

# Connections between sequential Bayesian inference and evolutionary dynamics

Sahani Pathiraja<sup>1</sup> and Philipp Wacker<sup>2</sup>

<sup>1</sup>School of Mathematics & Statistics, UNSW Sydney, Australia,

<https://orcid.org/0000-0002-0114-3164>

<sup>2</sup>Faculty of Engineering, University of Canterbury, New Zealand,

<https://orcid.org/0000-0001-8718-4313>

## Abstract

It has long been posited that there is a connection between the dynamical equations describing evolutionary processes in biology and sequential Bayesian learning methods. This manuscript describes new research in which this precise connection is rigorously established in the continuous time setting. Here we focus on a partial differential equation known as the Kushner-Stratonovich equation describing the evolution of the posterior density in time. Of particular importance is a piecewise smooth approximation of the observation path from which the discrete time filtering equations, which are shown to converge to a Stratonovich interpretation of the Kushner-Stratonovich equation. This smooth formulation will then be used to draw precise connections between nonlinear stochastic filtering and replicator-mutator dynamics. Additionally, gradient flow formulations will be investigated as well as a form of replicator-mutator dynamics which is shown to be beneficial for the misspecified model filtering problem. It is hoped this work will spur further research into exchanges between sequential learning and evolutionary biology and to inspire new algorithms in filtering and sampling.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Replicator-mutator equations . . . . .	3
1.2	Sequential Bayesian inference & replicator-mutator equations . . . . .	4
1.3	Research contributions . . . . .	5
1.4	Notation . . . . .	6
<b>2</b>	<b>The replicator equation &amp; continuous time Bayesian inversion</b>	<b>6</b>
<b>3</b>	<b>Time varying Crow-Kimura replicator-mutator and nonlinear filtering</b>	<b>9</b>
<b>4</b>	<b>Replicator-mutator equations &amp; filtering with misspecified models: the Linear-Gaussian case</b>	<b>14</b>
4.1	Replicator-mutator with $s = 0, r \neq 1$ & Inflated Kalman-Bucy filtering . . . . .	16
4.2	Local vs non-local replicator-mutator for misspecified model filtering . . . . .	17
4.3	Numerical experiments . . . . .	20

<b>5 Proofs</b>	<b>24</b>
5.1 Proof of Lemma 2.1 . . . . .	24
5.2 Proof of Lemma 3.1 . . . . .	25
5.3 Proof of Theorem 3.1 . . . . .	25
5.4 Proof of Lemma 4.1 . . . . .	33
5.5 Proof of Lemma 4.3 . . . . .	36
5.6 Proof of Lemma 4.4 . . . . .	37
5.7 Proof of Lemma 4.5 . . . . .	38
5.8 Proof of Lemma 4.6 . . . . .	40
5.9 Proof of Lemma 4.7 . . . . .	42
<b>6 Conclusion</b>	<b>43</b>
<b>A Technical lemmata</b>	<b>44</b>

# 1 Introduction

It has been posited that there is a connection between sequential Bayesian inference and dynamical models describing evolutionary biological processes. Understanding and studying this connection has the potential to provide valuable insights on improved algorithms for complex Bayesian inference and sampling tasks arising in a wide range of fields in the data science, engineering and machine learning. Specifically, the key connection to sequential Bayesian estimation is via the so-called replicator-mutator partial differential equations [Kim24; Hof85], describing the time evolution of a large population of individuals with certain traits or attributes due to mutation and reproduction (or selection). Broadly speaking, sequential Bayesian estimation procedures bear striking similarity to the way species respond to evolutionary pressure moderated by a fitness landscape. The correspondence is as follows:

- states or parameters  $\leftrightarrow$  traits
- prior distribution  $\leftrightarrow$  current population
- prediction (in the case of filtering or hidden markov models)  $\leftrightarrow$  mutation
- likelihood function  $\leftrightarrow$  fitness landscape governing selection or birth-death

This connection has been discussed most notably in [Mor97; Sha09; Har09b; Aky17; Czé+22], primarily in the context of discrete time and discrete trait space problems. Some of the earliest connections between discrete time particle based Bayesian updating and genetic mutation-selection models seem to have been in e.g. [Mor97; Mor04; DG05]. [Sha09] raised awareness to the similarity to replicator equations specifically; they show how Bayesian updating corresponds to one step of a discrete-time and continuous-trait replicator equation without mutation. Around the same time, a similar point was made in [Har09b]. This connection to replicator equations with mutation was further extended to the setting of sequential in time inference with hidden Markov models (also known as sequential filtering, discrete time data assimilation) in [Aky17; Czé+22]. They showed that discrete time replicator-mutator dynamics consists of a sequence of (discrete in time) alternating mutation and updating steps, as in sequential filtering. The PhD thesis [Zha17] draws some interesting connections to optimisation and interacting particle approaches to sequential filtering, e.g. the Feedback Particle Filter [Yan+12; Lau+14]. In this manuscript we focus on making this connection precise in the continuous time and continuous trait space case, which has not yet been

explored thoroughly in the literature. More recently, there has been interest in incorporating replicator or “birth-death” dynamics into sampling algorithms for optimisation and inversion tasks, see e.g. [LLN19; LSW23; Che+24a].

## 1.1 Replicator-mutator equations

The replicator-mutator equation is a broad class of dynamical systems modelling the response of a distribution of traits to evolutionary adaptation to an external fitness landscape. Early research on this class of models started with [Kim65; CK70] (sometimes also known as the Crow-Kimura equation), followed up by influential work in [Aki79; SS83] and many others. In these early works, the trait is often considered to be discrete (corresponding to discrete gene loci with a finite number of alleles) in the form  $\dot{\pi}_t(i) = \pi_t(i) \left( \sum_j w_{ij} \pi_t(j) - \sum_{rs} w_{rs} \pi_t(r) \pi_t(s) \right)$ , where  $\pi_t$  is a  $n$  dimensional probability vector of the relative frequency of each discrete trait, the matrix entries  $w_{ij}$  record the fitness benefit (or harm) of the presence of trait  $j$  for the proliferation of trait  $i$ , often assumed to be symmetric. The second term in the brackets ensures that  $\pi \in S^n$ , an  $n$ -dimensional probability simplex, i.e.,  $\pi(i) \geq 0$  and  $\sum_i \pi(i) = 1$  for all times. The dynamical properties of this system have been studied extensively see, e.g., [Bom90; OR01] and its relation to Fisher information and entropy has been studied in [BP16; Bae21]. This system also has a geometrical structure amenable to optimisation, as it can be seen to be a gradient flow of the total fitness with respect to the Shashahani metric, see [Har09a; Har09b; FA95; Cha+21]. There is also a rich connection to game theory, as in [CT14] (with traits being interpreted as game strategies). The continuous trait space setting has received comparatively less attention, but has been studied as early as [Kim65] or in [CHR06; Gil+17; Vla20]. These models are popular in the mathematical evolution, biology, and ecology literature, primarily in discrete time, as in [KM14; KNP18; MHK14]. Nevertheless, the continuous time form of these models has been subject to attention from a mathematical analysis perspective [CS09; Cha+20]. The unnormalised or normalised form of the continuous-time continuous-trait replicator-mutator partial differential equations (PDEs) is given by

$$\partial_t \mu_t(x) = \mu_t(x) \mathbb{E}_{z \sim \mu_t} [f_t(x, z)] + \mathcal{L}^* \mu_t(x) \quad (1.1)$$

$$\partial_t \rho_t(x) = \rho_t(x) (\mathbb{E}_{z \sim \rho_t} [f_t(x, z)] - \mathbb{E}_{\rho_t} [f_t]) + \mathcal{L}^* \rho_t(x), \quad (1.2)$$

where  $\rho_t(x) \geq 0$  and  $\mu_t(x) \geq 0$  denote the normalised and unnormalised density functions respectively, describing the distribution of traits  $x \in \mathbb{R}^n$  in the population. Here,  $\mathcal{L}^* q_t(x)$  is an optional mutation term, where  $\mathcal{L}^*$  is the adjoint generator of a diffusion process, e.g.,  $\mathcal{L}^* = -\Delta$  for mutation according to standard Brownian motion. If  $\mathcal{L}^* = 0$ , i.e., no mutation exists, we call this the (pure) replicator equation. The (potentially non-local & time-dependent) selection or fitness function is denoted by  $f_t(x, z)$  and the net birth-death rate for a given trait  $x$  at time  $t$  is given by  $\mathbb{E}_{z \sim \rho_t} [f_t(x, z)]$ , where the subscript indicates that only the expectation is taken over the  $z$  variable only. A simplified form of the replicator-mutator often appearing in the literature [TLK96; AC14; AC17] is the local PDE

$$\partial_t \rho_t(x) = \rho_t(x) (f_t(x) - \mathbb{E}_{\rho_t} [f_t]) + \mathcal{L}^* \rho_t(x), \quad (1.3)$$

where the non-local fitness function in (1.2) has been replaced by a fitness function that no longer depends on the current distribution of traits, only on the value of the trait itself. An ubiquitous example is the quadratic fitness,  $f_t(x) = -\frac{1}{2} \|Hx - y_t\|_{\Xi}^2$ , which penalises traits  $x$  that have a large misfit to the (potentially) time varying data  $y_t$ , is connected to least-squares estimation. The general idea here is that  $H$  maps traits  $x \in X$  to features  $Hx \in Y$ , and  $y_t$  represents an optimal feature

at time  $t$ , with fitness being quantified as a quadratic deviation, possibly preconditioned with a covariance matrix  $\Xi$ . A non-local version of this fitness function has been presented in [CHR06], i.e.

$$f_t(x, z) = -\frac{r}{2}\|Hx - y_t\|_{\Xi}^2 - \frac{r}{2}\|Hz - y_t\|_{\Xi}^2 + s\langle Hx - y_t, Hz - y_t \rangle_{\Xi} \quad (1.4)$$

with  $r > 0$ ,  $s < r$  (the case  $H = I, r = 1, y_t = 0$  was presented in [CHR06]). This fitness function takes into account both a given trait's fitness on its own, but also beneficial or adversary effects of the group fitness through the  $s\langle Hx - y_t, Hz - y_t \rangle_{\Xi}$  term. This non-local fitness function is studied further for the case of non-linear  $h(x)$  in Section 3 and in the linear setting in Section 4 of this manuscript.

## 1.2 Sequential Bayesian inference & replicator-mutator equations

We will show in Section 3 that there exists a direct connection between the non-linear filtering and replicator-mutator equations, which has not yet been made explicit in the literature, to the best of our knowledge. Below we summarise the main findings of Section 3 of this manuscript, paying special attention to the linear-Gaussian setting (although the results hold for the more general non-linear setting, as detailed in Section 3). We first describe the standard non-linear filtering problem. Consider Euclidean spaces  $X = \mathbb{R}^m$  and  $Y = \mathbb{R}^n$ , covariance matrices  $\Sigma \in \mathbb{R}^{m \times m}, \Xi \in \mathbb{R}^{n \times n}$ , sufficiently regular mapping  $g : X \rightarrow X$  and  $h : X \rightarrow Y$ , and the following signal-observation pair,

$$dX_t = g(X_t)dt + \Sigma^{1/2} dW_t; \quad (\text{signal}) \quad (1.5)$$

$$dZ_t = h(X_t)dt + \Xi^{1/2} dB_t \quad (\text{observation}). \quad (1.6)$$

The goal of filtering is to reconstruct the signal  $X_t$  by means of the noisy observation path  $\{Z_s\}_{s \leq t}$ . Since  $X_t$  cannot be uniquely identified from this data, the correct object to study is the conditional distribution of  $X_t$  from the data  $\{Z_s\}_{s \leq t}$ , which is known to evolve in time according to the Kushner-Stratonovich equation,

$$dp_t(x) = \mathcal{L}^* p_t(x) + p_t(x) (h(x) - \mathbb{E}_{p_t}[h])^\top \Xi^{-1} (dZ_t - \mathbb{E}_{p_t}[h]dt)$$

where  $\mathcal{L}^*$  denotes the adjoint of the infinitesimal generator of (1.5) (i.e. the Kolmogorov forward operator),

$$\mathcal{L}^* p_t(x) = -\nabla \cdot (p_t(x) g(x)) + \frac{1}{2} \nabla \cdot (\Sigma \nabla p_t(x)).$$

As we will detail in Section 3, this is equivalent to the replicator-mutator equation (specifically, the Crow-Kimura equation),

$$\partial_t \rho_t(x) = \underbrace{-\nabla \cdot (\rho_t(x) g(x)) + \frac{1}{2} \nabla \cdot (\Sigma \nabla \rho_t(x))}_{\text{mutation}} + \underbrace{\rho_t(x) (f_t(x) - \mathbb{E}_{\rho_t}[f_t])}_{\text{replication}} \quad (1.7)$$

with fitness function

$$f_t(x) = -\frac{1}{2} \|h(x) - y_t\|_{\Xi}^2 \quad (1.8)$$

A formal interpretation  $y(t) = \frac{dZ_t}{dt}$  (which is not a valid object if  $Z_t$  is indeed a sample path of (1.6)) allows to then see the structural equivalence to the Kushner-Stratonovich equation. We give more detail in Section 3 on how this comparison can be made more rigorous. Additionally, in Section

3 we will establish connections between a modified Kushner equation and the replicator-mutator equation with a non-local fitness function, in the form of (1.4) (and with non-linear  $h$ ). We will use this equivalence to demonstrate how this a modified Kushner equation can be beneficial for filtering with misspecified models (see Section 4).

The following simplified case is presented to further help motivate the proofs below. We set  $g \equiv 0$ , and  $h(x) = Hx$  for some matrix  $H \in \mathbb{R}^{m \times m}$ , which means that the tracked distribution  $\rho_t$  will remain Gaussian, if  $\rho_0$  is Gaussian and we use the notation  $\rho_t = \mathcal{N}(m_t, C_t)$ . In this case the (unnormalised) replicator-mutator equation with fitness function (1.4) can be re-written as

$$\partial_t \mu_t(x) = \mu_t(x) \cdot \left[ -\frac{s}{2} \|Hx - Hm_t\|_{\Xi}^2 - \frac{r-s}{2} \|Hx - y_t\|_{\Xi}^2 - \frac{r-s}{2} \|Hm_t - y_t\|_{\Xi}^2 \right] + \frac{1}{2} \nabla \cdot (\Sigma \nabla \mu_t(x)).$$

For the special case  $r = 1, s = 0$ , the above equation collapses to the familiar Kalman-Bucy equations. This formulation allows us to see that the fitness function takes into account both the considered trait's fitness as a quadratic deviation from the "observation"  $y_t$  in the form of  $-\frac{r-s}{2} \|Hx - y_t\|_{\Xi}^2$ , and also a "uniformity term"  $-\frac{s}{2} \|Hx - Hm_t\|_{\Xi}^2$  as measured by quadratic deviation from the distribution mean. In other words, the most adapted trait  $x$  is both close to the mean trait  $m_t$  (in a specific sense in observation space  $Y$ ), and also close to the observation  $y_t$  when mapped into observation space. The parameter  $s$  balances the relative strength of these contributions. The following cases for how  $s$  relates to  $r$  have very different behaviour: The case  $s = 0$  corresponds to penalty by deviation from the data only (analogous to the function of the log-likelihood in Bayesian inference or filtering), while  $s = r$  means that the most adapted trait is the mean trait, independent of its actual "feature"  $Hx$  and how it relates to the data  $y_t$  at time  $t$ . In fact, the consequence of setting  $r = s$  is that the fitness function becomes equal to  $-\frac{r}{2} \|Hx - Hm_t\|_{\Xi}^2$ , i.e., optimal traits  $x$  have the same feature  $Hx$  as the mean of the population  $m_t$ . For  $s \in (0, r)$ , a compromise between these two edge cases holds. In the domain  $s < 0$  on the other hand we observe evolutionary fitness attributed to non-uniformity, i.e., being further away from the population mean  $m_t$  than strictly necessary for pure adaptation to the data  $y_t$ . In the mathematical theory of (biological) evolution, this parameter  $s$  can easily be motivated by similarities to biological systems, where uniformity or uniqueness (as opposing poles) can lead to improved fitness, but to the best of our knowledge this has not been studied in depth in the filtering context. In section 4 we will explore this further in the context of misspecified model filtering.

### 1.3 Research contributions

We have the following main contributions in this manuscript:

1. A rigorous connection between the crow-kimura replicator-mutator equations and non-linear filtering is established in Theorem 3.1 (section 3). In particular, we clarify how a time varying fitness functional coinciding with a discrete time measurement model can be reconciled with the Kushner-Stratonovich partial differential equation arising in stochastic filtering.
2. Theorem 3.1 more broadly establishes a "generalised" stochastic filtering partial differential equation as the continuous time limit of a non-local replicator-mutator equation. This pde is the subject of further study in Section 4 where its utility in the context of misspecified model filtering is established.
3. In section 4.1 we demonstrate that for a specific choice of parameters ( $r > 0, s = 0$ ), this "generalised" stochastic filtering partial differential equation coincides with covariance inflation ensemble Kalman-Bucy filtering [And07; DSH20; MH00; BMP18; BD23] in the linear-Gaussian setting. Section 4.2 then demonstrates the benefit of this "generalised" filtering pde

(where  $s \neq 0$ ) for filtering in the presence of model errors. Specifically, this is established rigorously through a series of lemmas for the linear-Gaussian setting and where the misspecification arises through an unknown constant bias in the signal dynamics. We demonstrate that by optimally choosing  $(r, s)$ , one can minimise mean squared error and simultaneously obtain realistic uncertainty quantification, which has traditionally been difficult with classical covariance inflation.

## 1.4 Notation

We define some notation that will be used throughout the manuscript.

Set  $X = \mathbb{R}^m$  the trait space,  $H : X \rightarrow Y$  be a mapping, and  $Y = \mathbb{R}^n$ .

$m$  = dimension of state/trait space

$n$  = dimension of observation/fitness space

$$\|x\|_{\Xi}^2 = x^\top \Xi^{-1} x$$

$\dot{m}_t$  is used to denote  $\frac{dm_t}{dt}$  for any vector  $m_t \in \mathbb{R}^d$  depending only on  $t$ .

$\partial_t p_t(x)$  is used to denote  $\frac{\partial p_t(x)}{\partial t}$  for  $\rho_t(x) : [0, \infty) \times \mathbb{R}^m \rightarrow \mathbb{R}$

$\mathbb{E}_{p_t}[f] = \int f(x) \rho_t(x) dx$  and where necessary, we specify the variable in the subscript to indicate the variable to which the integration operation applies, i.e.  $\mathbb{E}_{z \sim \rho_t}[f_t(x, z)] = \int f_t(x, z) \rho_t(z) dz$ .

$f_t(x, z) : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$  is the non-local fitness landscape at time  $t$  (omission of the  $t$  subscript is used to indicate a time-independent fitness landscape). Frequency dependent fitness of trait  $x \in \mathbb{R}^m$  is indicated by the term  $\mathbb{E}_{z \sim \rho_t}[f_t(x, z)] \equiv \int f_t(x, z) \rho_t(z) dz$ .

$p_t$  and  $q_t$  is used to refer to the normalised and unnormalised filtering density satisfying the Kushner-Stratonovich and Zakai equations respectively.

$\rho_t$  and  $\mu_t$  is used to refer to the normalised and unnormalised density function from Crow Kimura replicator-mutator respectively.

## 2 The replicator equation & continuous time Bayesian inversion

Before presenting the rigorous connection between sequential filtering and replicator-mutator equations with time dependent fitness, we present the case of static fitness functions and continuous time Bayesian inversion. Inversion [EHN96] is the problem of inferring an unknown parameter  $x \in X = \mathbb{R}^m$  from a noisy measurement  $y \in Y = \mathbb{R}^n$  of form

$$y = h(x) + \varepsilon, \tag{2.1}$$

where  $H : X \rightarrow Y$  is the so-called forward operator and  $\varepsilon$  is measurement noise, commonly assumed to be zero-mean Gaussian,  $\varepsilon \sim \mathcal{N}(0, \Xi)$ . The operator  $H$  is analogous to the observation operator  $h(x)$  in the filtering setting. Unfavorable properties of the operator and the noise makes a naive inversion procedure ill-defined, so additional regularisation of the inversion procedure is necessary. The Bayesian approach to inversion ([Stu10]) requires a prior distribution  $q_0$  of the unknown parameter  $x$  and produces the Bayesian posterior measure  $\nu^y$  via

$$\frac{d\nu^y}{d\nu_0}(x) \propto \exp\left(-\frac{1}{2}\|h(x) - y\|_{\Xi}^2\right). \tag{2.2}$$

This inversion procedure constitutes a one-step method from the prior  $\nu_0$  to the posterior  $\nu^y$ . In the following, we assume that all measures have a Lebesgue density, which we denote  $p_0$  and  $p^y$ . A commonly adopted prior  $p_0$  is the multivariate Gaussian  $\mathcal{N}(m_0, C_0)$ .

When  $m, n$  are large and the prior and posterior are significantly “different” it can be advantageous to gradually transform the prior to the posterior, either over a finite or infinite time horizon. When this is done over a finite time interval, this procedure is known as tempering or annealing in the sequential monte carlo literature [GM98; Nea01; CCK24] and also the homotopy approach to Bayesian inversion (e.g. [Rei11; Blö+22; CST20]). More specifically, this approach modifies the single step from prior  $\mu_0$  to posterior  $\mu^y$  into a smooth transition by introducing a pseudotime  $t \in [0, 1]$ , and intermediate measures with density  $p_t$  via

$$p_t(x) \propto \exp\left(-\frac{t}{2}\|h(x) - y\|_{\Xi}^2\right) \cdot p_0(x)$$

such that  $p_1 = p^y$ . As outlined in e.g. [PR23] it can be seen that a family of probability densities defined via  $p_t(x) \propto \exp(tf(x))p_0(x)$  is the solution of the infinite-dimensional system of ODEs

$$\frac{dp_t(x)}{dt} = \frac{d}{dt} \left( \frac{\exp(tf(x))p_0(x)}{\mathbb{E}_{p_t}[f]} \right) = (f(x) - \mathbb{E}_{p_t}[f])p_t(x), \quad (2.3)$$

This is identical to the pure replicator equation, i.e., the replicator-mutator equation without any mutation, with the fitness function being the log-likelihood,

$$f(x) = -\frac{1}{2}\|h(x) - y\|_{\Xi}^2.$$

In the simplest setting,  $y$  is a time-independent feature, which then means that the fitness function is also static. From a sampling point of view, it is then possible to construct both deterministic and stochastic schemes which have the replicator equation as their density evolution equation. To do so, we switch to the so-called *Eulerian* perspective on this problem, as presented e.g. in [Rei11]. The goal is to construct a vector field  $(t, x) \mapsto v(t, x)$  such that the family of diffeomorphisms  $T_t : X \rightarrow X$  with  $T_t(x_0) := x(t)$  on trait space defined by solutions of the ODE

$$\dot{x}(t) = v(t, x(t)), \quad x(0) = x_0 \quad (2.4)$$

smoothly pushes forward the initial population distribution to the distribution at a later time via

$$(T_t)_\# p_0 = p_t. \quad (2.5)$$

The advantage of this perspective is that if we were able to find such a vector field  $v$ , then we can approximate the solution of (2.3) by sampling  $J$  particles  $\{x_0^{(i)}\}_{i=1}^J \sim p_0$ , evolve them according to (2.4), and the resulting ensemble  $\{x^{(i)}(t)\}_{i=1}^J$  then constitutes valid samples from  $p_t$ . [Vil21, Theorem 5.34] states that such a velocity field  $v$  satisfies the continuity equation

$$\partial_t p_t(x) = -\nabla_x \cdot (v(t, x)p_t(x)). \quad (2.6)$$

On the other hand, comparing the right hand sides of (2.3) and (2.6) means that the vector field needs to be a solution of the Poisson equation

$$-\nabla_x \cdot (v(t, x)p_t(x)) = (f(x) - \mathbb{E}_{p_t}[f])p_t(x). \quad (2.7)$$

which leads us back to the familiar pure replicator dynamics. It is worthwhile noting that this equation also arises in the construction of interacting particle filtering algorithms, where the fitness

function is time-dependent due to the time-varying data term [CX10; Lau+14; PRS21]. Finally, it is possible to include a mutation or “exploration term” to aid in generating samples when  $m$  is large (i.e. the underlying trait space is high dimensional), and is even necessary for particle-based implementations of the above (e.g. [Mor97; DDJ06; MD14; Che+24a; LLN19; LSW23]). In this setting, one arrives at a connection to the Crow-Kimura replicator-mutator equation, albeit with a static (time independent) fitness functional.

