kalman filter with adaptive covariance inflation*, arXiv preprint arXiv:1507.08319, (2015).
[43] R. Van Handel, *Uniform observability of hidden markov models and filter stability for unstable signals*,

(2009).
[44] X. Wang, C. H. Bishop, and S. J. Julier, *Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble?*, Monthly Weather Review, 132 (2004), pp. 1590–1605.
[45] H. Weyl, *Über die gleichverteilung von zahlen mod. eins*, Mathematische Annalen, 77 (1916), pp. 313–352.
[46] J. S. Whitaker and T. M. Hamill, *Ensemble data assimilation without perturbed observations*, Monthly weather review, 130 (2002), pp. 1913–1924.
[47] V. M. Zolotarev, *One-dimensional stable distributions*, vol. 65, American Mathematical Soc., 1986.