In the remainder of this section we discuss some geometric properties of the replicator equation, showing that the continuous trait-space replicator equation follows a gradient flow of the mean fitness energy functional with respect to the Fisher-Rao metric. This extends the well-known result that the replicator equation is a gradient flow with respect to the Shahshahani metric (i.e. the finite dimensional version of the Fisher-Rao metric), see for example Theorem 7.8.3 in [HS98], [FA95] (for the entropy rather than fitness functional), [Har09a], [CS09] and [Cha+21] (in the case of two traits/species) or [Bae21] and [BP16] (including some interesting remarks about speed and acceleration in this metric). To the best of our knowledge, this result has not been established in the continuous trait space setting, although a similar result for a specific fitness functional (the Kullback-Leibler divergence) has appear recently in the sampling literature [Che+24b; MM24; WN24; LLN19; LSW23]. More specifically, [LLN19] shows that a dynamical system similar to the Crow Kimura replicator-mutator equation is a gradient flow of the Kullback-Leibler divergence with respect to the Wasserstein-Fisher-Rao metric.

We start by reminding ourselves of the basics of information geometry. We define  $\mathcal{P}$  as the manifold of absolutely continuous probability measures on  $\mathbb{R}^n$ . Every  $p \in \mathcal{P}$  will be identified with its (Lebesgue) density  $p(x)$ . At a  $p \in \mathcal{P}$  such that  $p(x) > 0$  everywhere, the tangent space of  $\mathcal{P}$  is given by  $\mathcal{T}_p\mathcal{P} = \{\sigma \in C^\infty(\mathbb{R}^n) : \int \sigma(x) dx = 0\}$ . If  $p$  has support  $\text{supp}(p) \subsetneq \mathbb{R}^n$ , then the tangent space is  $\mathcal{T}_p\mathcal{P} = \{\sigma \in C^\infty(\mathbb{R}^n) : \sigma|_{\text{supp}(p)^c} = 0, \int \sigma(x) dx = 0\}$ . The associated cotangent space is given by  $\mathcal{T}_p^*\mathcal{P} = \{\phi \in C^\infty(\mathbb{R}^n)\} / \sim$ , where the equivalence relation  $\sim$  is defined as  $\phi_1 \sim \phi_2$  if and only if  $(\phi_1 - \phi_2)|_{\text{supp}(p)} \equiv \text{const}$ . The dual pairing between  $\phi \in \mathcal{T}_p^*\mathcal{P}$  and  $\sigma \in \mathcal{T}_p\mathcal{P}$  is then canonically defined as  $\langle \sigma, \phi \rangle = \int \sigma(x)\phi(x) dx$ . The Rao-Fisher metric on the tangent space is given by  $g_p(\sigma_1, \sigma_2) = \int \frac{\sigma_1(x)}{p(x)} \frac{\sigma_2(x)}{p(x)} dp(x)$ . This still makes sense on the boundary, i.e. if  $p(x) = 0$  for some  $x$ , because then  $\sigma(x) = 0$  for  $\sigma \in \mathcal{T}_p\mathcal{P}$  and we interpret the resulting expression  $\frac{0}{0} = 0$ . The metric defines an invertible metric tensor  $G(p) : \mathcal{T}_p\mathcal{P} \rightarrow \mathcal{T}_p^*\mathcal{P}$  via  $g_p(\sigma_1, \sigma_2) = \langle \sigma_1, G(p)[\sigma_2] \rangle$  (or equivalently,  $G(p)[\sigma] = g_p(\cdot, \sigma)$ ), which in this case is given explicitly by

$$(G(p)[\sigma])(x) = \begin{cases} \frac{\sigma(x)}{p(x)} & \text{if } p(x) > 0 \\ 0 & \text{else,} \end{cases}$$

which is consistent since  $\sigma \in \mathcal{T}_p\mathcal{P}$  is required to vanish on the set  $\text{supp}(p)^c$ , anyway. In fact, any map identical to  $G(p)$  up to a global constant on  $\text{supp}(p)$ , and with arbitrary values on  $\text{supp}(p)^c$ , would give a valid representative due to the quotient structure on  $\mathcal{T}_p^*\mathcal{P}$ .

The identity  $g_p(\sigma_1, \sigma_2) = \langle \sigma_1, G(p)[\sigma_2] \rangle$  then holds true since

$$\langle \sigma_1, G(p)[\sigma_2] \rangle = \int \sigma_1(x)G(p)[\sigma_2](x) dx = \int \sigma_1(x) \frac{\sigma_2(x)}{p(x)} dx = \int \frac{\sigma_1(x)}{p(x)} \frac{\sigma_2(x)}{p(x)} dp(x)$$

The metric tensor associated to the Rao-Fisher metric has inverse  $G(p)^{-1} : \mathcal{T}_p^*\mathcal{P} \rightarrow \mathcal{T}_p\mathcal{P}$

$$(G(p)^{-1}[\phi])(x) = \left( \phi(x) - \int \phi(y) dp(y) \right) p(x)$$



which is well-defined since  $G(p) [G(p)^{-1}[\phi]] = \phi - \int \phi(y) dp(y) \sim \phi$ . We now recall some facts on gradient flows. Let  $\mathcal{P}$  denote a linear space; then the time evolution of  $p_t \in \mathcal{P}$  is a gradient flow if it can be written as

$$\partial_t p_t = -\mathcal{K}(p_t) \mathcal{F}'(p_t)$$

where  $\mathcal{F} : \mathcal{P} \rightarrow \mathbb{R}$  is an energy functional,  $\mathcal{F}' : \mathcal{P} \rightarrow T_p^* \mathcal{P}$  is the Frechet derivative and  $\mathcal{K}(p) : T_p^* \mathcal{P} \rightarrow T_p \mathcal{P}$  is a linear operator characterising the dissipation mechanism (loosely, giving meaning to how quickly  $\mathcal{F}$  increases/decreases). In this context we are most interested in the case where  $\mathcal{K}$  is related to the metric tensor: The dissipation mechanism  $\mathcal{K}(p)$  is then taken to be  $\mathcal{G}(p)^{-1}$ . We are now ready to state our main result for this section, the proof of which can be found in section 5.1.

**Lemma 2.1.** *The replicator equation (1.2) with  $\mathcal{L}^*$  and with frequency dependent fitness  $\pi_{p_t}(x)$  performs a gradient flow with respect to the Fisher-Rao metric of the average fitness functional  $\mathcal{F}(p) : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$*

$$\mathcal{F}(p) = -\frac{1}{2} \iint f(x, z) p(z) p(x) dz dx \quad (2.8)$$

### 3 Time varying Crow-Kimura replicator-mutator and nonlinear filtering

We begin with background on non-linear filtering. The classical continuous time filtering problem is given by (1.5) and (1.6), restated here for convenience:

$$\begin{aligned} dX_t &= g(X_t) dt + \Sigma^{1/2} dW_t; \\ dZ_t &= h(X_t) dt + \Xi^{1/2} dB_t. \end{aligned}$$

The goal in nonlinear filtering is to estimate the conditional density of the hidden state  $X_t$  at time  $t$ , given the observation filtration  $\mathcal{Z}_t := \sigma(Z_s : s \leq t)$ , which we denote by  $p_t(x)$ . It is well known that under certain regularity conditions on  $g, h$  that  $p_t(x)$  evolves according to the Kushner-Stratonovich equation, a PDE driven by finite dimensional Ito noise,

$$dp_t(x) = \mathcal{L}^* p_t(x) + p_t(x) (h(x) - \mathbb{E}_{p_t}[h])^\top \Xi^{-1} (dZ_t(\omega) - \mathbb{E}_{p_t}[h] dt) \quad (3.1)$$

where  $\mathcal{L}^*$  denotes the adjoint of the infinitesimal generator of (1.5) (i.e. the Kolmogorov forward operator),

$$\mathcal{L}^* p_t(x) = -\nabla \cdot (p_t(x) g(x)) + \frac{1}{2} \nabla \cdot (\Sigma \nabla p_t(x)).$$

The notation  $(\omega)$  is used to emphasise that observation enters as a fixed, known realisation, but we will drop it for the remainder of this section. The unnormalised density  $q_t$  evolves according to the Zakai equation

$$dq_t = \mathcal{L}^* q_t(x) + q_t h(x)^\top \Xi^{-1} dZ_t \quad (3.2)$$

which can be equivalently expressed in Stratonovich form as (see e.g. [HKX02], [PRS21])

$$dq_t(x) = \mathcal{L}^* q_t(x) - \frac{1}{2} h(x)^\top \Xi^{-1} h(x) q_t(x) + q_t(x) h(x)^\top \Xi^{-1} \circ dZ_t \quad (3.3)$$

In order to connect to discrete measurement processes more commonly encountered in practice, as well as to the Crow-Kimura equation replicator-mutator equation, consider a piecewise smooth observation process. The piecewise smooth approach to approximating rough signals has a long history in robust filtering [Cri+13] and has been well-studied more generally by the so-called Wong-Zakai style theorems and in the context of rough path theory [KM16; Pat24]. In this section, we will consider a very specific form of piecewise smooth approximation using piecewise linear interpolations of Brownian noise. More precisely, consider a partition of the time interval  $[0, T]$ ,  $0 < t_1 < t_2 \dots < t_d = T$  with time-step  $t_{i+1} - t_i = \delta_d$  for all  $i$  and such that  $\delta_d \rightarrow 0$  as  $d \rightarrow \infty$  (i.e.  $\delta_d = \frac{T}{d}$ ). Define the piecewise linear approximation to a Brownian path as

$$B_t^d = B_{t_i} + \frac{t - t_i}{\delta_d}(B_{t_{i+1}} - B_{t_i}), \quad t \in [t_i, t_{i+1}),$$

which is piecewise differentiable with time derivative

$$\frac{dB_t^d}{dt} = \frac{1}{\delta_d}(B_{t_{i+1}} - B_{t_i}), \quad t \in [t_i, t_{i+1}).$$

Then a piecewise smooth version of (1.6) can be constructed, for all  $i = 1, 2, \dots, d$  at

$$\frac{dZ_t^d}{dt} = h(x_{t_i}^*) + \Xi^{1/2} \frac{dB_t^d}{dt}, \quad t \in [t_i, t_{i+1}) \quad (3.4)$$

where the notation  $x_t^*$  is used to denote the true hidden state trajectory or reference trajectory that generates the observed measurement. This form allows us to connect more easily to observation models more commonly encountered in practice, i.e.,

$$y_t(x) := \frac{dZ_t^d}{dt} = h(x_t) + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, R)$$

and as we will see in the remainder of the section, to also build a bridge between the replicator-mutator equation and the Kushner-Stratonovich equation. It is worth noting that we consider a piecewise smooth approximation of the observation noise term only, rather than a smooth approximation of the entire observation trajectory as is done more traditionally in the stochastic filtering literature [HKX02; CX10]. Specifically, [HKX02] consider the following approximation

$$Z_t^\Pi = Z_{t_i} = \frac{Z_{t_{i+1}} - Z_{t_i}}{t_{i+1} - t_i}(t - t_i), \quad t \in [t_i, t_{i+1})$$

where  $Z_t$  is a fixed realisation of (1.6) and  $Z_t^\Pi$  denotes the corresponding piecewise approximation. This then yields (assuming  $t_{i+1} - t_i = \delta_d$ ),

$$\frac{dZ_t^\Pi}{dt} = \frac{Z_{t_{i+1}} - Z_{t_i}}{t_{i+1} - t_i} = \frac{1}{\delta_d} \int_{t_i}^{t_{i+1}} dZ_s = \int_{t_i}^{t_{i+1}} h(x_s^*) ds + \frac{dB_t^d}{dt}, \quad t \in [t_i, t_{i+1})$$

so that the observation involves a time integrated version of the hidden state,  $\int_{t_i}^{t_{i+1}} h(x_s^*) ds$  rather than  $h(x_{t_i}^*)$  as in (3.4), which may be more practically relevant particularly when the time between observations is large. This distinction is primarily for comparison to measurement models encountered in practice, both approximate forms can be shown to have valid limiting forms.

In regards to the Crow-Kimura replicator-mutator equation, for the remainder of this section, we focus on the following time-varying quadratic fitness landscape with  $s < r$ ,  $r > 0$ ,

$$f_t(x, z) = -\frac{r}{2} \|h(x) - \frac{dZ_t^d}{dt}\|_\Xi^2 + s \left\langle h(x) - \frac{dZ_t^d}{dt}, h(z) - \frac{dZ_t^d}{dt} \right\rangle_\Xi \quad (3.5)$$

and time-varying optimal feature  $\frac{dZ_t^d}{dt}$  given by (3.4). Notice that with the definition of  $\frac{dZ_t^d}{dt}$  in (3.4),  $f_t(x, z)$  is a piecewise constant in time functional in  $x$ . In the remainder of this section, we will establish equivalence (in a sense to be made precise), between the the C-K replicator-mutator with (3.5) and a modified form of the Zakai equation. For the special case of  $r = 1, s = 0$  in (3.5), we will see that C-K replicator-mutator converges to the standard Zakai equation (3.2) as  $d \rightarrow \infty$ . The following reformulation of the Crow-Kimura replicator-mutator equation with (3.5) will be a useful aid. Its proof can be found in section 5.2.

**Lemma 3.1.** *Consider the Crow-Kimura replicator-mutator equation*

$$\partial_t \rho_t(x) = -\nabla \cdot (\rho_t(x)g(x)) + \frac{1}{2} \nabla \cdot (\Sigma \rho_t(x)) + \rho_t(x)(\mathbb{E}_{z \sim \rho_t}[f_t(x, z)] - \mathbb{E}_{\rho_t}[f_t]) \quad (3.6)$$

with fitness landscape given by (3.5). This equation can be expressed in the form

$$\begin{aligned} \partial_t \rho_t(x) = \mathcal{L}^* \rho_t(x) + (r - s) & \left( -\frac{1}{2} \left( h(x)^\top \Xi^{-1} h(x) - \mathbb{E}_{\rho_t}[h^\top \Xi^{-1} h] \right) + (h(x) - \mathbb{E}_{\rho_t}[h])^\top \Xi^{-1} \frac{dZ_t^d}{dt} \right) \rho_t(x) \\ & - \frac{s}{2} \left( \|h(x) - \mathbb{E}_{\rho_t}[h]\|_\Xi^2 - \mathbb{E}_{\rho_t}[\|h(x) - \mathbb{E}_{\rho_t}[h]\|_\Xi^2] \right) \rho_t(x). \end{aligned} \quad (3.7)$$

where  $\mathcal{L}^*$  denotes the adjoint of the generator of the diffusion process (1.5). Additionally, the unnormalised form is given by

$$\partial_t \mu_t(x) = \mathcal{L}^* \mu_t(x) + \left( -\frac{r}{2} h(x)^\top \Xi^{-1} h(x) + (r - s) h(x)^\top \Xi^{-1} \frac{dZ_t^d}{dt} \right) \mu_t(x) \quad (3.8)$$

Before presenting the main theorem of this section, consider the following simple motivating example to demonstrate why as  $d \rightarrow \infty$ , the Crow-kimura replicator-mutator with fitness landscape (3.5) converges to a pde driven by Stratonovich rather than Ito noise. It should be noted that although the pde is driven by Stratonovich noise, it can be transformed to an Ito version from which the familiar Kushner-Stratonovich equation can be recovered (for  $r = 1, s = 0$ ).

**Example 3.2. Simplified one dimensional replicator-mutator.** *Consider the 1d linear-Gaussian filtering problem with trivial signal dynamics, i.e.  $g(x) = 0, \Sigma = 0, h(x) = Hx, \Xi = 1$ . In this case, the unnormalised crow-kimura replicator-mutator equation (3.8) with  $r = 1, s = 0$  takes the form,*

$$\partial_t \mu_t^d(x) = -\frac{1}{2} (Hx)^2 \mu_t^d(x) + \mu_t^d(x) Hx \frac{dZ_t^d}{dt} \quad (3.9)$$

The Zakai equation (unnormalised filtering pde) (3.2) has the following (Stratonovich) representation,

$$\partial_t q_t(x) = -\frac{1}{2} (Hx)^2 q_t(x) + q_t(x) Hx \circ dZ_t \quad (3.10)$$

To help demonstrate why the limit of the smooth approximated noise in (3.9) must indeed be of Stratonovich type, consider the following pde driven by finite dimensional Ito noise,

$$\partial_t \rho_t(x) = -\frac{1}{2} (Hx)^2 \rho_t(x) + \rho_t(x) Hx dZ_t \quad (3.11)$$

The following numerical experiment demonstrates empirically the convergence of (3.9) to (3.10) rather than (3.11). A sequence of smooth observations over  $[0, T]$  with step size  $\delta_d$  is generated as

$$B_t^d = B_{t_n} + \left( \frac{t - t_n}{t_{n+1} - t_n} \right) (B_{t_{n+1}} - B_{t_n}), \quad t \in [t_n, t_{n+1}) \quad (3.12)$$

where  $B_t$  corresponds to a Brownian motion so that  $(B_{t_{n+1}} - B_{t_n}) \sim \mathcal{N}(0, \delta_d)$ . The obs increment  $\frac{dZ_t^d}{dt}$  is then defined as in (3.4). With the approximation (3.12),  $\frac{dZ_t^d}{dt}$  is constant in the time interval  $[t_n, t_{n+1})$  for every  $n = 0, 1, 2, \dots$ , and to emphasise the lack of dependence on  $t$ , we denote it by a random variable  $\xi_n$  where

$$\xi_n \sim \mathcal{N}\left(Hx_0^*, \frac{1}{\delta_d}\right)$$

for some fixed  $x_0^*$  denoting the true hidden state at time 0. By interpreting (3.9) as a linear pde of the form

$$\partial_t \mu_t^d(x) = A_t(x) \mu_t^d(x)$$

with piecewise smooth in time coefficient  $A_t(x) := -\frac{1}{2}(Hx)^2 + Hx\xi_n$  for  $t \in [t_n, t_{n+1})$ , it can be discretised via the usual euler scheme over  $[0, T]$  with time step  $\Delta$  and  $t_{i+1} - t_i = \Delta t < \delta_d$  for all  $i = 0, 1, 2, \dots$ . Let  $\tilde{\mu}_i(x)$  denote the approximation to  $\mu_t^d(x)$  at  $t = t_i$ ,

$$\tilde{\mu}_{i+1}(x) = \tilde{\mu}_i(x) + \Delta t A_i(x) \tilde{\mu}_i(x)$$

with  $A_i(x) = -\frac{1}{2}(Hx)^2 + Hx\xi_{n_i}$  for  $n_i$  such that  $t_i \in [t_{n_i}, t_{n_i+1})$ . Similarly, (3.10) and (3.11) can be simulated with Euler-Maruyama schemes with the same time step  $\Delta t$ . There the observation path is a solution of  $dZ_t = Hx_0^* dt + dB_t$ , simulated at a fine time interval. Figure 3.1 shows the results for a single time instant. Importantly, the crow-kimura replicator equation (3.7) (cyan line) coincides closely with (3.10) (black line) (up to normalisation), while the Ito version (3.11) (pink line) is significantly different.

The following theorem establishes the convergence of the replicator-mutator equation (in the form (3.7), as identified in Lemma 3.1) to a “generalised” form of the kushner-stratonovich equation from nonlinear filtering as  $d \rightarrow \infty$ . Convergence is studied via the unnormalised equations as this greatly simplifies the analysis but still yields the overall conclusion relating filtering and replicator-mutator equations due to the one to one correspondence between the unnormalised and normalised equations. Convergence to the standard filtering equations for the specific choice  $r = 1, s = 0$ ; the benefits of the generalised form (ie. when  $s \neq 0, r \neq 1$ ) will be further explored in the context of misspecified filtering in Section 4. The proof of Theorem 3.1 borrows many elements from the proof of Theorem 3.1 in [HKX02] and makes use of the representation formulae developed in [Kun82]. We extend their work to consider unbounded observation drifts  $h$  (where they had assumed uniform boundedness) and to the case of multivariate rather than scalar valued observations  $Z_t$ . Due to the representation formula used here, we do not need to rely on strong convergence of piecewise smooth approximations with unbounded diffusion coefficients as developed in e.g. [Pat24] (this aspect is discussed more specifically in the proof below). The price paid is that we focus on pointwise convergence of the density functions, rather than stronger  $L^p$  convergence, i.e. (i.e.  $\mathbb{E}[\|\mu_t^d - q_t(x)\|_p^p] \rightarrow 0$  as  $d \rightarrow \infty$  where  $\|f\|_p^p := \int_{\mathbb{R}^d} |f(x)|^p dx$ ). The weaker mode of convergence is still useful, particularly given that it allows us to relax restrictive assumptions on  $h$  which previously did not even cover the linear-Gaussian setting. The proof of the following theorem can be found in section 5.3.

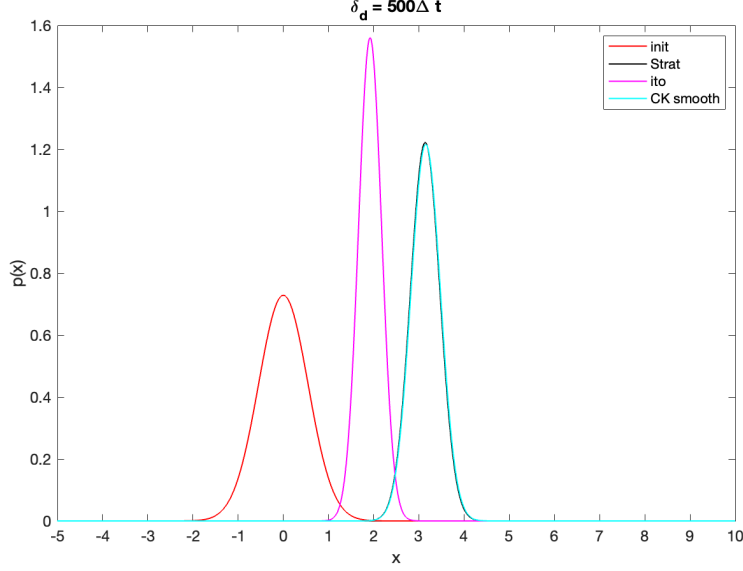


Figure 3.1: Snapshot in time for the above filtering problem with  $H = 2, m_0 = 0, P_0 = 0.3, x_0^* = 5, \Xi = 1$  and the smooth observations are constructed with  $\delta_d = 500\Delta t$ . Clearly the Ito interpretation (pink line) is not the correct limit for the crow-kimura with smooth approximation. The correct stratonovich interpretation (black line) aligns closely with the crow-kimura with smooth obs (cyan line). Note that the stratonovich interpretation (3.10) coincides with the familiar zakai equation from filtering (3.2).

**Theorem 3.1.** Assume that  $g(x) : \mathbb{R}^m \rightarrow \mathbb{R}^m$  and  $h(x) : \mathbb{R}^m \rightarrow \mathbb{R}^n$  are  $C^2$ , globally Lipschitz continuous functions satisfying linear growth conditions, i.e. there exists a constant  $C > 0$  such that

$$\begin{aligned} \|h(x) - h(y)\| + \|g(x) - g(y)\| &\leq C\|x - y\|, \quad x, y \in \mathbb{R}^m \\ \|h(x)\| + \|g(x)\| &\leq C(1 + \|x\|), \quad x \in \mathbb{R}^m \end{aligned}$$

Let  $\Sigma$  and  $\Xi$  be  $m \times m$  and  $n \times n$  positive definite matrices respectively. Denote by  $\mu_t^d(x)$  the solution to the unnormalised Crow Kimura replicator-mutator equation with time-varying fitness landscape (3.5) with  $s < r, r > 0$

$$\begin{aligned} \partial_t \mu_t^d(x) &= \mathcal{L}^* \mu_t^d(x) + \mu_t^d(x) \mathbb{E}_{z \sim \mu_t^d} [f_t(x, z)] \\ &= \mathcal{L}^* \mu_t^d(x) - \frac{r}{2} \left\| h(x) - \frac{dZ_t^d}{dt} \right\|_{\Xi}^2 \mu_t^d(x) + s \left\langle h(x) - \frac{dZ_t^d}{dt}, h(z) - \frac{dZ_t^d}{dt} \right\rangle_{\Xi} \mu_t^d(x) \end{aligned} \quad (3.13)$$

Let  $q_t(x)$  denote the solution of the modified Zakai equation (presented here in Ito form),

$$dq_t = \mathcal{L}^* q_t(x) - \frac{s}{2} h(x)^\top \Xi^{-1} h(x) q_t(x) + (r - s) q_t(x) h(x)^\top \Xi^{-1} dZ_t. \quad (3.14)$$

Suppose also that  $q_0(x) = \mu_0^d(x) = f(x)$  where  $f$  is a uniformly bounded  $C^\infty$  probability density function. Then for any  $T > 0$ ,

$$\lim_{d \rightarrow \infty} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |\mu_t^d(x) - q_t(x)|^p \right] = 0, \quad \forall x \in \mathbb{R}^m \quad (3.15)$$

for  $p \geq 1$  and  $r, s$  additionally satisfying  $r - s < \frac{2}{iC(1+\mathbb{E}^\mathbb{Q}[|x|^2])p\lambda_{\Xi}^2 r_1 r_2}$ , where  $r_1, r_2 > 1$  and  $1/r_1 + 1/r_2 = 1$  and  $C$  depends on the linear growth constant of  $h(x)$ . Importantly, when  $r = 1, s = 0$ , (3.13) converges to the standard Zakai equation (3.2).

## 4 Replicator-mutator equations & filtering with misspecified models: the Linear-Gaussian case

Throughout this section, we use the terminology *local* and *non-local* replicator-mutator equation to refer to (3.7) with  $s = 0, r > 0$  and  $s \neq 0, r > 0$ , respectively. The terminology is motivated by the fact that the case  $s \neq 0$  introduces a non-local term into the fitness function. Theorem 3.1 shows that  $r = 1, s = 0$  in the crow-kimura replicator-mutator equation (3.7) coincides with the non-linear filtering equation when the system dynamics is known perfectly. In this section, we focus on the linear-Gaussian setting to demonstrate both analytically and numerically the benefits of the non-local replicator-mutator model for inference in the presence of model misspecification. Additionally, we show in Section 4.1 that the local replicator-mutator equation with  $r \neq 1$  corresponds to the familiar covariance inflated ensemble Kalman-Bucy filter.

Before detailing these main insights, we first present the fundamental equations in the Linear-Gaussian setting and establish some useful findings on mean-field models corresponding to (3.7). Firstly, consider the (normalised) linear-Gaussian crow-kimura replicator-mutator equation in the limit  $\delta_d \rightarrow 0$  as derived in Theorem 3.1. That is, consider the normalised form of (3.14) with  $h(x) = Hx$ ,  $g(x) = Gx, p_0 = \mathcal{N}(x; m_0, C_0)$ ,

$$\begin{aligned} dp_t(x) = \mathcal{L}^* p_t(x) dt + (r - s) p_t(x) & \left( -(Hx - Hm_t)^\top \Xi^{-1} H m_t dt + (Hx - Hm_t)^\top \Xi^{-1} dZ_t \right) \\ & - \frac{s}{2} p_t(x) \left( \|Hx - Hm_t\|_{\Xi}^2 - \mathbb{E}_{p_t}[\|Hx - Hm_t\|_{\Xi}^2] \right) dt \end{aligned} \quad (4.1)$$

noticing that this coincides with the standard Kushner equation when  $r = 1, s = 0$ . The form with  $\delta_d > 0$  is given by

$$\begin{aligned} dp_t(x) = \mathcal{L}^* p_t(x) dt + (r - s) p_t(x) & \left( -\frac{1}{2} \left( x^\top H^\top \Xi^{-1} Hx - \mathbb{E}_{p_t} \left[ x^\top H^\top \Xi^{-1} Hx \right] \right) + (Hx - Hm_t)^\top \Xi^{-1} \frac{dZ_t^d}{dt} \right) \\ & - \frac{s}{2} p_t(x) \left( \|Hx - Hm_t\|_{\Xi}^2 - \mathbb{E}_{p_t}[\|Hx - Hm_t\|_{\Xi}^2] \right) dt \end{aligned} \quad (4.2)$$

where in the remainder of this section, we use the notation  $m_t := \mathbb{E}_{p_t}[x]$  and  $C_t = \mathbb{E}_{p_t}[xx^\top] - \mathbb{E}_{p_t}[x]\mathbb{E}_{p_t}[x^\top]$  for the mean and covariance respectively. It can be shown straightforwardly that the evolution of the mean and covariance of the crow-kimura replicator-mutator equation (4.1) is given by

$$dm_t = Gm_t dt + (r - s) C_t H^\top \Xi^{-1} (dZ_t - Hm_t dt) \quad (4.3)$$

$$\dot{C}_t = GC_t + C_t G^\top + \Sigma - r C_t H^\top \Xi^{-1} H C_t \quad (4.4)$$

and also that the mean equation for the case  $\delta_d > 0$  (with pde given by (4.2)) is given by

$$\dot{m}_t = Gm_t + (r - s) C_t H^\top \Xi^{-1} \left( \dot{Z}_t^d - Hm_t \right) \quad (4.5)$$

and the covariance equation coinciding with (4.4). Notice in particular that while both  $s$  and  $r$  affect the evolution of the mean, only  $r$  affects the evolution of the covariance. This aspect will be discussed

further in the context of misspecified filtering in section 4.2. In the below lemma, we consider mean-field processes associated to the replicator-mutator equation. Such mean-field processes arise in interacting particle implementations of Kalman Bucy filtering (the so-called ensemble Kalman-Bucy methods, in both the stochastic form [Lee20; HM97] and deterministic forms [BR12; Tag+17]). The mean-field process helps to demonstrate that the drift term involving  $s$  acts as a covariance deflation for  $s < 0$  and an additional inflation term for  $s \in (0, r)$ . The proof of the lemma can be found in section 5.4. We will utilise this characterisation to establish connections to covariance inflated Kalman-Bucy filtering specifically in section 4.1.

**Lemma 4.1.** *Consider the multivariate linear-Gaussian setting as described above and the following mean-field SDEs with  $s < r, r > 0$*

$$d\bar{X}_t = \underbrace{G\bar{X}_t + \Sigma^{1/2}dW_t}_{\text{mutation}} - \frac{s}{2}C_tH^\top \Xi^{-1}(H\bar{X}_t - Hm_t)dt + \underbrace{C_tH^\top \left(\frac{1}{r-s}\Xi\right)^{-1} \left(dZ_t - H\bar{X}_t dt + \left(\frac{1}{r-s}\Xi\right)^{1/2} d\bar{W}_t\right)}_{\text{Stochastic Kalman Innovation}} \quad (4.6)$$

$$d\bar{X}_t = \underbrace{G\bar{X}_t + \Sigma^{1/2}dW_t}_{\text{mutation}} - sHC_t\Xi^{-1}(dZ_t - Hm_t dt) + \underbrace{C_tH^\top \left(\frac{1}{r-s}\Xi\right)^{-1} \left(dZ_t - \frac{1}{2}(H\bar{X}_t + Hm_t)dt\right)}_{\text{Deterministic Kalman Innovation}} \quad (4.7)$$

where  $\bar{W}$  is a scalar Wiener process independent of  $W, B$  and  $\bar{X}_0 \sim \mathcal{N}(x; m_0, C_0)$ . The time evolution of the conditional density of  $\bar{X}_t$  given  $\mathcal{Z}_t$  for both processes (4.6) and (4.7) is given by (4.1). We may similarly define a mean-field process with smooth observations of the form

$$d\bar{X}_t = \underbrace{G\bar{X}_t + \Sigma^{1/2}dW_t}_{\text{mutation}} - \frac{s}{2}C_tH^\top \Xi^{-1}(H\bar{X}_t - Hm_t)dt + \underbrace{C_tH^\top \left(\frac{1}{r-s}\Xi\right)^{-1} \left(\frac{dZ_t^d}{dt} - H\bar{X}_t\right) dt + C_tH^\top \left(\frac{1}{r-s}\Xi\right)^{-1/2} d\bar{W}_t}_{\text{Stochastic Kalman Innovation}} \quad (4.8)$$

and similarly for the deterministic update version (4.7). The conditional density of (4.8) (as well as the deterministic update version) evolves according to the crow-kimura replicator-mutator equation given by (4.2).

It is well known that the Kalman-Bucy filter is the minimum variance unbiased estimator of  $X_t$ ,  $t > 0$ . This property carries over to the crow-kimura replicator mutator equation for  $s = 0, r = 1$  due to the pre-established equivalence to the Kalman Bucy filter. In the following result, we show that the unbiasedness property holds more generally for the replicator mutator equation (4.2), i.e. for any choice  $s < r, r > 0$ , when the system parameters are known perfectly. Obviously both from inspection of (4.4) and also from Lemma 4.1, it is clear that the minimum variance property is destroyed unless  $r = 1$ .

**Lemma 4.2. Unbiasedness of the non-local replicator-mutator with perfect system.** *Consider the signal-observation pair (1.5)-(1.6) with  $g(x) = Gx$ ,  $h(x) = Hx$  and Gaussian initial conditions. If  $m_0 = \mathbb{E}[X_0]$  then the generalised Kalman-Bucy filter with  $s < r, r > 0$  is unbiased, i.e.  $\mathbb{E}[m_t] = \mathbb{E}[X_t]$ ,  $t > 0$ .*

*Proof.* Begin with the mean equation from the generalised Kalman-Bucy filter

$$\begin{aligned} m_t &= m_0 + \int_0^t Gm_u du + (r-s) \int_0^t K_u dZ_u - (r-s) \int_0^t K_u H m_u du \\ &= m_0 + \int_0^t Gm_u du + (r-s) \int_0^t K_u H (X_u^* - m_u) du + (r-s) \int_0^t K_u \Xi^{1/2} dB_u \end{aligned}$$

where  $K_t := C_t H^\top \Xi^{-1}$  corresponds to the Kalman gain at time  $t$  and we have substituted in the form of the observations in the second equality. Then

$$\mathbb{E}[m_t] = m_0 + \int_0^t G\mathbb{E}[m_u] du + (r-s) \int_0^t K_u H (\mathbb{E}[X_u^*] - \mathbb{E}[m_u]) du$$

Also recall that

$$\mathbb{E}[X_t^*] = \mathbb{E}[X_0^*] + \int_0^t G\mathbb{E}[X_u^*] du$$

then

$$\mathbb{E}[m_t] - \mathbb{E}[X_t^*] = m_0 - \mathbb{E}[X_0^*] + \int_0^t (G - (r-s)K_s H) (\mathbb{E}[m_s] - \mathbb{E}[X_s^*]) ds$$

which corresponds to an ODE of the form  $\dot{y}_t = A_t y_t$  with  $y_t := \mathbb{E}[m_t] - \mathbb{E}[X_t^*]$ , which has solutions  $y_t = 0$ ,  $t > 0$  whenever  $y_0 = 0$ .  $\square$

#### 4.1 Replicator-mutator with $s = 0, r \neq 1$ & Inflated Kalman-Bucy filtering

The mean-field processes presented in Lemma 4.1, we see that in the linear-Gaussian setting, the replicator-mutator equation with  $s \neq 0$  involves both an adjustment of the observation error covariance  $\Xi$  as well as the inclusion of the term  $\frac{s}{2} C_t H^\top \Xi^{-1} (H \bar{X}_t - H m_t)$  which either acts as a mean-reversion or mean-repulsion depending on the chosen value of  $s$ . A similar term also appears in [WRS18] where the ensemble Kalman Bucy filter with so-called ‘‘deterministic noise’’ in place of the stochastic signal noise  $\Sigma$  is analysed. The use of such terms has a long history in the field of ensemble Kalman filtering, where it is more widely known as covariance inflation [And07; DSH20; MH00; TMK16; BD23]. Covariance inflation is an important heuristic tool used to improve numerical stability of the ensemble Kalman filter when the number of samples is low [BMP18; BD19] and also to account for model errors.

The below lemma establishes the equivalence of the crow-kimura replicator mutator equation for both  $\delta_d > 0$  and the limiting  $\delta_d \rightarrow 0$  case, i.e. (4.2) and (4.1) respectively, to a form of covariance inflation in the literature. The proof of the lemma can be found in section 5.5.

**Lemma 4.3.** *Consider the following mean field description of covariance inflated ‘‘stochastic’’ ensemble Kalman-Bucy filter see e.g. [BMP18; BD23],*

$$d\bar{X}_t = G\bar{X}_t + \Sigma^{1/2} dW_t + (C_t + \epsilon T) H^\top \Xi^{-1} \left( dZ_t - H \bar{X}_t dt + \Xi^{1/2} d\bar{W}_t \right) \quad (4.9)$$

*and the mean field description of covariance inflated ‘‘deterministic’’ ensemble Kalman-Bucy filter [BD23] with piecewise smooth observations,*

$$\frac{d\bar{X}_t}{dt} = G\bar{X}_t + \Sigma^{1/2} dW_t + (C_t + \epsilon T) H^\top \Xi^{-1} \left[ \frac{dZ_t^d}{dt} - H \left( \frac{\bar{X}_t + m_t}{2} \right) \right] \quad (4.10)$$

*where  $\epsilon > 0$  is a tuning parameter and  $T$  is a reference matrix guiding the inflation. The corresponding evolution pde describing the conditional density  $p(X_t | Z_t)$  for (4.9) and (4.10) is formally equivalent to (4.1) and (4.2) respectively, for the choices  $T = C_t$ ,  $r = 1 + \epsilon$  and  $s = 0$ .*



## 4.2 Local vs non-local replicator-mutator for misspecified model filtering

We now turn our attention to analysing the performance of the replicator-mutator equation for a filtering problem with misspecified signal dynamics. This section focuses on the  $\delta_d \rightarrow 0$  form of the replicator-mutator, (4.1), although the conclusions of this section are expected to similarly hold for  $\delta_d > 0$ . The misspecified model filtering problem is as follows. Consider the following linear-gaussian problem where the hidden state  $X_t \in \mathbb{R}^m$  evolves according to

$$dX_t = GX_t dt + bdt + \Sigma^{1/2} dW_t \quad (4.11)$$

with  $X_0 \sim \mathcal{N}(m_0, P_0)$  and  $b$  a constant vector. The hidden dynamics is known imperfectly and that the assumed model for the trait is instead

$$dX_t = GX_t dt + \Sigma^{1/2} dW_t \quad (4.12)$$

(i.e. (4.11) with  $b = 0$ ), and the observation process is as in (1.6). The replicator-mutator equation (4.2) is no longer expected to produce unbiased estimates due to the presence of the  $b$  term.

Firstly, we introduce the following important variables. The tracking error at time  $t$  is denoted by  $\varepsilon_t$ ,

$$\varepsilon_t := m_t - x_t^*$$

where  $m_t := \int x p_t(x) dx$  and  $p_t(x)$  is the solution of (4.1) and  $X_t^*$  the solution of (4.11), also known as the reference trajectory or true hidden state. Denote also by  $P_t$  the error covariance matrix,

$$P_t := \mathbb{E}[(\varepsilon_t - \mathbb{E}[\varepsilon_t])(\varepsilon_t - \mathbb{E}[\varepsilon_t])^T]$$

Note well that  $P_t$  is distinct from  $C_t$  which we use to denote the covariance from the replicator-mutator equation, also the covariance of the filtering equations. Recall that in the perfect knowledge Kalman Bucy filtering setting, since  $\mathbb{E}[\varepsilon_t] = 0$ , we have that  $C_t = \tilde{P}_t = P_t$ . When  $b \neq 0$ , even a standard Kalman-Bucy filter no longer produces an unbiased estimates of the hidden state, so that  $\tilde{P}_t \neq P_t$  and also,  $C_t \neq P_t$ . Finally, let

$$\nu_t := \|\mathbb{E}[\varepsilon_t]\|^2, \quad E_t := \mathbb{E}[\|\varepsilon_t\|^2]$$

denote the squared expected error (or squared bias) and expected squared error (mean squared error), respectively. Notice that  $\nu_t = \text{Tr}(\mathbb{E}[\varepsilon_t]\mathbb{E}[\varepsilon_t]^T)$ . We also define

$$\tilde{P}_t := \mathbb{E}[\varepsilon_t \varepsilon_t^T] \quad (4.13)$$

and notice that  $\text{Tr}(\tilde{P}_t) = E_t$ . Recall that when  $b = 0$ , we have that  $\mathbb{E}[\varepsilon_t] = 0$  for all  $t$  from lemma 4.2, so that  $P_t = \tilde{P}_t$ . Since our focus is on the case  $b \neq 0$ , it is necessary to further study  $\tilde{P}_t$  in its own right. Additionally, since  $\tilde{P}_t = P_t + \mathbb{E}[\varepsilon_t]\mathbb{E}[\varepsilon_t]^T$  the following bias-variance decomposition holds

$$\underbrace{E_t}_{=\text{m.s.e}} = \underbrace{\text{Tr}(P_t)}_{=\text{total error var.}} + \underbrace{\nu_t}_{=\text{squared bias}}$$

Finally, we use the shorthand notation

$$K_t := C_t H^T \Xi^{-1}$$

which for the case  $r = 1$  coincides with the familiar Kalman Gain from the Kalman-Bucy filter. The following lemma characterises the time evolution of bias, variance and mean squared error. As expected, the error variance  $P_t$  evolves independently of the unknown term  $b$ . The proof of the lemma can be found in section 5.6

**Lemma 4.4.** *Assume the system properties described by (4.11) for the true hidden state, (4.12) for the assumed hidden state and (1.6) for the observation model. Given  $\Xi^{-1/2}H$  an invertible matrix and  $r, s$  such that  $r < s, r > 0$ , we have the following evolution equations for the error covariance  $P_t$  and expected squared error  $\tilde{P}_t$ ,*

$$\frac{dP_t}{dt} = (G - (r - s)K_t H)P_t + P_t(G - (r - s)K_t H)^\top + \Sigma dt + (r - s)^2 K_t \Xi K_t^\top \quad (4.14)$$

$$\frac{d\tilde{P}_t}{dt} = (G - (r - s)K_t H)\tilde{P}_t + \tilde{P}_t(G - (r - s)K_t H)^\top + \Sigma + (r - s)^2 K_t \Xi K_t^\top - \mathbb{E}[\varepsilon_t]b^\top - b\mathbb{E}[\varepsilon_t^\top] \quad (4.15)$$

For the special case  $s = 0, r = 1$ , it holds that  $P_t = C_t, \forall t > 0$  if  $P_0 = C_0$  where  $C_t$  is the solution of the covariance equation of the replicator-mutator.

Furthermore, for any p.d.  $C_0$ , the evolution of the mean squared error  $E_t$  satisfies the following inequality

$$\frac{dE_t}{dt} \leq 2\|(A_t(r, s))\|_F E_t - 2Tr(\mathbb{E}[\varepsilon_t]b^\top) + Tr(\Sigma) + (r - s)^2 \lambda_{max}(H^\top \Xi^{-1}H)\|C_t\|_F \quad (4.16)$$

where  $A_t(r, s) := G - (r - s)K_t H$ .

To obtain further insights on optimal choices of  $r, s$ , we consider a simplified setting where  $C_0 = C_\infty$  (i.e. where the covariance is initialised at the steady state covariance matrix in (4.4)). This setting is still rich enough to provide insights on the role of  $r, s$  in the non-local replicator-mutator, particularly as we are primarily interested in the time asymptotic behaviour of mean squared error. Whilst the calculations in the previous section are applicable in the multivariate setting, from now on we focus purely on the fully scalar case, and leave the multivariate setting to future work. For the remainder of the section, we use  $\tilde{P}_t$  to refer to the mean square error. The following lemma gives explicit representations of the time asymptotic squared bias  $\nu_\infty$  and mean squared error  $E_\infty$  where  $E_t = \tilde{P}_t$  in the scalar case. The proof of the lemma can be found in section 5.7.

**Lemma 4.5. Steady state Bias-Variance.** *Consider the scalar setting where  $m = n = 1$ . Assume  $\mathbb{E}[\varepsilon_0 = 0]$  and  $C_0 = C_\infty$  where  $C_\infty$  satisfies*

$$0 = GC_\infty + C_\infty G^\top + \Sigma - rC_\infty H^\top \Xi^{-1}HC_\infty \quad (4.17)$$

Suppose  $r, s$  are specified such that  $A_\infty < 0$ . Then as  $t \rightarrow \infty, \nu_t \rightarrow \nu_\infty$  and  $\tilde{P}_t \rightarrow \tilde{P}_\infty$  where

$$\nu_\infty = \left(\frac{b}{A_\infty(r, s)}\right)^2 \quad (4.18)$$

$$\tilde{P}_\infty = -\frac{1}{2} \left( \Sigma + \left(\frac{G - A_\infty(s, r)}{H}\right)^2 \Xi \right) \frac{1}{A_\infty(s, r)} + \left(\frac{b}{A_\infty(s, r)}\right)^2 \quad (4.19)$$

and

$$A_\infty(s, r) = \frac{s}{r}G + \frac{(r - s)}{r}A_\infty(0, r) \quad (4.20)$$

$$A_\infty(0, r) = -\sqrt{G^2 + rH^2\Xi^{-1}\Sigma} \quad (4.21)$$

We are now ready to characterise the optimal  $r, s$  values that minimise mean square error. Recall that the minimal conditions on  $s, r$  are that  $r > 0$  and  $s < r$ . The below lemma establishes a range of allowable values of  $s$  in terms of  $r$  which guarantee the existence of a stable mean squared error value,  $\tilde{P}_\infty$ . Importantly, the lemma shows that there is not one unique pair  $(r, s)$  that minimises the asymptotic mean square error for a given system, but rather infinitely many pairs satisfying (4.26). These estimates hold true regardless of the stability characteristics of the hidden state and observation dynamics. Additionally, we demonstrate that for  $s \in (s^l, s^u)$  where  $s^l, s^u$  are as defined in the below lemma, there exists two possible  $r$  values for a given  $s$  that will yield the minimal asymptotic mean square error. This is particularly beneficial in terms of allowing for a realistic  $C_\infty$ , as will be explored further in Lemma 4.7. The proof of lemma 4.6 can be found in section 5.8.

**Lemma 4.6. Optimal  $r, s$  minimising m.s.e.** *Adopt the same conditions as in Lemma 4.5. Define,*

$$p := -(H^2 \Xi^{-1} \Sigma + G^2), \quad q := 4b^2 H^2 \Xi^{-1} \quad (4.22)$$

$$\tau := \frac{q^2}{4} + \frac{p^3}{27} \quad (4.23)$$

Suppose that  $G, H, \Sigma, \Xi, b$  are such that  $\tau \neq 0$ . Finally, define  $A_\infty^*$  depending on whether  $\tau < 0$  or  $\tau >$ , as

$$A_\infty^* := \begin{cases} \left(-\frac{q}{2} + \sqrt{\tau}\right)^{1/3} + \left(-\frac{q}{2} - \sqrt{\tau}\right)^{1/3}, & \tau > 0 \\ 2 \sqrt{-\frac{p}{3}} \cos \left[ \frac{1}{3} \cos^{-1} \left( \frac{3q}{2p} \sqrt{\frac{-3}{p}} \right) - \frac{4\pi}{3} \right], & \tau < 0 \end{cases} \quad (4.24)$$

Then the admissible values of  $s$  for any  $r > 0$  satisfy

$$s^l := -\frac{(G + A_\infty^*)^2}{4H^2 \Xi^{-1} \Sigma} < s < \min \left( r, \frac{\sqrt{G^2 + rH^2 \Xi^{-1} \Sigma}}{G + \sqrt{G^2 + rH^2 \Xi^{-1} \Sigma}} r \right) \quad (4.25)$$

where the upper bound guarantees the existence of  $\tilde{P}_\infty$ . The optimal  $s$  for any given  $r > 0$  is given by

$$s^{opt} = \frac{r(A_\infty^* + \sqrt{G^2 + rH^2 \Xi^{-1} \Sigma})}{G + \sqrt{G^2 + rH^2 \Xi^{-1} \Sigma}} \quad (4.26)$$

Alternatively, for any  $s \in (s^l, s^u)$  where  $s^u := \frac{G^2 + (A_\infty^* - G)|G| - GA_\infty^*}{H}$ , there are two possible values of  $r$  optimising mse, given by

$$r^{opt} = \frac{1}{4H^2 \Xi^{-1} \Sigma} \left( G - A_\infty^* \pm \sqrt{(G - A_\infty^*)^2 + 4(GA_\infty^* + sH^2 \Xi^{-1} \Sigma)} \right)^2 - \frac{G^2}{H^2 \Xi^{-1} \Sigma} \quad (4.27)$$

and for any  $s^u < s < \min \left( r, \frac{\sqrt{G^2 + rH^2 \Xi^{-1} \Sigma}}{G + \sqrt{G^2 + rH^2 \Xi^{-1} \Sigma}} r \right)$ , there is a unique optimal  $r$ , given by

$$r^{opt} = \frac{1}{4H^2 \Xi^{-1} \Sigma} \left( G - A_\infty^* + \sqrt{(G - A_\infty^*)^2 + 4(GA_\infty^* + sH^2 \Xi^{-1} \Sigma)} \right)^2 - \frac{G^2}{H^2 \Xi^{-1} \Sigma} \quad (4.28)$$

Importantly, the above lemma shows that there are infinitely many  $(s, r)$  satisfying (4.26),  $r > 0, s < r$  and  $A_\infty < 0$ , and these will all yield the minimum mean squared error  $\tilde{P}_\infty$  as  $t \rightarrow \infty$ , as well as the same error variance  $P_\infty$  and squared bias  $\nu_\infty$  (since each of these terms depend only on  $A_\infty$ ). However,  $C_\infty$  will clearly be different, due to the dependence on the chosen  $r$  value (see

(5.20)). As can be seen in the next lemma (and also in section 4.3), larger values of  $r$  lead to smaller  $C_\infty$ . From an inference perspective, smaller  $C_\infty$  indicates greater confidence in the estimator, which can be problematic especially when  $C_\infty$  is smaller than the minimum covariance achievable in the perfect model setting. To analyse this phenomenon further, define

$$\hat{C}_\infty := \frac{G + \sqrt{G^2 + H^2 \Xi^{-1} \Sigma}}{H^2 \Xi^{-1}} \quad (4.29)$$

which is the steady state covariance coinciding with the Kalman-Bucy filter (the optimal filter) in the case of perfect knowledge, (i.e. when  $r = 1, s = 0, b = 0$ ). Any choice of  $r$  for which  $C_\infty < \hat{C}_\infty$  should be avoided when there is model misspecification, as we cannot expect to be more confident than when we have no misspecification. Recall that in the perfect knowledge setting, we have that  $\tilde{P}_\infty = \hat{C}_\infty$ , in other words, the asymptotic covariance from the Kalman-Bucy filter coincides with the mean squared error. The following lemma establishes a relationship between  $r, s$  that ensures the specific pair satisfies  $C_\infty = \tilde{P}_\infty$  in the misspecified model setting where  $b \neq 0$ . This highlights that the non-local replicator-mutator equation, unlike the regular inflated ensemble Kalman method, is capable of simultaneously minimise mean squared error and providing realistic uncertainty estimates through  $C_\infty$ . In particular, (4.31) in the below lemma demonstrates that the standard inflated ensemble Kalman method requires  $r > 1$  to minimise mse. It will be demonstrated numerically in section 4.3 that this often coincides with an under-representation of the uncertainty (i.e.  $\frac{C_\infty(r_0^{opt})}{\hat{C}_\infty} \ll 1$ ). The last claim of the below lemma takes a step towards this claim analytically, by demonstrating that  $\frac{C_\infty(r)}{\hat{C}_\infty} \rightarrow 0$  as  $r \rightarrow \infty$  whenever  $s = 0$  (see (4.32)). The proof can be found in section 5.9.

**Lemma 4.7.** *Adopt the same conditions as in Lemma 4.5. Then the following holds,*

1. *The optimal  $(r, s)$  such that  $\tilde{P}_\infty = C_\infty$  satisfies (4.26) and*

$$r^{opt} - s^{opt} = \frac{(A_\infty^*)^2 (G - A_\infty^*)}{-0.5 A_\infty^* (H^2 \Xi^{-1} \Sigma + (G - A_\infty^*)^2) + b^2 H^2 \Xi^{-1}} \quad (4.30)$$

where  $A_\infty^*$  is given by (4.24).

2. *For the case  $\tau > 0$ , choosing  $s = 0$  corresponds to*

$$r_0^{opt} > \frac{2(G^2 + H^2 \Xi^{-1} \Sigma)}{H^2 \Xi^{-1} \Sigma} > 1 \quad (4.31)$$

*independently of the unknown  $b$  and for any  $G, H \neq 0$  and  $\Xi, \Sigma > 0$ . Since  $r_0^{opt} > 1$ , this also means that*

$$\frac{C_\infty(r_0^{opt})}{\hat{C}_\infty} < \frac{1}{\sqrt{r_0^{opt}}} \left( \frac{2|G| + \sqrt{H^2 \Xi^{-1} \Sigma}}{(G + \sqrt{G^2 + H^2 \Xi^{-1} \Sigma})} \right) \quad (4.32)$$

### 4.3 Numerical experiments

The following experiment aims to provide further insights on the role of  $s, r$  in (3.13) with fitness landscape  $f_t(x, z)$  given by (3.5) for the misspecified model setting from the previous section (which we repeat here for convenience). Although the analysis in the previous section has been done for the limiting case  $\delta_d \rightarrow 0$ , here we focus on the practically relevant discrete case and show that much

of the analysis holds for  $\delta_d$  small enough. That is, consider a partition of the time interval  $[0, T]$ ,  $0 < t_1 < t_2 \cdots < t_d = T$  with time-step  $t_{i+1} - t_i = \delta_d$ . Synthetic observations of the form

$$\frac{dZ_t^d}{dt} = Hx_{t_i}^*(\omega) + \Xi^{1/2} \frac{dB_t^d}{dt}, \quad t \in [t_i, t_{i+1}) \quad (4.33)$$

(as in (3.4)) are constructed, where  $x_t^*(\omega)$  is a solution for a realisation  $\omega$  at time  $t$  of

$$dX_t = GX_t dt + bdt + \Sigma^{1/2} dW_t \quad (4.34)$$

with  $X_0 \sim \mathcal{N}(m_0, P_0)$ . The process in (4.11) describes the evolution of the actual optimal trait and  $\frac{dZ_t^d}{dt}$  corresponds to noisy observations of it. Suppose the assumed model for the trait is instead

$$dX_t = GX_t dt + \Sigma^{1/2} dW_t \quad (4.35)$$

(i.e. (4.11) with  $b = 0$ ), so that the corresponding replicator-mutator equation takes the form

$$\partial p_t(x) = -G\nabla \cdot p_t(x) + \frac{1}{2} \nabla \cdot (\Sigma \nabla p_t(x)) + p_t(x) (\mathbb{E}_{z \sim p_t} [f_t(x, z)] - \mathbb{E}_{x, z \sim p_t} [f_t(x, z)]) \quad (4.36)$$

with  $f_t(x, z)$  given by (3.5). The remainder of this section will focus on the following experimental settings, which have been randomly generated.

Parameter	System 1 ( $\tau > 0$ )	System 2 ( $\tau < 0$ )
$G$	0.5	2.5
$H$	8.5	2.9
$\Sigma$	0.8	18
$\Xi$	6.3	26
$b$	9.9	1.2

Note that throughout, we assume  $C_0 = C_\infty$  given by (5.20). The settings in System 1 and 2 correspond to the case where  $\tau > 0$  (i.e. (5.23) has one real root) and  $\tau < 0$  (i.e. (5.23) has three real roots), respectively. Notice that in both systems, the hidden state evolves according to unstable dynamics and the crow- Kimura replicator mutator is capable of tracking an unstable signal as  $A_t < 0$ . We restrict the time domain to one where machine precision doesn't become an issue. We adopt  $\delta_d = 10^{-3}$  and use a simulation time step of  $10^{-4}$  to construct the true hidden state as well as to discretise the mean and covariance equation (4.3) and (4.4) using forward euler. We have the following main insights.

**Verification of lemma 4.6.** Figures 4.1 and 4.2 show the empirical estimate of the asymptotic mean square error  $\tilde{P}_\infty$  for different  $(r, s)$  pairs for system 1 and 2 respectively. The empirical mse at time  $t$ ,  $\mathcal{E}_t$ , was calculated as

$$\mathcal{E}_t = \frac{1}{N_s} \sum_{j=1}^{N_s} (m_t^j - x_t^*)^2 \quad (4.37)$$

where  $x_t^*$  is a single fixed realisation of (4.11) and  $m_t^j$  is a solution of (4.5) with  $C_0 = C_\infty$  and the index  $j$  referring to a single realisation of the smooth observation path  $Z_t^d$ . We used a total of with  $N_s = 5000$ . Figures 4.1 and 4.2 show that in both experiments, the analytic expressions in (4.26)

and for the smallest optimal  $s$  value,  $s^l$  as given in (4.25) matches quite closely. Notice also that in system 1, Figure 4.1 we see that there are two optimal  $r$  values for every  $s^l < s < 0$ , since 0 is the approx value of  $s^u$  in this case, as also described in Lemma 4.6. Conversely, for system 2 this only happens for  $0 < r < 1.09$  as  $s^l = -0.047$  is fairly close to zero, and is barely visible in the figure. Nevertheless, for system 1 in particular it is apparent that choosing  $s < 0$  allows to choose smaller  $r$  values that can simultaneously reduce mean square error and provide a realistic representation of uncertainty via  $C_\infty$ .

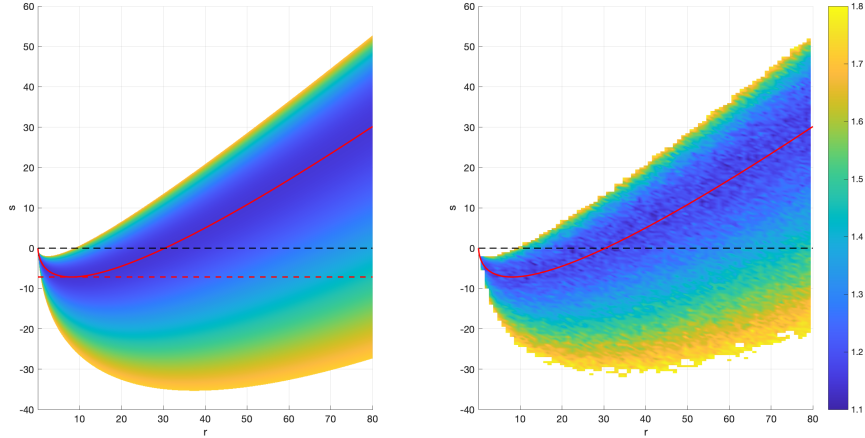


Figure 4.1: Plot of MSE for various  $s$  vs  $r$  values for System 1. The (optimal in terms of mse) values are indicated by the red line, calculated using (4.26). Colourbar shows corresponding values of the asymptotic MSE  $\tilde{P}_\infty$ . The dashed red line on the left plot shows the theoretical expression for  $s^l$  as in (4.25)

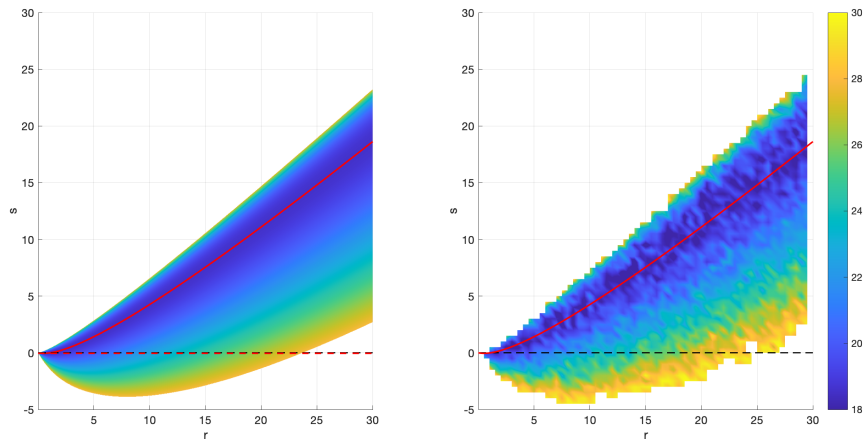


Figure 4.2: Plot of MSE for various  $s$  vs  $r$  values for System 2. The (optimal in terms of mse) values are indicated by the red line, calculated using (4.26). Colourbar shows corresponding values of the asymptotic MSE  $\tilde{P}_\infty$ .

**Uncertainty quantification with  $C_\infty$ .** Recall that we may further constrain the optimal  $(r, s)$  values by enforcing the requirement that  $C_\infty = \tilde{P}_\infty$ . This allows us to construct a filter for which the covariance of the estimate coincides with the actual mean squared error, as one obtains from the regular Kalman-Bucy filter in the perfect model setting. Figure 4.4 shows the variation in asymptotic covariance  $C_\infty$  for different  $r$  values, obtained from (5.20). Notice here that in the regular inflation case  $s = 0$ , the asymptotic covariance for the corresponding optimal  $r_0^{\text{opt}}$  is ( $\tilde{C}_\infty = 0.05$ ), which is considerably smaller than the covariance in the perfect model setting ( $\tilde{C}_\infty = 0.31$ ), indicating overconfidence in the estimation. The non-local replicator mutator on the other hand allows to obtain estimates that simultaneously minimise mse and provide a realistic representation of uncertainty. More specifically, the choice  $r = 0.13$  (which coincides with  $s^{\text{opt}} = -1.18$  gives a  $C_\infty$  which coincides with the MSE, so that the covariance produced by the estimation algorithm provides us with useful uncertainty quantification. In particular, it represents an increase in uncertainty over the perfect knowledge case (pink line), which should be reflected given the unknown bias in the system. Figure 4.3 also verifies the relation between  $r, s$  in (4.30) given in Lemma 4.7 (see cyan line). This extra criterion can be used to identify a single optimal  $(r, s)$  pair (as the cyan line intersects with the blue line at only one point, see Figure 4.3). In particular, for system 1, choosing  $s^{\text{opt}} = -1.18$  can also yield a corresponding  $r^{\text{opt}} = 27.9$  (see left plot in Figure 4.3). This  $r$  value is close to that of the  $s = 0$  case (regular covariance inflation), where it was shown that one obtains over-confident estimates. Similar conclusions can be drawn for System 2, but the optimal  $(r, s)$  such that  $C_\infty = \tilde{P}_\infty^{\text{opt}}$  is  $s^{\text{opt}} = -0.012, r^{\text{opt}} = 0.99$ , is not significantly different from the regular inflation case where  $s = 0, r^{\text{opt}} = 1.09$ . Similar to System 1, the other possible choice for  $r$  when  $s = -0.012, r = 0.07$ , coincides with an overestimation of uncertainty ( $C_\infty \approx 210$  vs  $\tilde{P}_\infty = 18$ ). We leave a study of specific system characteristics that would benefit most from the non-local replicator-mutator approach to future work.

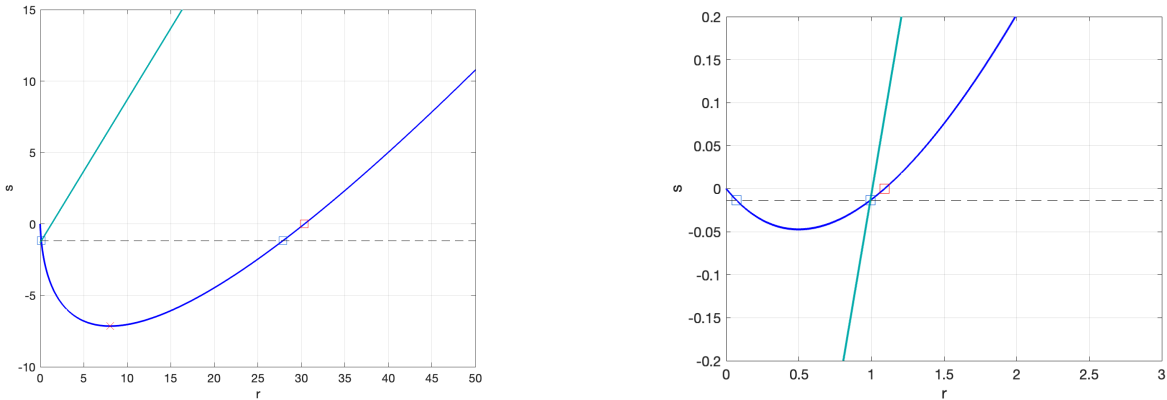


Figure 4.3: Plot of optimal  $(r, s)$  values for System 1 (left plot) and System 2 (right plot). The blue line indicates the  $(r, s)$  pairs minimising mse only, obtained from (4.26) and the cyan line indicates the  $(r, s)$  pairs such that  $C_\infty = \tilde{P}_\infty$ , obtained from (4.30). The point of intersection of the two lines indicates the  $(r, s)$  pair that achieves both. The red square indicates the optimal  $r$  value corresponding to the regular covariance inflation case ( $s = 0$ ). The blue squares indicate the possible  $r$  values corresponding to the optimal  $s$  in terms of both mse and  $C_\infty = \tilde{P}_\infty$ . In System 1, the standard inflation leads to overconfident estimates ( $C_\infty$  too small), whereas in System 2, it is not so far off from the optimal choice of  $s = -0.0135, r = 0.99$ .

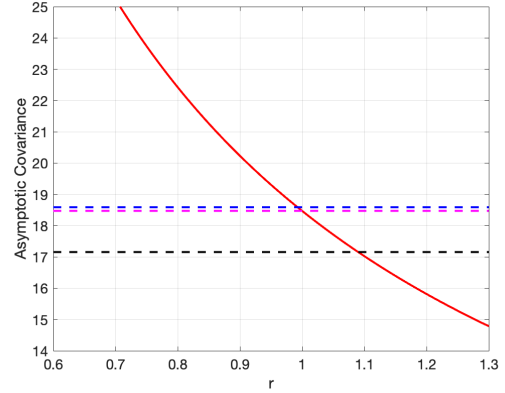
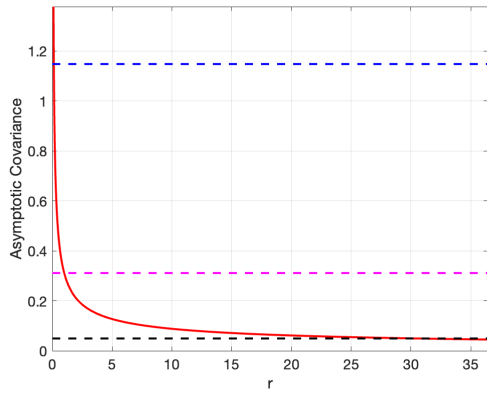


Figure 4.4: Demonstration of more realistic/representative covariances that can be obtained with the non-local replicator mutator, i.e. with  $s \neq 0$ . Left plot indicates system 1, right plot indicates system 2. The red solid line indicates  $C_\infty$  for  $r$  using (5.20). The blue horizontal dashed line indicates the value of  $\tilde{P}_\infty^{\text{opt}}$ , the minimum mean squared error, the pink horizontal dashed line indicates  $C_\infty$  for  $r = 1$  (i.e.  $\hat{C}_\infty$  as defined in (4.29) , the covariance in the perfect model case). Finally, the black dashed line indicates  $C_\infty$  for  $(s = 0, r^{\text{opt}})$ , i.e. the regular inflation case.



## 5 Proofs

### 5.1 Proof of Lemma 2.1

*Proof.* We now restrict  $\mathcal{P}$  to be the set of smooth probability density functions

$$\mathcal{P}(\mathbb{R}^d) := \left\{ p \in C^\infty(\mathbb{R}^d) : \int p(x) dx = 1, p \geq 0 \right\}$$

whose tangent space is given by<sup>1</sup>

$$T_p\mathcal{P}(\mathbb{R}^d) = \left\{ \sigma \in C^\infty(\mathbb{R}^d) : \int \sigma(x) dx = 0 \right\}$$

We start by computing the Fréchet derivative of (2.8) at  $p$ : Set  $\varepsilon > 0$  and  $q \in \mathcal{M}$ .

$$\begin{aligned} \frac{\mathcal{F}(q + \varepsilon\tilde{q}) - \mathcal{F}(q)}{\varepsilon} &= - \iint f(x, z) \frac{(q(x) + \varepsilon\tilde{q}(x))(q(z) + \varepsilon\tilde{q}(z)) - q(x)q(z)}{2\varepsilon} dz dx \\ &= -\frac{1}{2} \iint f(x, z) [q(x)\tilde{q}(z) + q(z)\tilde{q}(x) + \varepsilon\tilde{q}(x)\tilde{q}(z)] dz dx. \end{aligned}$$

This means that, because  $f$  is symmetric in its components,

$$D_q\mathcal{F}[\tilde{q}] = - \iint f(x, z) q(z)\tilde{q}(x) dz dx = - \int \pi_q(x)\tilde{q}(x) dx = -\langle \pi_q, \tilde{q} \rangle$$

which shows that

$$\mathcal{F}'(p) = -\pi_p$$

(this is to be understood as a linear operator). For the dissipation mechanism, consider the Fisher-Rao metric defined as

$$\begin{aligned} g_p^{FR}(\sigma_1, \sigma_2) &= \int \frac{\sigma_1}{p(x)} \frac{\sigma_2}{p(x)} dp(x) = \int \sigma_1 \frac{\sigma_2}{p(x)} dx \\ &\equiv \int \sigma_1 \frac{\sigma_2}{p(x)} dx - \int \sigma_1 dx \cdot \int \sigma_2 dx \end{aligned}$$

since  $\sigma \in T_p\mathcal{P}$  satisfies  $\int \sigma dx = 0$ . Its corresponding isomorphism/dual action  $\mathcal{G}^{FR}(p) : T_p\mathcal{P} \rightarrow T_p^*\mathcal{P}$  is given (by inspection) as

$$\mathcal{G}^{FR}(p)\sigma = \frac{\sigma}{p} - \int \sigma dx \tag{5.1}$$

The inverse mapping is given by

$$\mathcal{G}^{FR}(p)^{-1}\Phi = \left( \Phi - \int \Phi p(x) dx \right) p. \tag{5.2}$$

Straightforwardly, we have

$$-\mathcal{G}^{FR}(p)^{-1}\mathcal{F}'(p) = \left( \pi_p(x) - \int \pi_p(x)p(x) dx \right) p(x)$$

which yields the right-hand side of the replicator equation. □

---

<sup>1</sup>Actually, if  $p$  has a non-trivial support (e.g., there is an interval on which  $p$  vanishes), then the tangent space needs to be replaced by a tangent cone  $T_p\mathcal{P}(\mathbb{R}^d) = \left\{ \sigma \in C_c^\infty(\text{supp}(p)) : \int \sigma(x) dx = 0 \right\}$ , see [MS18], but we forgo the technical details here.

## 5.2 Proof of Lemma 3.1

*Proof.* The result follows from a simple re-arrangement of (3.13). Throughout, we use the shorthand notation  $y_t := \frac{dZ_t^d}{dt}$  and recall that  $\|v(x)\|_{\Xi}^2 := v(x)^\top \Xi^{-1} v(x)$  for vector valued functions  $v(x) : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . Firstly,

$$\begin{aligned} \mathbb{E}_{z \sim \rho_t}[f_t(x, z)] - \mathbb{E}_{\rho_t}[f_t] &= \int \left( -\frac{r}{2} \|h(x) - y_t\|_{\Xi}^2 + s \langle h(x) - y_t, h(z) - y_t \rangle_{\Xi} \right) \rho_t(z) dz \\ &\quad - \int \int \left( -\frac{r}{2} \|h(x) - y_t\|_{\Xi}^2 + s \langle h(x) - y_t, h(z) - y_t \rangle_{\Xi} \right) \rho_t(x) \rho_t(z) dx dz \\ &= -\frac{r}{2} (\|h(x) - y_t\|_{\Xi}^2 - \mathbb{E}_{\rho_t}[\|h - y_t\|_{\Xi}^2]) + s \langle h(x) - y_t, \mathbb{E}_{\rho_t}[h] - y_t \rangle_{\Xi} - s \|\mathbb{E}_{\rho_t}[h] - y_t\|_{\Xi}^2 \\ &=: I_1(x) + sI_2(x) + sI_3(x) \end{aligned}$$

For the first term, we have

$$\begin{aligned} I_1(x) &= -\frac{r}{2} \left( h(x)^\top \Xi^{-1} h(x) - h(x)^\top \Xi^{-1} y_t - y_t^\top \Xi^{-1} h(x) - \mathbb{E}_{\rho_t}[h^\top \Xi^{-1} h] + \mathbb{E}_{\rho_t}[h^\top] \Xi^{-1} y_t + y_t^\top \Xi^{-1} \mathbb{E}_{\rho_t}[h] \right) \\ &= -\frac{r}{2} \left( h(x)^\top \Xi^{-1} h(x) - \mathbb{E}_{\rho_t}[h(x)^\top \Xi^{-1} h(x)] \right) + r (h(x) - \mathbb{E}_{\rho_t}[h])^\top \Xi^{-1} y_t \end{aligned}$$

For the remaining terms, we have that

$$\begin{aligned} I_2(x) + I_3(x) &= \langle h - \mathbb{E}_{\rho_t}[h], \mathbb{E}_{\rho_t}[h] - y_t \rangle_{\Xi} \\ &= -\frac{1}{2} \|h - \mathbb{E}_{\rho_t}[h]\|_{\Xi}^2 - \frac{1}{2} \mathbb{E}_{\rho_t}[h^\top] \Xi^{-1} \mathbb{E}_{\rho_t}[h] + \frac{1}{2} h^\top \Xi^{-1} h - \frac{1}{2} \mathbb{E}_{\rho_t}[h^\top \Xi^{-1} h] \\ &\quad + \frac{1}{2} \mathbb{E}_{\rho_t}[h^\top \Xi^{-1} h] - (h - \mathbb{E}_{\rho_t}[h])^\top \Xi^{-1} y_t \\ &= -\frac{1}{2} \|h - \mathbb{E}_{\rho_t}[h]\|_{\Xi}^2 + \frac{1}{2} \left( \mathbb{E}_{\rho_t}[h^\top \Xi^{-1} h] - \mathbb{E}_{\rho_t}[h^\top] \Xi^{-1} \mathbb{E}_{\rho_t}[h] \right) \\ &\quad + \frac{1}{2} (h^\top \Xi^{-1} h - \mathbb{E}_{\rho_t}[h^\top \Xi^{-1} h]) - (h - \mathbb{E}_{\rho_t}[h])^\top \Xi^{-1} y_t. \end{aligned}$$

Combining yields

$$\begin{aligned} I_1(x) + sI_2(x) + sI_3(x) &= -\frac{1}{2} (r - s) \left( h(x)^\top \Xi^{-1} h(x) - \mathbb{E}_{\rho_t}[h(x)^\top \Xi^{-1} h(x)] \right) + (r - s) (h(x) - \mathbb{E}_{\rho_t}[h])^\top \Xi^{-1} y_t \\ &\quad - \frac{s}{2} \|h - \mathbb{E}_{\rho_t}[h]\|_{\Xi}^2 + \frac{s}{2} \left( \mathbb{E}_{\rho_t}[h^\top \Xi^{-1} h] - \mathbb{E}_{\rho_t}[h^\top] \Xi^{-1} \mathbb{E}_{\rho_t}[h] \right). \end{aligned}$$

Substituting the above into (3.13) yields (3.7). □

## 5.3 Proof of Theorem 3.1

*Proof.* Start with the reformulation of (3.13) as derived in Lemma 3.1, (repeating here for convenience)

$$\partial_t \mu_t^d(x) = \mathcal{L}^* \mu_t^d(x) + \left( -\frac{r}{2} h(x)^\top \Xi^{-1} h(x) + (r - s) h(x)^\top \Xi^{-1} \frac{dZ_t^d}{dt} \right) \mu_t^d(x) \quad (5.3)$$

and the Stratonovich form of (3.14),

$$dq_t = \mathcal{L}^* q_t(x) dt - \frac{r}{2} h(x)^\top \Xi^{-1} h(x) q_t(x) dt + (r - s) q_t(x) h(x)^\top \Xi^{-1} \circ dZ_t. \quad (5.4)$$

We proceed with the following steps. The proof below is inspired by the proof of Theorem 3.1 in [HKX02] except for the following important extensions: 1) we no longer assume  $h$  is uniformly bounded; 2)  $h$  is no longer a scalar valued function but may be vector valued; 3) we make use of the forward stochastic Feynman-Kac style representation formula in [Kun82] rather than the backward formula. Existence and uniqueness of density valued solutions to the zakai equation in  $L^2(\mathbb{R}^m)$  with  $g, h$  unbounded has been studied by a number of authors [BBH83; BKK95], building on extensive works in the unbounded case (see e.g. the excellent summary in [BC09]).

**Step 1.** Use (stochastic) Feynman-Kac type formulae to obtain a probabilistic representation of solutions to the Zakai equation and replicator-mutator equation, as given in Theorem A.1. Specifically, we make use of the formulae developed in [Kun82; Kun81] as was done in [HKX02] although we rely on the forward rather than backward representation formulae. Recall that  $\mathcal{L}^*$  denotes the adjoint operator of the generator of the diffusion process

$$dX_t = g(X_t)dt + \sigma(X_t)dW_t$$

By expanding the adjoint operator, we may express it as

$$\begin{aligned} \mathcal{L}^* \mu_t^d(x) &= - \sum_{i=1}^m \frac{\partial}{\partial x^i} \left( g^i(x) \mu_t^d(x) \right) + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2}{\partial x^i \partial x^j} \left( (\sigma \sigma^\top)^{ij} \mu_t^d(x) \right) \\ &= - \sum_{i=1}^m \left( \mu_t^d(x) \frac{\partial g^i(x)}{\partial x^i} + g^i(x) \frac{\partial \mu_t^d(x)}{\partial x^i} \right) + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \left( \mu_t^d \frac{\partial^2 (\sigma \sigma^\top)^{ij}}{\partial x^i \partial x^j} + (\sigma \sigma^\top)^{ij} \frac{\partial^2 \mu_t^d}{\partial x^i \partial x^j} \right) \\ &\quad + \sum_{i=1}^m \frac{\partial \mu_t^d(x)}{\partial x^i} \sum_{j=1}^m \frac{\partial (\sigma \sigma^\top)^{ij}}{\partial x^j} \\ &= \mathcal{G} \mu_t^d(x) + \left( - \sum_{i=1}^m \frac{\partial g^i(x)}{\partial x^i} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 (\sigma \sigma^\top)^{ij}}{\partial x^i \partial x^j} \right) \cdot \mu_t^d(x) \end{aligned}$$

where in the second line we have used that  $\sigma \sigma^\top$  is symmetric and  $\mathcal{G}$  denotes the infinitesimal generator of the diffusion process

$$dY_t = (b(Y_t) - g(Y_t)) dt + \sigma(Y_t) dW_t^y \tag{5.5}$$

where  $W_t^y$  is a Wiener process independent of  $W_t$ ,  $b^i(x) = \nabla \cdot (\sigma(x) \sigma(x)^\top)^i$  and the superscript  $i$  denotes the  $i$ th row of the matrix  $\sigma \sigma^\top$ . That is,

$$\mathcal{G} \mu_t^d(x) := - \sum_{i=1}^m g^i(x) \frac{\partial \mu_t^d(x)}{\partial x^i} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m (\sigma \sigma^\top)^{ij} \frac{\partial^2 \mu_t^d}{\partial x^i \partial x^j} + \sum_{i=1}^m \frac{\partial \mu_t^d(x)}{\partial x^i} \sum_{j=1}^m \frac{\partial (\sigma \sigma^\top)^{ij}}{\partial x^j}$$

This decomposition of  $\mathcal{L}^*$  will be used in both (5.3) and (5.4); starting with (5.4) and using (1.6) yields,

$$\begin{aligned} dq_t(x) &= \mathcal{L}^* q_t(x) - \frac{r}{2} h(x)^\top \Xi^{-1} h(x) q_t(x) + (r-s) q_t(x) h(x)^\top \Xi^{-1} \circ dZ_t \\ &= \mathcal{G} q_t(x) + \left( - \sum_{i=1}^m \frac{\partial g^i(x)}{\partial x^i} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 (\sigma \sigma^\top)^{ij}}{\partial x^i \partial x^j} - \frac{r}{2} h(x)^\top \Xi^{-1} h(x) + (r-s) h(x)^\top \Xi^{-1} h(x) dt \right) \cdot q_t(x) \\ &\quad + (r-s) q_t(x) h(x)^\top \Xi^{-1/2} \circ dB_t \end{aligned}$$

This equation now takes the form of (1.1) in Theorem A.1 with

$$\begin{aligned}
l^{ik} &= 0 \\
c^k &= (r-s)(h(x)^\top \Xi^{-1/2})^k, \quad k = 1, 2, \dots, n \\
c^0 &= -\sum_{i=1}^m \frac{\partial g^i(x)}{\partial x^i} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 (\sigma \sigma^\top)^{ij}}{\partial x^i \partial x^j} - \frac{r}{2} h(x)^\top \Xi^{-1} h(x) + (r-s) h(x)^\top \Xi^{-1} h(x_t^*) \\
a &= \sigma \\
b^i &= -g^i(x) + \sum_{j=1}^m \sum_{k=1}^m \sigma^{ik} \frac{\partial \sigma^{kj}}{\partial x^j} + \frac{1}{2} \sigma^{kj} \frac{\partial \sigma^{ik}}{\partial x^j}
\end{aligned}$$

recalling that  $x_t^*$  is treated as a fixed realisation. Then by Theorem A.1, there exists another probability space equipped with the measure  $\mathbb{Q}$  (from here on we use the notation  $\mathbb{E}^\mathbb{Q}$  denote the expectation with respect to this measure) such that the solution can be represented as

$$q_t(x) = \mathbb{E}^\mathbb{Q} \left[ f(\xi_t(x)) \exp \left( (r-s) \int_0^t h(\xi_u(x))^\top \Xi^{-1} \circ dZ_u - \frac{r}{2} \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(\xi_u(x)) du + \int_0^t \tilde{c}(\xi_u(x)) du \right) \right]$$

where  $\xi_s(x)$  is a vector-valued function of  $x$  denoting the solution of an SDE in the form of (1.2) in Theorem A.1 with  $b, a, l$  as defined above, i.e.

$$d\xi_t(x) = \sum_{i=1}^m b^i(\xi_t(x)) dt + \sum_{j=1}^m \sum_{i=1}^m \sigma^{ij}(\xi_t(x)) \circ dW_t^j \quad (5.6)$$

and  $f(x) = \lim_{t \rightarrow 0} q_t(x)$  denotes the initial density and

$$\tilde{c}(x) := -\sum_{i=1}^m \frac{\partial g^i(x)}{\partial x^i} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 (\sigma(x) \sigma^\top(x))^{ij}}{\partial x^i \partial x^j} \quad (5.7)$$

We can similarly apply Theorem A.1 to (5.3) with

$$\begin{aligned}
c^k &= 0 \\
c^0 &= -\sum_{i=1}^m \frac{\partial g^i(x)}{\partial x^i} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 (\sigma \sigma^\top)^{ij}}{\partial x^i \partial x^j} - \frac{r}{2} h(x)^\top \Xi^{-1} h(x) + (r-s) h(x)^\top \Xi^{-1} \frac{dZ_t^d}{dt}
\end{aligned}$$

and  $l, a, b$  as defined previously, to obtain the representation

$$\mu_t^d(x) = \mathbb{E}^\mathbb{Q} \left[ f(\xi_t(x)) \exp \left( (r-s) \int_0^t h(\xi_u(x))^\top \Xi^{-1} \dot{Z}_u^d du - \frac{r}{2} \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(\xi_u(x)) du + \int_0^t \tilde{c}(\xi_u(x)) du \right) \right]$$

where  $\tilde{c}(x)$  is given in (5.7) and  $\xi_s$  is as defined previously since  $l^{ik} = 0$  in both cases and from now onwards we use the shorthand notation  $\dot{Z}_t^d \equiv \frac{dZ_t^d}{dt}$ . Recall also that both (5.3) and (5.4) are assumed to be initialised by the same density  $f(x)$ .

**Step 2.** We are now ready to prove pointwise convergence using the above representation formulae. Firstly, we have

$$\begin{aligned}
\mu_t^d(x) - q_t(x) &= \mathbb{E}^\mathbb{Q} \left[ f(\xi_t(x)) \exp \left( -\frac{r}{2} \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(\xi_u(x)) du + (r-s) \int_0^t h(\xi_u(x))^\top \Xi^{-1} \dot{Z}_u^d du + \int_0^t \tilde{c}(\xi_u(x)) du \right) \right. \\
&\quad \left. - \mathbb{E}^\mathbb{Q} \left[ f(\xi_t(x)) \exp \left( -\frac{r}{2} \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(\xi_u(x)) du + (r-s) \int_0^t h(\xi_u(x))^\top \Xi^{-1} \circ dZ_u + \int_0^t \tilde{c}(\xi_u(x)) du \right) \right] \right. \\
&=: \mathbb{E}^\mathbb{Q} [f(\xi_t(x)) I_2(x) I_1(x; \mathcal{Z})]
\end{aligned}$$

where

$$\begin{aligned}
I_2(x) &= \exp\left(\int_0^t \tilde{c}(\xi_s(x)) - \frac{r}{2} h(\xi_s(x))^\top \Xi^{-1} h(\xi_s(x)) ds\right) \\
I_1(x; \mathcal{Z}) &= \exp\left((r-s) \int_0^t h(\xi_s(x))^\top \Xi^{-1} \dot{Z}_s^d ds\right) - \exp\left((r-s) \int_0^t h(\xi_s(x))^\top \Xi^{-1} \circ dZ_s\right) \\
&=: \exp(I_3(x; \mathcal{Z})) - \exp(I_4(x; \mathcal{Z}))
\end{aligned}$$

and the  $\mathcal{Z}$  notation is used to denote the dependence of  $I_1$  on the observation path. In the below, let  $\mathbb{E}$  refer to the expectation on the original space, i.e. wrt to the observation noise  $W$ . Then for any fixed  $x \in \mathbb{R}^m$ , in other words, a realisation of the initialisation which has probability density  $f(x)$ ,

$$\begin{aligned}
\mathbb{E}\left[|\mu_t^d(x) - q_t(x)|^p\right] &\leq \mathbb{E}\left[\mathbb{E}^\mathbb{Q}\left[|f(\xi_t(x))I_2(x)I_1(x; \mathcal{Z})|^p\right]\right] \\
&= \mathbb{E}^\mathbb{Q}\left[|f(\xi_t(x))I_2(x)|^p \mathbb{E}\left[|I_1(x; \mathcal{Z})|^p\right]\right] \quad (\text{Fubini}) \\
&\leq C\left(\mathbb{E}^\mathbb{Q}\left[|f(\xi_t(x))I_2(x)|^{pr_2}\right]\right)^{1/r_2} \cdot \left(\mathbb{E}^\mathbb{Q}\left[\left(\mathbb{E}\left[|I_1(x; \mathcal{Z})|^p\right]\right)^{r_1}\right]\right)^{1/r_1} \quad (\text{Hoelder inequality}) \\
&=: CI_5^{1/r_2} \cdot I_6^{1/r_1}
\end{aligned}$$

for a constant  $C > 0$  independent of  $t$  and with  $1/r_1 + 1/r_2 = 1$  and  $r_1, r_2 > 1$ . Starting with  $I_2$ , since  $g(x)$  is assumed to be  $C^2$  and globally Lipschitz continuous and  $\Sigma$  is a constant, it holds that  $\tilde{c}(x)$  as defined in (5.7) is uniformly bounded. Additionally, it holds that  $h(x)^\top \Xi^{-1} h(x) > 0$ ,  $\forall x \in \mathbb{R}^m$  since  $\Xi$  is a positive definite matrix, so that there exists some  $C_2 > 0$  independent of  $t$ . Combining, we have that

$$|I_2(x)|^{pr_2} = \exp\left(pr_2 \int_0^t \tilde{c}(\xi_u(x)) du\right) \cdot \exp\left(-pr_2 \frac{r}{2} \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(\xi_u(x)) du\right) \leq \exp(C_1 t) \cdot C_2$$

from which we obtain

$$\begin{aligned}
I_5 &\leq \mathbb{E}^\mathbb{Q}\left[|f(\xi_t(x))|^{pr_2} |I_2(x)|^{pr_2}\right] \\
&\leq C_3(t)
\end{aligned}$$

using the fact that  $f$  is uniformly bounded. Now turning to  $I_6$ , using the identity

$$\begin{aligned}
|\exp(x) - \exp(y)|^p &\leq (\exp(x) + \exp(y))^p |x - y|^p, \quad x, y \in \mathbb{R} \\
&\leq C(\exp(px) + \exp(py)) |x - y|^p
\end{aligned}$$

along with Hoelder inequality yields

$$\begin{aligned}
I_6 &\leq C\mathbb{E}^\mathbb{Q}\left[\left(\mathbb{E}\left[\left(\exp(pI_3(x; \mathcal{Z})) + \exp(pI_4(x; \mathcal{Z}))\right)|I_3(x; \mathcal{Z}) - I_4(x; \mathcal{Z})|^p\right]\right)^{r_1}\right] \\
&\leq C\mathbb{E}^\mathbb{Q}\left[\left(\mathbb{E}\left[\exp(pI_3(x; \mathcal{Z}))\right]|I_3(x; \mathcal{Z}) - I_4(x; \mathcal{Z})|^p\right)^{r_1} + \left(\mathbb{E}\left[\exp(pI_4(x; \mathcal{Z}))\right]|I_3(x; \mathcal{Z}) - I_4(x; \mathcal{Z})|^p\right)^{r_1}\right] \\
&\leq C\mathbb{E}^\mathbb{Q}\left[\left(\mathbb{E}\left[\exp(r_2 p I_3(x; \mathcal{Z}))\right]\right)^{r_1/r_2} \cdot \mathbb{E}\left[|I_3(x; \mathcal{Z}) - I_4(x; \mathcal{Z})|^{pr_1}\right]\right] \\
&+ C\mathbb{E}^\mathbb{Q}\left[\left(\mathbb{E}\left[\exp(r_2 p I_4(x; \mathcal{Z}))\right]\right)^{r_1/r_2} \cdot \mathbb{E}\left[|I_3(x; \mathcal{Z}) - I_4(x; \mathcal{Z})|^{pr_1}\right]\right] \\
&\leq C\left[\left(\mathbb{E}^\mathbb{Q}\left[I_9(x)^{r_1}\right]\right)^{1/r_2} + \left(\mathbb{E}^\mathbb{Q}\left[I_8(x)^{r_1}\right]\right)^{1/r_2}\right] \left(\mathbb{E}^\mathbb{Q}\left[I_7(x)^{r_1}\right]\right)^{1/r_1}
\end{aligned}$$

with

$$I_7(x) := \mathbb{E}[|I_3(x; \mathcal{Z}) - I_4(x; \mathcal{Z})|^{p r_1}], \quad I_8(x) := \mathbb{E}[\exp(r_2 p I_4(x; \mathcal{Z}))], \quad I_9(x) := \mathbb{E}[\exp(r_2 p I_3(x; \mathcal{Z}))]$$

**Step 3.** The remainder of the proof will focus on showing that the terms involving  $I_8$  and  $I_9$  can be bounded by constants (depending on  $t$  only). The term involving  $I_7$  will be shown to go to zero as  $d \rightarrow \infty$ , yielding the desired convergence result. Starting with  $I_8(x)$  and using the shorthand notation  $p_r := p(r - s)$ ,

$$\begin{aligned} \mathbb{E}[\exp(r_2 p_r I_4(x; \mathcal{Z}))] &= \mathbb{E}\left[\exp\left(r_2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1} \circ dZ_u\right)\right] \\ &= \mathbb{E}\left[\exp\left(r_2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(x_u^*) du + r_2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1/2} dB_u\right)\right] \\ &\leq \left(\mathbb{E}\left[\exp\left(r_1 r_2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(x_u^*) du\right)\right]\right)^{1/r_1} \cdot \left(\mathbb{E}\left[\exp\left(r_2^2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1/2} dB_u\right)\right]\right)^{1/r_2} \\ &\leq \left(\mathbb{E}\left[\exp\left(r_1 r_2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(x_u^*) du\right)\right]\right)^{1/r_1} \cdot \left(\prod_{k=1}^n \mathbb{E}\left[\exp\left(nr_2^2 p_r \int_0^t (h(\xi_u(x))^\top \Xi^{-1/2})^k dB_u^k\right)\right]\right)^{1/(r_2 n)} \\ &\leq \left(\mathbb{E}\left[\exp\left(r_1 r_2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(x_u^*) du\right)\right]\right)^{1/r_1} \cdot \left(\exp\left(\frac{nr_2^2 p_r}{2} \sum_{k=1}^n \int_0^t ((h(\xi_u(x))^\top \Xi^{-1/2})^k)^2 du\right)\right)^{1/(r_2 n)} \\ &= \left(\mathbb{E}\left[\exp\left(r_1 r_2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(x_u^*) du\right)\right]\right)^{1/r_1} \cdot \left(\exp\left(\frac{nr_2^2 p_r}{2} \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(\xi_u(x)) du\right)\right)^{1/(r_2 n)} \end{aligned} \tag{5.8}$$

where in the last line, we have used Lemma (A.1) and the fact that  $\xi_s$  is defined on a different probability space to the signal process. Then we have

$$\begin{aligned} \left(\mathbb{E}^\mathbb{Q}[I_8(x)^{r_1}]\right)^{1/r_2} &\leq \left(\mathbb{E}^\mathbb{Q}\left[\mathbb{E}\left[\exp\left(r_1 r_2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(x_u^*) du\right)\right] \cdot \exp\left(\frac{r_1 r_2 p_r}{2} \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(\xi_u(x)) du\right)\right]\right)^{1/r_2} \\ &= \left(\mathbb{E}^\mathbb{Q}\left[\exp\left(r_1 r_2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(x_u^*) du\right) \cdot \exp\left(\frac{r_1 r_2 p_r}{2} \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(\xi_u(x)) du\right)\right]\right)^{1/r_2} \\ &= \left(\mathbb{E}^\mathbb{Q}\left[\exp\left(r_1 r_2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1} \left(h(x_u^*) + \frac{1}{2} h(\xi_u(x))\right) du\right)\right]\right)^{1/r_2} \\ &=: I_{10}(x)^{1/r_2} \end{aligned}$$

since the expectation  $\mathbb{E}$  is with respect to the observation noise and  $x_s^*$  is taken as a fixed realisation.

To proceed further, we will make use of Lemma A.4 which allows us to bound exponential moments of a non-decreasing process by its raw moments. Generally, it is not possible to bound exponential moments in terms of polynomial moments, as the exponential term grows faster. The crucial point of this lemma is in a careful specification of the factor in the exponential (c.f.  $L$  in Lemma A.4) which acts to “dampen” the growth relative to the growth of the raw moments. We have using the shorthand notation  $\tilde{L} := \frac{r_1 r_2 p_r}{2}$ ,

$$I_{10}(x) \leq \mathbb{E}^\mathbb{Q}\left[\exp\left(\tilde{L} \int_0^t \tilde{h}_u(\xi_u(x))^\top \Xi^{-1} \tilde{h}_u(\xi_u(x)) du\right)\right]$$

where  $\tilde{h}_u(y) := h(y) + h(x_u^*)$ . Furthermore,

$$\tilde{h}_u(x)^\top \Xi^{-1} \tilde{h}_u(x) = |\Xi^{-1/2} \tilde{h}_u(x)|^2 \leq \lambda_{\Xi}^2 |\tilde{h}_u(x)|^2, \quad \forall x \in \mathbb{R}^m,$$

where  $\lambda_{\Xi}$  is the smallest eigenvalue of  $\Xi^{1/2}$ . Combining, we have letting  $L = \tilde{L} \lambda_{\Xi}^2$ ,

$$I_{10}(x) \leq \mathbb{E}^{\mathbb{Q}} \left[ \exp \left( L \int_0^t |\tilde{h}_u(\xi_u(x))|^2 du \right) \right]$$

Let  $Y_t(x) := \int_0^t |\tilde{h}_u(\xi_u(x))|^2 du$ . Clearly this is a non-decreasing (and adapted) process, so that we may apply lemma A.4. In particular,

$$\mathbb{E}^{\mathbb{Q}}[Y_t(x) - Y_\tau(x)] = \int_\tau^t \mathbb{E}^{\mathbb{Q}} \left[ |\tilde{h}_u(\xi_u(x))|^2 \right] du \leq tC(1 + \mathbb{E}^{\mathbb{Q}}[|x|^2]) =: K(t), \quad \forall \tau \in [0, t]$$

where  $C$  is a constant depending on the growth properties of  $h$  and for the last inequality, we have used that since  $g, h$  satisfy lipschitz and linear growth assumptions,

$$\mathbb{E}^{\mathbb{Q}}[|\tilde{h}_u(\xi_u(x))|^2] \leq C(1 + \mathbb{E}^{\mathbb{Q}}[|\xi_u(x)|^2]) \leq C(1 + \mathbb{E}^{\mathbb{Q}}[|x|^2])$$

Therefore by Lemma A.4, whenever

$$\frac{r_1 r_2 (r - s) p \lambda_{\Xi}^2}{2} < \frac{1}{K(t)} \tag{5.9}$$

where  $K(t)$  is a constant depending on time, the second moment of the initial density  $f(x)$  and the linear growth constant of  $h$ , we have

$$I_{10}(x) \leq \mathbb{E}^{\mathbb{Q}} \left[ \exp \left( L \int_0^t |\tilde{h}_s(\xi_s(x))|^2 ds \right) \right] < \frac{1}{1 - LK(t)}.$$

Notice that condition (5.9) can be satisfied whenever  $r, s$  are chosen such that  $(r - s)$  is small enough. As will be seen in Section 4.2, this is at least possible in the linear-Gaussian setting whilst maintaining optimality (in the mean squared error sense) even in the case of a misspecified model. Finally, we have

$$\left( \mathbb{E}^{\mathbb{Q}}[I_8(x)^{r_1}] \right)^{1/r_2} \leq (1 - LK(t))^{-\frac{1}{r_2}}$$

The term involving  $I_9(x)$  can be analysed in much the same way as for  $I_8(x)$ , with the only difference being that the stochastic integral. In particular, we have letting  $j$  refer to the index such that  $t \in (t_j, t_{j+1}]$ ,

$$\begin{aligned} \mathbb{E}[\exp(r_2 p_r I_3(x; \mathcal{Z}))] &= \mathbb{E} \left[ \exp \left( r_2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1} \dot{Z}_u^d du \right) \right] \\ &\leq \left( \mathbb{E} \left[ \exp \left( r_1 r_2 p_r \int_0^t h(\xi_u(x))^\top \Xi^{-1} h(x_u^*) du \right) \right] \right)^{1/r_1} \cdot \left( \prod_{k=1}^n \mathbb{E} \left[ \exp \left( n r_2^2 p_r \int_0^t (h(\xi_u(x))^\top \Xi^{-1/2})^k \dot{B}_u^k du \right) \right] \right)^{1/(r_2 n)}. \end{aligned}$$

where with a slight abuse of notation, we let  $\dot{B}_t \equiv \dot{B}_t^d$  and  $\dot{B}_t^k$  denotes the  $k$ th component of  $\dot{B}_t$ . Furthermore,

$$\begin{aligned}
& \mathbb{E} \left[ \exp \left( nr_2^2 p_r \int_0^t (h(\xi_u(x)))^\top \Xi^{-1/2})^k \dot{B}_\tau^k d\tau \right) \right] \\
&= \mathbb{E} \left[ \exp \left( nr_2^2 p_r \sum_{i=1}^j \frac{1}{\delta_d} \int_{t_{i-1}}^{t_i} \left( \int_{t_{i-1}}^{t_i} (h(\xi_u(x)))^\top \Xi^{-1/2})^k du \right) dB_\tau^k + nr_2^2 p_r \int_{t_j}^t (h(\xi_s(x)))^\top \Xi^{-1/2})^k \dot{B}_\tau^k d\tau \right) \right] \\
&= \prod_{i=1}^j \mathbb{E} \left[ \exp \left( nr_2^2 p_r \frac{1}{\delta_d} \left( \int_{t_{i-1}}^{t_i} (h(\xi_u(x)))^\top \Xi^{-1/2})^k du \right) \int_{t_{i-1}}^{t_i} dB_\tau^k \right) \right] \mathbb{E} \left[ \exp \left( nr_2^2 p_r \left( \int_{t_j}^t (h(\xi_u(x)))^\top \Xi^{-1/2})^k du \right) \int_{t_j}^{t_{j+1}} \dot{B}_\tau^k d\tau \right) \right] \\
&\leq \prod_{i=1}^j \exp \left( \frac{nr_2^2 p_r}{2} \int_{t_{i-1}}^{t_i} (h(\xi_u(x)))^\top \Xi^{-1/2})^k du \right) \cdot \exp \left( \frac{nr_2^2 p_r}{2} \int_{t_j}^t (h(\xi_u(x)))^\top \Xi^{-1/2})^k du \right) \\
&= \exp \left( \frac{nr_2^2 p_r}{2} \int_0^t (h(\xi_u(x)))^\top \Xi^{-1/2})^k du \right) \\
&\leq \exp \left( \frac{nr_2^2 p_r}{2} \int_0^t |(h(\xi_u(x)))^\top \Xi^{-1/2})^k|^2 du \right) \\
&\leq \exp \left( \frac{nr_2^2 p_r}{2} \int_0^t |(h(\xi_u(x)))^\top \Xi^{-1/2})^k|^2 du \right)
\end{aligned}$$

where the second equality holds due to independence of brownian increments and in the first inequality holds due to lemma A.1. Combining this result with the same calculations as for  $I_8$  yields an upper bound on  $I_9$  which is identical to (5.8). Therefore, following the same reasoning as in for  $I_8$ , we have

$$\left( \mathbb{E}^{\mathbb{Q}}[I_9(x)^{r_1}] \right)^{1/r_2} \leq (1 - LK(t))^{-\frac{1}{r_2}}$$

Finally, for  $I_7$ , first note that for  $t \in [t_j, t_{j+1})$ ,

$$\begin{aligned}
\int_0^t h(\xi_u(x))^\top \Xi^{-1} \dot{Z}_u^d du &= \sum_{i=1}^j \int_{t_{i-1}}^{t_i} h(\xi_u(x))^\top \Xi^{-1} \left( h(x_{t_{i-1}}^*) + \frac{1}{\delta_d} \Xi^{1/2} (B_{t_i} - B_{t_{i-1}}) \right) du \\
&\quad + \int_{t_j}^t h(\xi_u(x))^\top \Xi^{-1} \left( h(x_{t_j}^*) + \frac{1}{\delta_d} \Xi^{1/2} (B_{t_{j+1}} - B_{t_j}) \right) du
\end{aligned}$$

similarly,

$$\begin{aligned}
\int_0^t h(\xi_u(x))^\top \Xi^{-1} \circ dZ_u &= \int_0^t h(\xi_u(x))^\top \Xi^{-1} dZ_u \\
&= \sum_{i=1}^j \int_{t_{i-1}}^{t_i} h(\xi_u(x))^\top \Xi^{-1} (h(x_u^*) du + \Xi^{1/2} dB_u) + \int_{t_j}^t h(\xi_u(x))^\top \Xi^{-1} (h(x_u^*) du + \Xi^{1/2} dB_u)
\end{aligned}$$

Then

$$\begin{aligned}
|I_3 - I_4|^{pr_1} &\leq |2(r-s)|^{pr_1} \left( \left| \sum_{i=1}^j \int_{t_{i-1}}^{t_i} h(\xi_u(x))^\top \Xi^{-1} (h(x_{t_{i-1}}^*) - h(x_u^*)) du + \int_{t_j}^t h(\xi_u(x))^\top \Xi^{-1} (h(x_{t_j}^*) - h(x_u^*)) du \right|^{pr_1} \right. \\
&\quad \left. + \left| \int_0^t h(\xi_u(x))^\top \Xi^{-1/2} \dot{B}_u^d du - \int_0^t h(\xi_u(x))^\top \Xi^{-1/2} dB_u \right|^{pr_1^2} \right) \\
&=: |2(r-s)|^{pr_1} (I_{11} + I_{12})
\end{aligned}$$



As we are required to upper bound  $\mathbb{E}[|I_3 - I_4|^{pr_1}]$ , we obtain the following bound for  $\mathbb{E}[I_{12}]$  following the same reasoning as in pg 41 of [HKX02] (bound on  $I_5^\Pi(x)$  in their proof),

$$\begin{aligned}
\mathbb{E}[I_{12}] &= \mathbb{E} \left[ \left| \sum_{k=1}^n \int_0^t (h(\xi_u(x))^\top \Xi^{-1/2})^k \dot{B}_u^k du - \int_0^t (h(\xi_u(x))^\top \Xi^{-1/2})^k dB_u^k \right|^{pr_1} \right] \\
&\leq n^{pr_1} \sum_{k=1}^n \mathbb{E} \left[ \left| \int_0^t (h(\xi_u(x))^\top \Xi^{-1/2})^k \dot{B}_u^k du - \int_0^t (h(\xi_u(x))^\top \Xi^{-1/2})^k dB_u^k \right|^{pr_1} \right] \\
&\leq C \frac{1}{\delta_d} \sum_{i=1}^j \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t_i} \mathbb{E}[|(h(\xi_u(x))^\top \Xi^{-1/2})^k - (h(\xi_\tau(x))^\top \Xi^{-1/2})^k|^{pr_1}] d\tau du \\
&\quad + C \delta_d^{\frac{pr_1}{2}-1} \int_{t_j}^t \mathbb{E}[|(h(\xi_\tau(x))^\top \Xi^{-1/2})^k|^{pr_1}] d\tau
\end{aligned}$$

where  $C$  is a constant depending on  $n$ . It is worthwhile clarifying that classical Wong-Zakai/piecewise smooth convergence results are focused on stochastic integrals of the form  $|\int h(X_s) \circ dB_s - \int h(X_s^d) \dot{B}_s^d ds|$  where the integrand  $X_s$  is dependent on  $B_s^d$ . As the coefficient here  $h(\xi_s(x))$  evolves independently of  $B$ , we can resort to simpler convergence tools than used in e.g. [Pat24]. Starting with the second term, we have under the assumptions on  $b, g, h$  using lemma A.3,

$$\begin{aligned}
\mathbb{E}[|(h(\xi_\tau(x))^\top \Xi^{-1/2})^k|^{pr_1}] &\leq \mathbb{E}[|(h(\xi_\tau(x))^\top \Xi^{-1/2})^k - (h(x)^\top \Xi^{-1/2})^k|^{pr_1}] + \mathbb{E}[|(h(x)^\top \Xi^{-1/2})^k|^{pr_1}] \\
&\leq C(x) |\tau|^{pr_1/2} + C(x)
\end{aligned}$$

so that

$$\begin{aligned}
C \delta_d^{\frac{pr_1}{2}-1} \int_{t_j}^t \mathbb{E}[|(h(\xi_\tau(x))^\top \Xi^{-1/2})^k|^{pr_1}] d\tau &\leq C \delta_d^{\frac{pr_1}{2}-1} C(x) \int_{t_j}^t (|\tau|^{pr_1/2} + 1) d\tau \\
&\leq C \delta_d^{\frac{pr_1}{2}-1} C(x) (\delta_d^{pr_1/2+1} + \delta_d) \\
&\leq C(x) \delta_d^{pr_1/2}
\end{aligned}$$

Then for the first term on the rhs of the inequality, making use of lemma A.3 and that  $h$  is lipschitz continuous and at most linear growth, we have

$$\begin{aligned}
&C \frac{1}{\delta_d} \sum_{i=1}^j \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t_i} \mathbb{E}[|(h(\xi_u(x))^\top \Xi^{-1/2})^k - (h(\xi_\tau(x))^\top \Xi^{-1/2})^k|^{pr_1}] d\tau du \\
&\leq C(x) \frac{1}{\delta_d} \sum_{i=1}^j \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t_i} |\tau - u|^{pr_1/2} d\tau du \\
&=\leq C(x) \frac{1}{\delta_d} \sum_{i=1}^j \delta_d^{pr_1/2} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t_i} d\tau du \\
&\leq C(x) \frac{1}{\delta_d} \sum_{i=1}^j \delta_d^{pr_1/2+2} \\
&= C(x) \frac{1}{\delta_d} \frac{t_j}{\delta_d} \delta_d^{pr_1/2+2} = C(x) \delta_d^{pr_1/2}
\end{aligned}$$

Combining the two yields

$$\mathbb{E}[I_{12}(x)] \leq C(x) \delta_d^{pr_1/2}$$

Finally, for  $I_{11}(x)$  we have using the notation  $[x_t^*] = x_{t_i}^*$ ,  $t \in [t_{i-1}, t_i]$  and Lemma A.3 and the linear growth assumption on  $h$ ,

$$\begin{aligned}
\mathbb{E}[I_{11}(x)] &= \mathbb{E} \left[ \left| \int_0^{t_j} h(\xi_s(x))^\top \Xi^{-1}(h(\lfloor x_u^* \rfloor) - h(x_u^*)) du + \int_{t_j}^t h(\xi_u(x))^\top \Xi^{-1}(h(x_{t_j}^*) - h(x_u^*)) du \right|^{pr_1} \right] \\
&\leq C \int_0^{t_j} \mathbb{E} \left[ \left| h(\xi_u(x))^\top \Xi^{-1}(h(\lfloor x_u^* \rfloor) - h(x_u^*)) \right|^{pr_1} \right] du + C \int_{t_j}^t \mathbb{E} \left[ \left| h(\xi_u(x))^\top \Xi^{-1}(h(x_{t_j}^*) - h(x_u^*)) \right|^{pr_1} \right] du \\
&\leq n^{pr_1} C(1 + C_h(x)) \sum_{i=1}^j \int_{t_{i-1}}^{t_i} \mathbb{E} \left[ \left| h(x_{t_{i-1}}^*) - h(x_u^*) \right|^{pr_1} \right] du + \int_{t_j}^t \mathbb{E} \left[ \left| h(x_{t_j}^*) - h(x_u^*) \right|^{pr_1} \right] du \\
&\leq n^{pr_1} C(1 + C_h(x)) \sum_{i=1}^{j+1} \int_{t_{i-1}}^{t_i} \delta_d^{pr_1/2} du \\
&\leq n^{pr_1} C(1 + C_h(x)) \frac{t_{j+1}}{\delta_d} \delta_d^{pr_1/2+1} \\
&\leq n^{pr_1} C(t)(1 + C_h(x)) \delta_d^{pr_1/2}
\end{aligned}$$

We are now ready to bound the remaining term involving  $I_7$ , i.e.

$$\begin{aligned}
\left( \mathbb{E}^\mathbb{Q}[I_7(x)^{r_1}] \right)^{1/r_1} &\leq (2(r-s))^p \left( \mathbb{E}^\mathbb{Q}[(\mathbb{E}[I_{11}(x)] + \mathbb{E}[I_{12}(x)])^{r_1}] \right)^{1/r_1} \\
&\leq C(r-s)^p \left( \mathbb{E}^\mathbb{Q}[C(x)^{r_1}] \delta_d^{pr_1^2/2} \right)^{1/r_1} \\
&\leq C(r-s)^p \delta_d^{pr_1/2}
\end{aligned}$$

under the assumption of finite  $p > 1$  moments of the initial density  $f(x)$ . Choosing  $r_1 > 1$  large enough for a given  $p$  gives us the required decay as  $\delta_d \rightarrow 0$ .  $\square$

## 5.4 Proof of Lemma 4.1

*Proof.* Here we focus on the proof of the fully scalar case for clarity, although the calculations extend similarly to the multivariate setting with extra matrix algebra. Starting with the proof of the second claim, notice that (4.8) in the scalar case takes the form

$$d\bar{X}_t = \mu_t(x)dt + \Sigma^{1/2}dW_t + \sigma_t d\bar{W}_t$$

with

$$\begin{aligned}
\sigma_t &= \sqrt{r-s} H C_t \Xi^{-1/2}; \\
\mu_t(x) &= Gx - \frac{s}{2} H^2 C_t \Xi^{-1}(x - m_t) + (r-s) H C_t \Xi^{-1} \left( \frac{dZ_t^d}{dt} - Hx \right)
\end{aligned}$$

The (conditional) forward Kolmogorov equation is then given by

$$\partial_t \rho_t(x) = -\partial_x(\rho_t(x)\mu_t(x)) + \frac{1}{2}(\sigma_t^2 + \Sigma)\partial_{xx}\rho_t(x) \quad (5.10)$$

$$= -\mu_t(x)\partial_x\rho_t(x) - \rho_t(x)\partial_x\mu_t(x) + \frac{1}{2}(\sigma_t^2 + \Sigma)\partial_{xx}\rho_t(x) \quad (5.11)$$

and once again, we know that  $\rho_t(x) = \mathcal{N}(x; m_t, C_t)$ . First define

$$\mathcal{L}^* \rho_t(x) := Gx \frac{1}{C_t} (x - m_t) \rho_t(x) - \rho_t(x) g(x) + \frac{1}{2} \Sigma \left( -\frac{1}{C_t} + \frac{1}{C_t^2} (x - m_t)^2 \right) \rho_t(x)$$

and using Gaussianity, we have

$$\begin{aligned} \partial_x \mu_t(x) &= G - \frac{s}{2} H^2 C_t \Xi^{-1} - (r - s) H^2 C_t \Xi^{-1} \\ \partial_x \rho_t(x) &= -\frac{1}{C_t} (x - m_t) \rho_t(x) \\ \partial_{xx} \rho_t(x) &= -\frac{1}{C_t} \rho_t(x) + \frac{1}{C_t^2} (x - m_t)^2 \rho_t(x) \end{aligned}$$

Substituting the above expressions into (5.14) yields

$$\begin{aligned} \partial_t \rho_t(x) &= \left( Gx - \frac{s}{2} H^2 C_t \Xi^{-1} (x - m_t) + (r - s) H C_t \Xi^{-1} \left( \frac{dZ_t^d}{dt} - Hx \right) \right) \frac{1}{C_t} (x - m_t) \rho_t(x) \\ &\quad - \left( G - \frac{s}{2} H^2 C_t \Xi^{-1} - (r - s) H^2 C_t \Xi^{-1} \right) \rho_t(x) \\ &\quad + \frac{1}{2} \left( (r - s) H^2 C_t^2 \Xi^{-1} + \Sigma \right) \left( -\frac{1}{C_t} \rho_t(x) + \frac{1}{C_t^2} (x - m_t)^2 \rho_t(x) \right) \\ &= \mathcal{L}^* \rho_t(x) + \left( -\frac{s}{2} H^2 C_t \Xi^{-1} (x - m_t) + (r - s) H C_t \Xi^{-1} \left( \frac{dZ_t^d}{dt} - Hx \right) \right) \frac{1}{C_t} (x - m_t) \rho_t(x) \\ &\quad + \left( \frac{s}{2} H^2 C_t \Xi^{-1} + (r - s) H^2 C_t \Xi^{-1} \right) \rho_t(x) \\ &\quad + \frac{1}{2} \left( (r - s) H^2 C_t^2 \Xi^{-1} \right) \left( -\frac{1}{C_t} \rho_t(x) + \frac{1}{C_t^2} (x - m_t)^2 \rho_t(x) \right) \\ &= \mathcal{L}^* \rho_t(x) + (r - s) (Hx - Hm_t) \Xi^{-1} \frac{dZ_t^d}{dt} \rho_t(x) \\ &\quad + \left( \frac{r}{2} H^2 C_t \Xi^{-1} + \left( \frac{r}{2} - s \right) H^2 \Xi^{-1} (x - m_t)^2 - (r - s) H^2 \Xi^{-1} x (x - m_t) \right) \rho_t(x) \end{aligned} \quad (5.12)$$

Also, (4.2) can be simplified to obtain

$$\begin{aligned} \partial_t p_t(x) &= \mathcal{L}^* p_t(x) + (r - s) (Hx - Hm_t) \Xi^{-1} \frac{dZ_t^d}{dt} p_t(x) + \\ &\quad + \left( -(r - s) \frac{1}{2} ((Hx)^2 - \mathbb{E}_{p_t}[(Hx)^2]) - \frac{s}{2} ((Hx - Hm_t)^2 - H^2 C_t) \right) \Xi^{-1} p_t(x) \end{aligned}$$

Then using

$$x^2 - \mathbb{E}_{p_t}[x^2] = -(x - m_t)^2 + 2x(x - m_t) - C_t \quad (5.13)$$

we obtain

$$\begin{aligned} &- (r - s) \frac{1}{2} ((Hx)^2 - \mathbb{E}_{p_t}[(Hx)^2]) - \frac{s}{2} ((Hx - Hm_t)^2 - H^2 C_t) \\ &= \frac{r}{2} H^2 C_t \Xi^{-1} + \left( \frac{r}{2} - s \right) H^2 \Xi^{-1} (x - m_t)^2 - (r - s) H^2 \Xi^{-1} x (x - m_t) \end{aligned}$$

Therefore, (4.2) is formally equivalent to (5.12), as desired.

The proof of the first claim follows from a very similar line of reasoning, nevertheless, we present the calculations here for completeness. Starting from (4.6), which in the scalar case takes the form

$$d\bar{X}_t = \mu_t(x)dt + \Sigma^{1/2}dW_t + \sigma_t d\bar{W}_t + \sigma_t^z dZ_t$$

with

$$\begin{aligned}\sigma_t &= \sqrt{r-s}C_t H^\top \Xi^{-1/2}; \\ \mu_t(x) &= Gx - \frac{s}{2}C_t H^\top \Xi^{-1}H(x - m_t) - (r-s)C_t H^\top \Xi^{-1}Hx \\ \sigma_t^z &= (r-s)C_t H^\top \Xi^{-1}\end{aligned}$$

The (observation conditioned) forward Kolmogorov equation (see e.g. [PRS21]) is then given by

$$\begin{aligned}d\rho_t(x) &= -\partial_x(\rho_t(x)\mu_t(x))dt + \frac{1}{2}(\sigma_t\sigma_t^\top + \Sigma)\partial_{xx}\rho_t(x)dt - \partial_x(\rho_t(x)\sigma_t^z)dZ_t \\ &= -(\partial_x\rho_t(x))\mu_t(x)dt - \rho_t(x)\partial_x\mu_t(x)dt - (\partial_x\rho_t(x)) \cdot \sigma_t^z dZ_t + \frac{1}{2}(\sigma_t^2 + \Sigma)\partial_{xx}\rho_t(x)\end{aligned}\quad (5.14)$$

It is well known that  $\rho_t(x) = \mathcal{N}(x; m_t, C_t)$  when  $\mu_t(x)$  is linear and  $\rho_0$  is also Gaussian (c.f. ensemble Kalman-Bucy filter). Making use of Gaussianity of  $\rho_t(x)$ , we have

$$\begin{aligned}\partial_x\mu_t(x) &= \text{tr}(G - \frac{s}{2}C_t H^\top \Xi^{-1}H - (r-s)C_t H^\top \Xi^{-1}H) \\ \partial_x\rho_t(x) &= -C_t^{-1}(x - m_t)\rho_t(x) \\ \partial_{xx}\rho_t(x) &= -\text{tr}(C_t^{-1})\rho_t(x) + (x - m_t)^\top C_t^{-1}C_t^{-1}(x - m_t)\rho_t(x)\end{aligned}$$

and define

$$\mathcal{L}^*\rho_t(x) := (x - m_t)C_t^{-1}Gx\rho_t(x) - \text{tr}(G)\rho_t(x) + \frac{1}{2}\Sigma\left(-\text{tr}(C_t^{-1})\rho_t(x) + (x - m_t)^\top C_t^{-1}C_t^{-1}(x - m_t)\right)\rho_t(x)$$

substituting the derivative expressions into (5.14) yields

$$\begin{aligned}d\rho_t(x) &= (x - m_t)^\top C_t^{-1}\left(Gx - \frac{s}{2}C_t H^\top \Xi^{-1}H(x - m_t) - (r-s)C_t H^\top \Xi^{-1}Hx\right)\rho_t(x)dt \\ &+ \frac{1}{2}\left((r-s)C_t H^\top \Xi^{-1}HC_t + \Sigma\right)\left(-\text{tr}(C_t^{-1}) + (x - m_t)^\top C_t^{-1}C_t^{-1}(x - m_t)\right)\rho_t(x)dt \\ &= \mathcal{L}^*\rho_t(x)dt + (x - m_t)^\top C_t^{-1}\left(-\frac{s}{2}C_t H^\top \Xi^{-1}H(x - m_t) - (r-s)C_t H^\top \Xi^{-1}Hx\right)\rho_t(x)dt \\ &+ \left(\text{tr}\left(\frac{s}{2}C_t H^\top \Xi^{-1}H + (r-s)C_t H^\top \Xi^{-1}H\right)\right)\rho_t(x)dt + (r-s)(x - m_t)^\top H^\top \Xi^{-1}dZ_t \\ &+ \frac{1}{2}\left((r-s)C_t H^\top \Xi^{-1}HC_t\right)\left(-\text{tr}(C_t^{-1}) + (x - m_t)^\top C_t^{-1}C_t^{-1}(x - m_t)\right)\rho_t(x)dt \\ &= \mathcal{L}^*\rho_t(x)dt + (r-s)\rho_t(x)(x - m_t)^\top H^\top \Xi^{-1}dZ_t - (r-s)(x - m_t)^\top H^\top \Xi^{-1}Hx\rho_t(x)dt \\ &+ (x - m_t)^\top C_t^{-1}\left(-\frac{s}{2}C_t H^\top \Xi^{-1}H(x - m_t)\right)\rho_t(x)dt + \left(r - \frac{s}{2}\right)\left(C_t H^\top \Xi^{-1}H\right)\rho_t(x)dt \\ &+ \frac{1}{2}\left((r-s)C_t H^\top \Xi^{-1}HC_t\right)\left(-C_t^{-1} + (x - m_t)^\top C_t^{-1}C_t^{-1}(x - m_t)\right)\rho_t(x)dt \\ &= \mathcal{L}^*\rho_t(x)dt + (r-s)\rho_t(x)(x - m_t)^\top H^\top \Xi^{-1}dZ_t - (r-s)(x - m_t)^\top H^\top \Xi^{-1}Hx\rho_t(x)dt \\ &+ \left(r - \frac{s}{2}\right)(x - m_t)^\top H^\top \Xi^{-1}H(x - m_t)\rho_t(x)dt + \frac{r}{2}C_t H^\top \Xi^{-1}H\rho_t(x)dt \\ &= \mathcal{L}^*\rho_t(x)dt + (r-s)\rho_t(x)(x - m_t)^\top H^\top \Xi^{-1}dZ_t \\ &- \left(\frac{(r-s)}{2}((Hx)^2 - \mathbb{E}_{\rho_t}[(Hx)^2])\Xi^{-1} - \frac{s}{2}C_t H^2 \Xi^{-1} + \frac{r}{2}(Hx - Hm_t)^2 \Xi^{-1}\right)\rho_t(x)dt\end{aligned}$$

where in the last line we have used the identity (5.13). Also, (4.1) can be re-written in the form

$$\begin{aligned} dp_t(x) &= \mathcal{L}^* p_t(x) dt + (r-s)p_t(x)(Hx - Hm_t)^\top \Xi^{-1} dZ_t \\ &\quad - \left( (r-s)(Hx - Hm_t)^\top \Xi^{-1} Hm_t dt + \frac{s}{2}(Hx - Hm_t)^2 \Xi^{-1} - \frac{s}{2} C_t H^2 \Xi^{-1} \right) p_t(x) dt \end{aligned}$$

Then using once again using (5.13) we obtain

$$\begin{aligned} &- \left( \frac{(r-s)}{2} ((Hx)^2 - \mathbb{E}_{\rho_t}[(Hx)^2]) \Xi^{-1} - \frac{s}{2} C_t H^2 \Xi^{-1} + \frac{r}{2} (Hx - Hm_t)^2 \Xi^{-1} \right) \\ &= - \left( (r-s)(Hx - Hm_t)^\top \Xi^{-1} Hm_t dt + \frac{s}{2}(Hx - Hm_t)^2 \Xi^{-1} - \frac{s}{2} C_t H^2 \Xi^{-1} \right) \end{aligned}$$

so that the first claim holds.  $\square$

### 5.5 Proof of Lemma 4.3

*Proof.* Starting with (4.10), by similar reasoning as in Lemma 4.1, we have that the conditional forward Kolmogorov has as solution  $\rho_t(x) = \mathcal{N}(m_t, C_t)$  and is given by

$$\partial_t \rho_t(x) = \mathcal{L}^* \rho_t(x) - \nabla \cdot \left( \rho_t(C_t + \epsilon T) H^T \Xi^{-1} \left[ \frac{dZ_t^\delta}{dt} - H \left( \frac{\bar{X}_t + m_t}{2} \right) \right] \right)$$

Starting with the second term on the rhs of the above equation, using the shorthand notation

$$B := (C_t + \epsilon T) H^\top \Xi^{-1}, \quad v := \frac{dZ_t^\delta}{dt} - H \left( \frac{\bar{X}_t + m_t}{2} \right)$$

$$\begin{aligned} -\nabla \cdot (\rho_t(x) B v) &= -\rho_t(x) \nabla \cdot (B v) - (\nabla \rho_t(x)) \cdot B v \\ &= -\rho_t(x) \nabla \cdot (B v) + C_t^{-1} (x - m_t) \cdot (C_t + \epsilon T) H^T \Xi^{-1} \left( \frac{dZ_t^\delta}{dt} - \frac{1}{2} H(x + m_t) \right) \rho_t(x) \\ &= -\rho_t(x) \nabla \cdot (B v) + (C_t^{-1} (x - m_t))^T (C_t + \epsilon T) H^T \Xi^{-1} \left( \frac{dZ_t^\delta}{dt} - \frac{1}{2} H(x + m_t) \right) \rho_t(x) \\ &= -\rho_t(x) \nabla \cdot (B v) + (x - m_t)^T C_t^{-1} (C_t + \epsilon T) H^T \Xi^{-1} \left( \frac{dZ_t^\delta}{dt} - \frac{1}{2} H(x + m_t) \right) \rho_t(x) \\ &= \frac{1}{2} \text{Tr}((C_t + \epsilon T) H^T \Xi^{-1} H) \rho_t(x) + (x - m_t)^T C_t^{-1} (C_t + \epsilon T) H^T \Xi^{-1} \left( \frac{dZ_t^\delta}{dt} - \frac{1}{2} H(x + m_t) \right) \rho_t(x) \\ &= \frac{1}{2} \text{Tr}(C_t H^T \Xi^{-1} H) \rho_t(x) + (x - m_t)^T H^T \Xi^{-1} \left( \frac{dZ_t^\delta}{dt} - \frac{1}{2} H(x + m_t) \right) \rho_t(x) \\ &\quad + \frac{\epsilon}{2} \text{Tr}(T H^T \Xi^{-1} H) \rho_t(x) + \epsilon (x - m_t)^T C_t^{-1} T H^T \Xi^{-1} \left( \frac{dZ_t^\delta}{dt} - \frac{1}{2} H(x + m_t) \right) \rho_t(x) \\ &= -\frac{1}{2} x^T H^T \Xi^{-1} H x + \frac{1}{2} \mathbb{E}_{\rho_t} [(xH)^T \Xi^{-1} H x] + (x - m_t)^T H^T \Xi^{-1} \frac{dZ_t^\delta}{dt} \\ &\quad + \frac{\epsilon}{2} \text{Tr}(T H^T \Xi^{-1} H) \rho_t(x) + \epsilon (x - m_t)^T C_t^{-1} T H^T \Xi^{-1} \left( \frac{dZ_t^\delta}{dt} - \frac{1}{2} H(x + m_t) \right) \rho_t(x) \end{aligned}$$

where in the last line we have used the cyclic property of the trace,

$$\text{Tr}(C_t H^T \Xi^{-1} H) = \mathbb{E}_{\rho_t} [(xH)^T \Xi^{-1} H x] - m_t^\top H^T \Xi^{-1} H m_t$$

Recall the quadratic replicator-mutator with  $r, s = 0$  is given by

$$\begin{aligned}\partial_t p_t(x) &= \mathcal{L}^* p_t(x) + r \left( -\frac{1}{2} \left( x^\top H^\top \Xi^{-1} H x - \mathbb{E}_{p_t} \left[ x^\top H^\top \Xi^{-1} H x \right] \right) + (Hx - Hm_t)^\top \Xi^{-1} \frac{dZ_t^\delta}{dt} \right) p_t(x) \\ &=: \mathcal{L}^* p_t(x) + \mathcal{M}_r p_t(x)\end{aligned}$$

Returning to the forward Kolmogorov equation, we have

$$\partial_t \rho_t(x) = \mathcal{L}^* \rho_t(x) + \mathcal{M}_1 \rho_t(x) + \frac{\epsilon}{2} \text{Tr}(T H^\top \Xi^{-1} H) \rho_t(x) + \epsilon (x - m_t)^\top C_t^{-1} T H^\top \Xi^{-1} \left( \frac{dZ_t^\delta}{dt} - \frac{1}{2} H(x + m_t) \right) \rho_t(x)$$

that is, the rhs coincides with the rhs of the replicator-mutator plus two extra terms. For the special case  $T = C_t$ , we have

$$\begin{aligned}-\nabla \cdot (\rho_t(x) B v) &= -\frac{1}{2} x^\top H^\top \Xi^{-1} H x + \frac{1}{2} \mathbb{E}_{\rho_t} \left[ (xH)^\top \Xi^{-1} H x \right] + (x - m_t)^\top H^\top \Xi^{-1} \frac{dZ_t^\delta}{dt} \\ &\quad + \frac{\epsilon}{2} \left( \mathbb{E}_{\rho_t} \left[ (xH)^\top \Xi^{-1} H x \right] - m_t^\top H^\top \Xi^{-1} H m_t \right) \rho_t(x) + \epsilon (x - m_t)^\top H^\top \Xi^{-1} \left( \frac{dZ_t^\delta}{dt} - \frac{1}{2} H(x + m_t) \right) \rho_t(x) \\ &= (1 + \epsilon) \left( -\frac{1}{2} x^\top H^\top \Xi^{-1} H x + \frac{1}{2} \mathbb{E}_{\rho_t} \left[ (xH)^\top \Xi^{-1} H x \right] + (x - m_t)^\top H^\top \Xi^{-1} \frac{dZ_t^\delta}{dt} \right) \rho_t(x) \\ &= \mathcal{M}_{1+\epsilon} \rho_t(x)\end{aligned}$$

that is, for the special case  $T = C_t$ , the forward Kolmogorov equation for (4.10) is given by

$$\partial_t \rho_t(x) = \mathcal{L}^* \rho_t(x) + \mathcal{M}_{1+\epsilon} \rho_t(x)$$

proving the claim for (4.10). The proof for (4.9) follows from a similar line of reasoning and is therefore omitted.  $\square$

## 5.6 Proof of Lemma 4.4

*Proof.* First recall the evolution equation for the mean of the linear-Gaussian replicator-mutator equation in the limit  $\delta_d \rightarrow 0$ ,

$$dm_t = Gm_t dt - (r - s)K_t(Hm_t dt - dZ_t)$$

Then

$$\begin{aligned}d\varepsilon_t &= dm_t - dX_t^* \\ &= Gm_t dt - (r - s)K_t(Hm_t dt - dZ_t) - GX_t^* dt - bdt + \Sigma^{1/2} dW_t \\ &= Gm_t dt - (r - s)K_t(Hm_t dt - HX_t^* dt + \Xi^{1/2} dB_t) - GX_t^* dt - bdt + \Sigma^{1/2} dW_t \\ &= G\varepsilon_t dt - bdt + \Sigma^{1/2} dW_t - (r - s)K_t H \varepsilon_t dt + (r - s)K_t \Xi^{1/2} dB_t \\ &= (G - (r - s)K_t H) \varepsilon_t dt - bdt + \Sigma^{1/2} dW_t + (r - s)K_t \Xi^{1/2} dB_t\end{aligned}\tag{5.15}$$

and

$$d\mathbb{E}[\varepsilon_t] = (G - (r - s)K_t H) \mathbb{E}[\varepsilon_t] dt - bdt\tag{5.16}$$

from which we obtain

$$\begin{aligned} d\mathbb{E}[\varepsilon_t]\mathbb{E}[\varepsilon_t^\top] &= \mathbb{E}[\varepsilon_t]d\mathbb{E}[\varepsilon_t^\top] + d\mathbb{E}[\varepsilon_t]\mathbb{E}[\varepsilon_t^\top] \\ &= \mathbb{E}[\varepsilon_t]\mathbb{E}[\varepsilon_t^\top](G - (r - s)K_t H)^\top dt + (G - (r - s)K_t H)\mathbb{E}[\varepsilon_t]\mathbb{E}[\varepsilon_t^\top]dt - \mathbb{E}[\varepsilon_t]b^\top dt - b\mathbb{E}[\varepsilon_t^\top]dt \end{aligned}$$

Since  $\tilde{P}_t = \mathbb{E}[\varepsilon_t\varepsilon_t^\top]$  and using Ito formula, we have

$$\begin{aligned} d\tilde{P}_t &= \mathbb{E}[\varepsilon_t(d\varepsilon_t)^\top] + \mathbb{E}[(d\varepsilon_t)\varepsilon_t^\top] + \mathbb{E}[d\varepsilon_t(d\varepsilon_t)^\top] \\ &= \mathbb{E}[\varepsilon_t\varepsilon_t^\top(G - (r - s)K_t H)^\top dt - \varepsilon_t b^\top dt] + \mathbb{E}[(G - (r - s)K_t H)\varepsilon_t\varepsilon_t^\top dt - b\varepsilon_t^\top dt] + \Sigma dt + (r - s)^2 K_t \Xi K_t^\top dt \\ &= \tilde{P}_t(G - (r - s)K_t H)^\top dt + (G - (r - s)K_t H)\tilde{P}_t dt - \mathbb{E}[\varepsilon_t]b^\top dt - b\mathbb{E}[\varepsilon_t^\top]dt + \Sigma dt + (r - s)^2 K_t \Xi K_t^\top dt \end{aligned}$$

which yields (4.15). Using,  $P_t = \tilde{P}_t - \mathbb{E}[\varepsilon_t]\mathbb{E}[\varepsilon_t^\top]$  and combining the above expressions yields (4.14).

For the second claim, consider that when  $s = 0$ , we have

$$dP_t = (G - rK_t H)P_t dt + P_t(G - rK_t H)^\top dt + \Sigma dt + r^2 K_t \Xi K_t^\top dt$$

whose solution at any given time  $t$  is formally equivalent to  $C_t$  (whose time evolution is given by 4.4) if  $P_0 = C_0$  and  $r = 1$ , since  $K_t \Xi K_t^\top = K_t H C_t$ . Notice that due to the  $r^2 K_t \Xi K_t^\top$  rather than  $rK_t \Xi K_t^\top$  term, this equivalence only holds when  $r = 1$  in addition to  $s = 0$ .

Finally, for (4.16), since  $Tr(\tilde{P}_t) = E_t$ , we start with

$$\frac{dTr(\tilde{P}_t)}{dt} = 2Tr(A_t^\top \tilde{P}_t) - 2Tr(\mathbb{E}[\varepsilon_t]b^\top) + Tr(\Sigma) + (r - s)^2 Tr(K_t \Xi K_t^\top)$$

where  $A_t(r, s) := G - (r - s)K_t H$ . Note that by the cyclic property of the trace and using  $Tr(AB^\top B) \leq \lambda_{max}(A)\|B\|_F$  for any  $A$  p.d.,

$$Tr(K_t \Xi K_t^\top) \leq \lambda_{max}(H^\top \Xi^{-1} H)\|C_t\|_F$$

if  $\Xi^{-1/2}H$  is an invertible matrix. For the remainder of the proof, we drop the  $(r, s)$  in  $A_t(r, s)$ . Also,

$$Tr(A_t^\top \tilde{P}_t) \leq \|A_t^\top\|_F \|\tilde{P}_t\|_F \leq \|A_t^\top\|_F Tr(\tilde{P}_t)$$

Note also that we have an explicit solution for  $\mathbb{E}[\varepsilon_t]$ , since starting from (5.15),

$$d\mathbb{E}[\varepsilon_t] = (G - (r - s)K_t H)\mathbb{E}[\varepsilon_t]dt - bdt \quad (5.17)$$

$$\mathbb{E}[\varepsilon_t] = \exp\left(\int_0^t A_u du\right)\mathbb{E}[\varepsilon_0] - \int_0^t \exp\left(\int_u^t A_v dv\right)b du \quad (5.18)$$

combining all yields (4.16). □

## 5.7 Proof of Lemma 4.5

*Proof.* When  $C_0 = C_\infty$  we have that  $K_t = K_\infty = C_\infty H^\top \Xi^{-1}$  for all  $t \geq 0$ . Using the evolution equation for the mean error (5.16), we obtain the explicit solution for  $\mathbb{E}[\varepsilon_t]$ , using the shorthand notation  $A_\infty(r, s) = G - (r - s)K_\infty H$ ,

$$\mathbb{E}[\varepsilon_t] = \exp(tA_\infty(s, r))\mathbb{E}[\varepsilon_0] + A_\infty^{-1}(s, r)[I - \exp(tA_\infty(s, r))]b \quad (5.19)$$

When  $A_\infty(r, s) < 0$ , it holds that  $t \rightarrow \infty$ ,  $\mathbb{E}[\varepsilon_t] \rightarrow -A_\infty(r, s)^{-1}b$ , from which we obtain (4.18). For the scalar case, (4.17) is explicitly solvable with

$$C_\infty = \frac{G + \sqrt{G^2 + rH^2\Xi^{-1}\Sigma}}{rH^2\Xi^{-1}} \quad (5.20)$$

and

$$\begin{aligned} A_\infty(s, r) &= G - \frac{(r-s)}{r} \left( G + \sqrt{G^2 + rH^2\Xi^{-1}\Sigma} \right) \\ &= \frac{s}{r}G - \frac{(r-s)}{r} \sqrt{G^2 + rH^2\Xi^{-1}\Sigma} \end{aligned} \quad (5.21)$$

Notice that  $A_\infty(s, r)$  is a weighted linear combination  $G$  and  $A_\infty(0, r) = -\sqrt{G^2 + rH^2\Xi^{-1}\Sigma}$ , i.e.

$$A_\infty(s, r) = \frac{s}{r}G + \frac{(r-s)}{r}A_\infty(0, r)$$

where from now onwards we use the shorthand notation  $A_\infty(0) := A_\infty(s=0, r)$ . Then returning to (4.15), we have

$$\frac{d\tilde{P}_t}{dt} = 2A_\infty(r, s)\tilde{P}_t + \mathcal{B}_t$$

where

$$\mathcal{B}_t := \Sigma + (r-s)^2(K_\infty)^2\Xi - 2\mathbb{E}[\varepsilon_t]b$$

which has explicit solution

$$\begin{aligned} \tilde{P}_t &= \exp(t2A_\infty(r, s))\tilde{P}_0 + \int_0^t \exp((t-u)2A_\infty(r, s)) \mathcal{B}_u du \\ &= \exp(t2A_\infty(r, s))\tilde{P}_0 + (\Sigma + (r-s)^2K_\infty^2\Xi) \int_0^t \exp((t-u)2A_\infty(r, s)) du - 2b \int \exp((t-u)2A_\infty(r, s)) \mathbb{E}[\varepsilon_u] du \end{aligned}$$

to simplify further, we need to evaluate (dropping dependence on  $r, s$  in  $A_\infty(r, s)$ ),

$$\begin{aligned} \int_0^t \exp((t-u)2A_\infty) \mathbb{E}[\varepsilon_u] du &= \int_0^t \exp((t-u)2A_\infty) [\exp(uA_\infty)\mathbb{E}[\varepsilon_0] + A_\infty^{-1}[I - \exp(uA_\infty)]b] du \\ &= \exp(2tA_\infty) \int_0^t \exp(-uA_\infty) (\mathbb{E}[\varepsilon_0] - A_\infty^{-1}b) + A_\infty^{-1}b \exp(-2uA_\infty) du \\ &= \exp(2tA_\infty) \left( -(\mathbb{E}[\varepsilon_0] - A_\infty^{-1}b)A_\infty^{-1}(\exp(-tA_\infty) - 1) - \frac{1}{2}A_\infty^{-2}b(\exp(-2tA_\infty) - 1) \right) \\ &= A_\infty^{-2}b \left( \exp(tA_\infty) - \exp(2tA_\infty) - \frac{1}{2} + \frac{1}{2} \exp(2tA_\infty) \right) \\ &= A_\infty^{-2}b \left( \exp(tA_\infty) - \frac{1}{2} \exp(2tA_\infty) - \frac{1}{2} \right) \\ &= -\frac{1}{2}A_\infty^{-2}b(1 - \exp(tA_\infty))^2 \end{aligned}$$

where in the last line we set  $\mathbb{E}[\varepsilon_0] = 0$  for simplicity. Returning to the explicit solution for  $\tilde{P}_t$ ,

$$\tilde{P}_t = \exp(t2A_\infty)\tilde{P}_0 - \frac{1}{2}(\Sigma + (r-s)^2K_\infty^2\Xi)A_\infty^{-1}(I - \exp(2tA_\infty)) + A_\infty^{-2}b^2(1 - \exp(tA_\infty))^2 \quad (5.22)$$

from which (4.19) follows immediately whenever  $A_\infty < 0$  and using  $(r-s)K_\infty = \frac{G-A_\infty}{H}$ .  $\square$



## 5.8 Proof of Lemma 4.6

*Proof.* In the below we drop the  $(s, r)$  dependency in  $A_\infty(s, r)$  where there is no ambiguity. Recall also that in the scalar case  $E_t = \tilde{P}_t$ , and we will use the notation  $\tilde{P}_\infty$  in the below to refer to the asymptotic mse. Since we require  $A_\infty < 0$  to ensure the existence of  $\tilde{P}_\infty$ , any choice of  $s$  must satisfy

$$\frac{s}{r}G - \frac{(r-s)}{r}\sqrt{G^2 + rH^2\Xi^{-1}\Sigma} < 0$$

from which the upper bound in (4.25) immediately follows (along with the requirement that  $s < r$ ). Before characterising the lower bound, we obtain an expression for the optimal  $s$  for any given  $r > 0$  minimising (4.19) by solving

$$\frac{\partial \tilde{P}_\infty}{\partial s} = \frac{\partial \tilde{P}_\infty}{\partial A_\infty} \frac{\partial A_\infty}{\partial s} = 0$$

where

$$\begin{aligned} \frac{\partial \tilde{P}_\infty}{\partial A_\infty} &= -2b^2 A_\infty^{-3} + A_\infty^{-1} \left( \frac{G - A_\infty}{H} \right) \frac{\Xi}{H} + \frac{1}{2} A_\infty^{-2} \left( \Sigma + \left( \frac{G - A_\infty}{H} \right)^2 \Xi \right) \\ \frac{\partial A_\infty}{\partial s} &= \frac{G}{r} + \frac{\sqrt{G^2 + rH^2\Xi^{-1}\Sigma}}{r} \end{aligned}$$

Since  $\frac{\partial A_\infty}{\partial s}$  is clearly never zero, we need only solve

$$g(A_\infty) := A_\infty^3 - \frac{(H^2\Sigma + \Xi G^2)}{\Xi} A_\infty + \frac{4b^2 H^2}{\Xi} = 0 \quad (5.23)$$

which takes the form of a depressed cubic  $A_\infty^3 + pA_\infty + q = 0$  with  $p = -(H^2\Xi^{-1}\Sigma + G^2)$ ,  $q = 4b^2 H^2 \Xi^{-1}$ , and we have that  $q > 0, p < 0$  always. The nature of the roots can be characterised in the usual way via the discriminant  $\tau$ , i.e. when  $\tau > 0$ , (5.23) has one real root and two complex roots, whilst when  $\tau < 0$ , it has three real roots. We deal with these two cases below separately.

**Case 1:**  $\tau > 0$ . First we show that the real root here is strictly negative. From inspection, we have that  $g(0) > 0$ . Also, (5.23) always has two real extremal (turning) points, one of which is strictly negative and the other is strictly positive, since  $g'(A_\infty) = 3A_\infty^2 + p = 0$  implies the extremal points occur at  $A_\infty = \pm\sqrt{\frac{-p}{3}}$ . Furthermore,  $g'(0) = p < 0$ . Combining these properties implies that (5.23) has one negative real root when  $\tau > 0$ . When  $\tau > 0$ , we can use Cardano's formula to obtain an expression for the real root,

$$A_\infty^* = \left( -\frac{q}{2} + \sqrt{\tau} \right)^{1/3} + \left( -\frac{q}{2} - \sqrt{\tau} \right)^{1/3}$$

from which we then obtain (4.24), and substituting the expression for  $A_\infty^*$  into (4.20) and rearranging yields (4.26). It can be verified straightforwardly that this is indeed a minimum since  $\frac{\partial^2 \tilde{P}_\infty}{\partial A_\infty^2}(A_\infty^*) > 0$ . In order to characterise the lower bound  $s^l$  and to obtain  $r^{\text{opt}}$ , we work with the following change of variable

$$y := \sqrt{G^2 + rH^2\Xi^{-1}\Sigma} \quad (5.24)$$

which when substituted into (4.26) with  $r = \frac{y^2 - G^2}{H^2 \Xi^{-1} \Sigma}$  yields

$$s^{\text{opt}}(y) = \frac{1}{H^2 \Xi^{-1} \Sigma} (y - G)(A_\infty^* + y)$$

The minimiser of  $s^{\text{opt}}(y)$  wrt  $y$ , which we denote by  $y^*$ , can be found straightforwardly by solving  $\frac{ds^{\text{opt}}(y)}{dy} = 0$ , giving

$$y^* = \frac{G - A_\infty^*}{2}$$

which implies that for any  $r > 0$ ,

$$s^{\text{opt}} \geq -\frac{(G + A_\infty^*)^2}{4H^2 \Xi \Sigma} =: s^l$$

since  $y$  is monotonically related to  $r$  (notice also that  $s^l$  is independent of  $r$ . Furthermore,  $y^* > 0$  for any choice of  $G, H, \Sigma, \Xi$  satisfying the conditions of the lemma since from the definition of  $A_\infty$ ,  $G - A_\infty = \frac{(r-s)}{r}(G + \sqrt{G^2 + rH^2 \Xi^{-1} \Sigma}) > 0$ . Therefore, the calculated minimum point is admissible.

Finally, we can obtain expressions for  $r^{\text{opt}}$  for a given  $s$  satisfying (4.25) by solving

$$y^2 + (A_\infty^* - G)y - GA_\infty^* - sH^2 \Xi^{-1} \Sigma = 0 \quad (5.25)$$

which combined with (5.24) gives

$$y = \frac{(G - A_\infty^*) \pm \sqrt{(G - A_\infty^*)^2 + 4(GA_\infty^* + sH^2 \Xi^{-1} \Sigma)}}{2}$$

Since  $r > 0$ , we can only consider  $y > |G|$ , and since  $s^{\text{opt}}(y)$  is a convex quadratic in  $y$ , it holds that for  $s^l < s < s^{\text{opt}}(|G|)$ , we have two possible values for the optimal  $y$ ,

$$y = \frac{(G - A_\infty^*) \pm \sqrt{(G - A_\infty^*)^2 + 4(GA_\infty^* + sH^2 \Xi^{-1} \Sigma)}}{2}$$

where  $s^{\text{opt}}(|G|) = \frac{G^2 + (A_\infty^* - G)|G| - GA_\infty^*}{H}$ , whilst for  $s^{\text{opt}}(|G|) < s < \min\left(r, \frac{\sqrt{G^2 + rH^2 \Xi^{-1} \Sigma}}{G + \sqrt{G^2 + rH^2 \Xi^{-1} \Sigma}} r\right)$ , there is only one optimal  $y$  value given by

$$y = \frac{(G - A_\infty^*) + \sqrt{(G - A_\infty^*)^2 + 4(GA_\infty^* + sH^2 \Xi^{-1} \Sigma)}}{2}$$

Combining these expressions with (5.24) and rearranging yields the final claim of the lemma.

**Case 2:**  $\tau < 0$ . In this case, (5.23) has three real roots due to the usual condition on the discriminant. First we characterise the signs of the roots. We have directly from (5.23) that  $g(0) > 0$ . Also, (5.23) always has two real extremal (turning) points, one of which is strictly negative and the other is strictly positive, since  $g'(A_\infty) = 3A_\infty^2 + p = 0$  implies the extremal points occur at  $A_\infty = \pm \sqrt{\frac{-p}{3}}$ . Furthermore,  $g'(0) = p < 0$ . Combining these properties implies that (5.23) has two positive real roots and one negative real root when  $\tau < 0$ .

When  $\tau < 0$ , the following trigonometric formula holds for the characterisation of the three real roots,  $A_\infty^k$ ,  $k = 0, 1, 2$

$$A_\infty^k = 2\sqrt{-\frac{p}{3}} \cos \left[ \frac{1}{3} \cos^{-1} \left( \frac{3q}{2p} \sqrt{\frac{-3}{p}} \right) - \frac{2\pi k}{3} \right] \quad \text{for } k = 0, 1, 2$$

To determine the negative root, first notice that  $-\frac{2\pi k}{3} < \frac{1}{3} \cos^{-1}(y) - \frac{2\pi k}{3} < -\frac{2\pi k}{3} + \frac{\pi}{3}$  for all  $-1 < y < 1$ . Furthermore,  $\cos(\theta) < 0$  for all  $-\frac{3\pi}{2} < \theta < -\frac{\pi}{2}$  and it holds that for  $\theta := \frac{1}{3} \cos^{-1}(y) - \frac{2\pi k}{3}$  and  $k = 2$ ,  $-\frac{4}{3}\pi < \theta < -\pi$ , so that  $\cos(\theta) < 0$  for  $k = 2$ . This yields the remaining case in (4.24).  $\square$

## 5.9 Proof of Lemma 4.7

*Proof.* The first claim follows from a simple re-arrangement of the identity  $C_\infty = \tilde{P}_\infty$  with  $A_\infty = A_\infty^*$ , i.e.

$$-\frac{1}{2} \left( \Sigma + \left( \frac{G - A_\infty^*}{H} \right)^2 \Xi \right) \frac{1}{A_\infty^*} + \left( \frac{b}{A_\infty^*} \right)^2 = \frac{G + \sqrt{G^2 + rH^2\Xi^{-1}\Sigma}}{rH^2\Xi^{-1}}$$

additionally, rearranging (5.21) yields

$$G + \sqrt{G^2 + rH^2\Xi^{-1}\Sigma} = \frac{r(G - A_\infty^*)}{(r - s)}$$

substituting into the expression for  $\tilde{P}_\infty = C_\infty$  yields the result. Additionally, we have

$$r = \frac{1}{4H^2\Xi^{-1}\Sigma} \left( G - A_\infty^* \pm \sqrt{(G - A_\infty^*)^2 + 4(GA_\infty^* + sH^2\Xi^{-1}\Sigma)} \right)^2 - \frac{G^2}{H^2\Xi^{-1}\Sigma}$$

so that

$$\begin{aligned} r - s &= \frac{1}{4H^2\Xi^{-1}\Sigma} \left( G - A_\infty^* \pm \sqrt{(G - A_\infty^*)^2 + 4(GA_\infty^* + sH^2\Xi^{-1}\Sigma)} \right)^2 - \frac{G^2}{H^2\Xi^{-1}\Sigma} - s \\ &= \frac{(A_\infty^*)^2(G - A_\infty^*)}{-0.5A_\infty^* (H^2\Xi^{-1}\Sigma + (G - A_\infty^*)^2) + b^2H^2\Xi^{-1}} \end{aligned}$$

For the second claim, when  $s = 0$ , it follows directly from (4.26) that

$$r_0^{\text{opt}} = \frac{(A_\infty^*)^2 - G^2}{H^2\Xi^{-1}\Sigma} \quad (5.26)$$

First establish the following bound on  $A_\infty^*$ ,

$$\begin{aligned} A_\infty^* &= \left( -\frac{q}{2} + \sqrt{\tau} \right)^{1/3} + \left( -\frac{q}{2} - \sqrt{\tau} \right)^{1/3} \\ &< \left( -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}} \right)^{1/3} + \left( -\frac{q}{2} - \sqrt{\tau} \right)^{1/3} \\ &< \left( -\left( -\frac{p^3}{27} \right)^{1/2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}} \right)^{1/3} + \left( -\frac{q}{2} - \sqrt{\tau} \right)^{1/3} \\ &< \left( -\left( -\frac{p^3}{27} \right)^{1/2} \right)^{1/3} + \left( -\frac{q}{2} - \sqrt{\tau} \right)^{1/3} \end{aligned}$$

recalling that  $p = -(H^2\Xi^{-1}\Sigma + G^2)$ ,  $q = 4b^2H^2\Xi^{-1}$ . The condition  $\tau > 0$  also implies

$$-\frac{q}{2} < -\left( -\frac{p^3}{27} \right)^{1/2}$$

Then using the inequality  $\sqrt{2}(a_1 + a_2)^{1/2} \geq \sqrt{a_1} + \sqrt{a_2}$ ,

$$-\frac{q}{2} - \sqrt{\tau} < -\frac{q}{2} - \frac{1}{\sqrt{2}} \left( \frac{q}{2} + \left| \frac{p^3}{27} \right|^{1/2} \right) = -\frac{\sqrt{2} + 1}{\sqrt{2}} \frac{q}{2} - \frac{1}{\sqrt{2}} \left| \frac{p^3}{27} \right|^{1/2} < -\left( \frac{\sqrt{2} + 2}{\sqrt{2}} \right) \left| \frac{p^3}{27} \right|^{1/2}$$

Putting altogether, we have

$$A_\infty^* < -\frac{2\sqrt{2} + 2}{\sqrt{6}} \sqrt{|p|}$$

Then letting  $c := \frac{2\sqrt{2}+2}{\sqrt{6}}$  and substituting into (5.26) yields

$$(A_\infty^*)^2 > c^2(H^2\Xi^{-1}\Sigma + G^2)$$

and

$$r_0^{\text{opt}} > \frac{(c^2 - 1)G^2 + c^2H^2\Xi^{-1}\Sigma}{H^2\Xi^{-1}\Sigma} > \frac{(c^2 - 1)(G^2 + H^2\Xi^{-1}\Sigma)}{H^2\Xi^{-1}\Sigma} > 1$$

since  $c^2 > 1$ , which yields (4.31). Then for the remainder of the claim, we have

$$\begin{aligned} \frac{C_\infty}{\hat{C}_\infty} &= \frac{1}{r} \left( \frac{G + \sqrt{G^2 + rH^2\Xi^{-1}\Sigma}}{G + \sqrt{G^2 + H^2\Xi^{-1}\Sigma}} \right) \\ &\leq \frac{1}{r} \left( \frac{2|G| + \sqrt{rH^2\Xi^{-1}\Sigma}}{G + \sqrt{G^2 + H^2\Xi^{-1}\Sigma}} \right) \\ &= \frac{1}{r} \frac{2|G|}{(G + \sqrt{G^2 + H^2\Xi^{-1}\Sigma})} + \frac{1}{\sqrt{r}} \frac{\sqrt{H^2\Xi^{-1}\Sigma}}{(G + \sqrt{G^2 + H^2\Xi^{-1}\Sigma})} \\ &\leq \frac{1}{\sqrt{r}} \left( \frac{2|G| + \sqrt{H^2\Xi^{-1}\Sigma}}{(G + \sqrt{G^2 + H^2\Xi^{-1}\Sigma})} \right) \end{aligned}$$

whenever  $r > 1$ . □

## 6 Conclusion

We presented a detailed investigation of connections between continuous time, continuous trait Crow-Kimura replicator-mutator dynamics [Kim65] and the fundamental equation of nonlinear filtering, the Kushner-Stratonovich partial differential equation. Inspired by a non-local fitness functional presented in the mathematical biology literature [CHR06], we extended this connection to obtain a “modified” Kushner-Stratonovich equation. This equation was shown to be beneficial for filtering with misspecified models and a specific choice of parameters in the fitness functional was shown to coincide with covariance inflated Kalman Bucy filtering, in the linear-Gaussian setting. Additionally, we considered the misspecified model filtering problem, with linear-Gaussian dynamics and where the misspecification arises through an unknown constant bias in the signal dynamics. We proved that through a judicious choice of parameters in the fitness functional, mean squared error and uncertainty quantification (through the covariance) could be improved via this modified Kushner-Stratonovich equation. Estimation is improved over traditional covariance inflation techniques, as well as over the standard filtering setup (assuming perfect model knowledge).

There are several avenues for further work, most notably, the analysis on misspecified models in Section 4 has primarily focused on the scalar setting which has simplified the analysis. In future works, the multivariate setting, as well as extensions to nonlinear dynamics should be explored. Additionally, it would be worthwhile to extend the mode of convergence in Theorem 3.1 to  $L^p$  convergence rather than pointwise convergence.

## Acknowledgements

SP gratefully acknowledges PW for discussions on this work and for providing supporting simulations.

## A Technical lemmata

**Theorem A.1. Forward representation formula** (Theorem 3.1 in [Kun82].) *Let  $u_t(x, \omega)$  for  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  denote a measurable stochastic process on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with  $\omega \in \Omega$  (from now on we suppress the  $\omega$  notation). Suppose its time evolution is given by*

$$\partial_t u_t(x) = \mathcal{L}u_t(x) + \sum_{k=1}^n \mathcal{M}^k u_t(x) \circ dB_t^k, \quad u_0(x) = f(x) \quad (1.1)$$

where  $B_t^k$  is a Brownian motion wrt  $\mathbb{P}$  and  $f$  a bounded  $C^2$  function with bounded derivatives and

$$\begin{aligned} \mathcal{L} &:= \frac{1}{2} \sum_{j=1}^m \left( \sum_{i=1}^m a^{ij}(x) \frac{\partial}{\partial x^i} \right)^2 + \sum_{i=1}^m b^i(x) \frac{\partial}{\partial x^i} + c^0 \\ &= \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \left( \sum_k a^{ik} a^{jk} \right) \frac{\partial^2}{\partial x^i \partial x^j} + \frac{1}{2} \sum_{i,j} a^{ij} \sum_k \left( \frac{\partial a^{kj}}{\partial x^i} \right) \frac{\partial}{\partial x^k} + \sum_{i=1}^m b^i(x) \frac{\partial}{\partial x^i} + c^0 \\ \mathcal{M}^k &:= \sum_{i=1}^m l^{ik}(x) \frac{\partial}{\partial x^i} + c^k \end{aligned}$$

with  $c^k(x)$  being uniformly bounded  $C^2$  functions with bounded derivatives in  $x$  and  $a^{ij}(x), b^i(x), m^{ik}(x)$  being uniformly bounded  $C^4$  functions with bounded first derivatives in  $x$ .

Then there exists another probability space  $(\tilde{\Omega}, \mathcal{B}, \mathbb{Q})$  on which the Brownian motion  $W_t = [W_t^1, \dots, W_t^m]^\top$  is defined and an SDE on the product space  $(\Omega \otimes \tilde{\Omega}, \mathcal{F} \otimes \mathcal{B}, \mathbb{P} \otimes \mathbb{Q})$  given by

$$d\xi_t(x) = \sum_{i=1}^d b^i(\xi_t(x)) dt + \sum_{j=1}^m \sum_{i=1}^m a^{ij}(\xi_t(x)) \circ dW_t^j + \sum_{k=1}^n \sum_{i=1}^m l^{ik}(\xi_t(x)) \frac{\partial \xi_t(x)}{\partial x^i} \circ dB_t^k \quad (1.2)$$

where the notation  $\xi_t(x)$  is used to denote the solution of the SDE with initial condition  $\xi_0 = x$  and  $t > 0$ . The solution of (1.1) has the representation

$$u_t(x) = \mathbb{E}^{\mathbb{Q}} \left[ f(\xi_t(x)) \exp \left( \sum_{k=1}^n \int_0^t c^k(\xi_s(x)) \circ dB_s^k + \int_0^t c^0(\xi_s(x)) ds \right) \right], \quad x \in \mathbb{R}^m$$

The following form of Ito-Stratonovich correction will also be useful. Consider the following Stratonovich SDE

$$d\xi_t = X_0(t, \xi_t)dt + \sum_{j=1}^m X_j(t, \xi_t) \circ dW_t^j$$

where  $X_j$  is a lipschitz cts function with derivatives in the second argument up to second order. Let  $\xi_{0,r}(x)$  denote the solution of the above Stratonovich SDE at time  $r$  with initial condition  $\xi_0 = x$ . Then the following relation between the ito and stratonovich integral holds (backward form!)

$$\int_s^t X^k(f(\xi_{r,t}(x))) \circ dB_r^k = \int_s^t X^k(f(\xi_{r,t}(x)))dW_r^k + \frac{1}{2} \int_s^t (X^k(f(\xi_{r,t}(x))))^2 dr \quad (1.3)$$

where  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is a  $C^3$  function.

Some well-known results from stochastic analysis are presented below.

**Lemma A.1. Moment generating function.** Consider

$$Y_t = \int_0^t \alpha_s dW_s$$

where  $\alpha_s$  is a deterministic, scalar-valued continuous function of  $s$  and  $W_t$  is a real-valued Brownian motion wrt a probability measure  $\mathbb{P}$ . Then

$$\mathbb{E}^{\mathbb{P}}[\exp(\lambda Y_t)] = \exp\left(\frac{\lambda}{2} \int_0^t \alpha_s^2 ds\right)$$

**Lemma A.2.** Consider a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  on which a scalar valued Wiener process is defined. Suppose  $f(s, \omega)$  for  $\omega \in \Omega$  is an  $\mathcal{F}_t$ -adapted process. Then the following inequality holds for  $q > 1$

$$\mathbb{E} \left[ \left| \int_0^t f(s, \omega) dW_s \right|^{2q} \right] \leq t^{q-1} (q(2q-1))^q \mathbb{E} \left[ \int_0^t |f(s, \omega)|^{2q} ds \right] \quad (1.4)$$

The following well-known lemma will also be used

**Lemma A.3. Continuity of solutions of SDEs** Let  $g, \sigma$  be lipschitz continuous functions satisfying linear growth conditions. Denote by  $\xi_t(x)$  the unique solution to

$$d\xi_t = g(\xi_s)ds + \sigma(\xi_s)dB_s$$

with  $\xi_0 = x$ . Suppose  $h$  is a globally lipschitz continuous real vector valued function. Then it holds that

$$\mathbb{E} [|h(\xi_s(x)) - h(\xi_r(x))|^q] \leq C(x) |s - r|^{q/2}$$

**Lemma A.4. Exponential moment bounds.** Let  $Y_s$  for  $s \in [0, t]$  denote a non-decreasing scalar-valued adapted process such that  $\mathbb{E}[Y_t - Y_s | \mathcal{F}_s] \leq K$  for all  $s \in [0, t]$ . Then for any  $L < \frac{1}{K}$ ,

$$\mathbb{E}[\exp(LY_t)] < \frac{1}{1 - LK}$$

## References

- [AC14] Matthieu Alfaro and Rémi Carles. “Explicit solutions for replicator-mutator equations: extinction vs. acceleration”. en. In: *SIAM Journal on Applied Mathematics* 74.6 (Jan. 2014). arXiv:1405.2768 [math, q-bio], pp. 1919–1934. ISSN: 0036-1399, 1095-712X. DOI: [10.1137/140979411](https://doi.org/10.1137/140979411). URL: <http://arxiv.org/abs/1405.2768> (visited on 01/24/2024).
- [AC17] Matthieu Alfaro and Rémi Carles. “Replicator-mutator equations with quadratic fitness”. In: *Proceedings of the American Mathematical Society* 145.12 (2017), pp. 5315–5327. ISSN: 0002-9939. DOI: [10.1090/proc/13669](https://doi.org/10.1090/proc/13669). arXiv: [1611.06119](https://arxiv.org/abs/1611.06119).
- [Aki79] E Akin. “The Geometry of Population Genetics”. In: *Lec. Notes in Biomath.* 31 (1979).
- [Aky17] Ömer Deniz Akyıldız. *A probabilistic interpretation of replicator-mutator dynamics*. 2017. URL: <http://arxiv.org/abs/1712.07879>.
- [And07] Jeffrey L. Anderson. “An adaptive covariance inflation error correction algorithm for ensemble filters”. en. In: *Tellus A: Dynamic Meteorology and Oceanography* 59.2 (Jan. 2007), p. 210. ISSN: 1600-0870. DOI: [10.1111/j.1600-0870.2006.00216.x](https://doi.org/10.1111/j.1600-0870.2006.00216.x). (Visited on 03/05/2024).
- [Bae21] John C Baez. “The fundamental theorem of natural selection”. In: *Entropy* 23.11 (2021), p. 1436.
- [BBH83] J. Baras, G. Blankenship, and W. Hopkins. “Existence, uniqueness, and asymptotic behavior of solutions to a class of Zakai equations with unbounded coefficients”. en. In: *IEEE Transactions on Automatic Control* 28.2 (Feb. 1983), pp. 203–214. ISSN: 0018-9286. DOI: [10.1109/TAC.1983.1103218](https://doi.org/10.1109/TAC.1983.1103218). URL: <http://ieeexplore.ieee.org/document/1103218/> (visited on 11/18/2024).
- [BC09] Alan Bain and Dan Crisan. *Fundamentals of Stochastic Filtering*. en. Ed. by B. Rozovskiĭ and G. Grimmett. Vol. 60. Stochastic Modelling and Applied Probability. New York, NY: Springer New York, 2009. ISBN: 978-0-387-76895-3 978-0-387-76896-0. DOI: [10.1007/978-0-387-76896-0](https://doi.org/10.1007/978-0-387-76896-0). URL: <http://link.springer.com/10.1007/978-0-387-76896-0> (visited on 04/23/2024).
- [BD19] Adrian N. Bishop and Pierre Del Moral. “On the stability of matrix-valued Riccati diffusions”. In: *Electronic Journal of Probability* 24 (2019), pp. 1–40. ISSN: 10836489. DOI: [10.1214/19-ejp342](https://doi.org/10.1214/19-ejp342). arXiv: [1808.00235](https://arxiv.org/abs/1808.00235).
- [BD23] Adrian N. Bishop and Pierre Del Moral. “On the mathematical theory of ensemble (linear-Gaussian) Kalman–Bucy filtering”. en. In: *Mathematics of Control, Signals, and Systems* 35.4 (Dec. 2023), pp. 835–903. ISSN: 0932-4194, 1435-568X. DOI: [10.1007/s00498-023-00357-2](https://doi.org/10.1007/s00498-023-00357-2). URL: <https://link.springer.com/10.1007/s00498-023-00357-2> (visited on 01/30/2024).
- [BKK95] Abhay G. Bhatt, G. Kallianpur, and Rajeeva L. Karandikar. “Uniqueness and Robustness of Solution of Measure-Valued Equations of Nonlinear Filtering”. en. In: *The Annals of Probability* 23.4 (Oct. 1995). ISSN: 0091-1798. DOI: [10.1214/aop/1176987808](https://doi.org/10.1214/aop/1176987808). (Visited on 11/18/2024).
- [Blö+22] Dirk Blömker et al. “Continuous time limit of the stochastic ensemble Kalman inversion: strong convergence analysis”. In: *SIAM Journal on Numerical Analysis* 60.6 (2022), pp. 3181–3215. DOI: [10.1137/21M1437561](https://doi.org/10.1137/21M1437561).

- [BMP18] Adrian N Bishop, Pierre del Moral, and Sahani Pathiraja. “Perturbations and projections of Kalman–Bucy semigroups”. en. In: *Stochastic Processes and their Applications* 128 (2018). DOI: <https://doi.org/10.1016/j.spa.2017.10.006>.
- [Bom90] Immanuel M Bomze. “Dynamical aspects of evolutionary stability”. In: *Monatshefte für Mathematik* 110 (1990), pp. 189–206.
- [BP16] John C Baez and Blake S Pollard. “Relative entropy in biological systems”. In: *Entropy* 18.2 (2016), p. 46.
- [BR12] Kay Bergemann and Sebastian Reich. “An ensemble Kalman–Bucy filter for continuous data assimilation”. In: *Meteorologische Zeitschrift* 21.3 (2012), pp. 213–219. ISSN: 09412948. DOI: [10.1127/0941-2948/2012/0307](https://doi.org/10.1127/0941-2948/2012/0307).
- [CCK24] Nicolas Chopin, Francesca Crucinio, and Anna Korba. “A connection between Tempering and Entropic Mirror Descent”. In: *Proceedings of the 41st International Conference on Machine Learning*. Ed. by Ruslan Salakhutdinov et al. Vol. 235. Proceedings of Machine Learning Research. PMLR, 21–27 Jul 2024, pp. 8782–8800. URL: <https://proceedings.mlr.press/v235/chopin24a.html>.
- [Cha+20] Fabio A. C. C. Chalub et al. *Gradient flow formulations of discrete and continuous evolutionary models: a unifying perspective*. en. arXiv:1907.01681 [math, q-bio]. Oct. 2020. URL: <http://arxiv.org/abs/1907.01681> (visited on 01/04/2024).
- [Cha+21] Fabio A.C.C. Chalub et al. “Gradient Flow Formulations of Discrete and Continuous Evolutionary Models: A Unifying Perspective”. In: *Acta Applicandae Mathematicae* 171.1 (2021), pp. 1–68. ISSN: 15729036. DOI: [10.1007/s10440-021-00391-9](https://doi.org/10.1007/s10440-021-00391-9). arXiv: [1907.01681](https://arxiv.org/abs/1907.01681).
- [Che+24a] Yifan Chen et al. *Efficient, Multimodal, and Derivative-Free Bayesian Inference With Fisher-Rao Gradient Flows*. en. arXiv:2406.17263 [cs]. Oct. 2024. URL: <http://arxiv.org/abs/2406.17263> (visited on 11/20/2024).
- [Che+24b] Yifan Chen et al. *Sampling via Gradient Flows in the Space of Probability Measures*. en. arXiv:2310.03597 [stat]. Mar. 2024. URL: <http://arxiv.org/abs/2310.03597> (visited on 11/18/2024).
- [CHR06] Ross Cressman, Josef Hofbauer, and Frank Riedel. “Stability of the replicator equation for a single species with a multi-dimensional continuous trait space”. In: *Journal of Theoretical Biology* 239.2 (2006), pp. 273–288. DOI: [10.1016/j.jtbi.2005.07.022](https://doi.org/10.1016/j.jtbi.2005.07.022).
- [CK70] J.F. Crow and M. Kimura. *An introduction to Population Genetics Theory*. Harper and Row, 1970.
- [Cri+13] D. Crisan et al. “Robust filtering: Correlated noise and multidimensional observation”. en. In: *The Annals of Applied Probability* 23.5 (Oct. 2013). arXiv:1201.1858 [math]. ISSN: 1050-5164. DOI: [10.1214/12-AAP896](https://doi.org/10.1214/12-AAP896). URL: <http://arxiv.org/abs/1201.1858> (visited on 01/04/2024).
- [CS09] Fabio A.C.C. Chalub and Max O. Souza. “From discrete to continuous evolution models: A unifying approach to drift-diffusion and replicator dynamics”. In: *Theoretical Population Biology* 76.4 (2009), pp. 268–277. DOI: [10.1016/j.tpb.2009.08.006](https://doi.org/10.1016/j.tpb.2009.08.006). arXiv: [0811.0203](https://arxiv.org/abs/0811.0203).
- [CST20] Neil K. Chada, Andrew M. Stuart, and Xin T. Tong. “Tikhonov Regularization within Ensemble Kalman Inversion”. In: *SIAM Journal on Numerical Analysis* 58.2 (2020), pp. 1263–1294. DOI: [10.1137/19M1242331](https://doi.org/10.1137/19M1242331).



- [CT14] Ross Cressman and Yi Tao. “The replicator equation and other game dynamics”. In: *Proceedings of the National Academy of Sciences* 111.supplement\_3 (2014), pp. 10810–10817. DOI: <https://doi.org/10.1073/pnas.1400823111>.
- [CX10] Dan Crisan and Jie Xiong. “Approximate McKean–Vlasov representations for a class of SPDEs”. en. In: *Stochastics* 82.1 (Feb. 2010), pp. 53–68. ISSN: 1744-2508, 1744-2516. DOI: [10.1080/17442500902723575](https://doi.org/10.1080/17442500902723575). (Visited on 01/04/2024).
- [Czé+22] Dániel Czégel et al. “Bayes and Darwin: How replicator populations implement Bayesian computations”. In: *BioEssays* 44.4 (2022), pp. 1–11. ISSN: 15211878. DOI: [10.1002/bies.202100255](https://doi.org/10.1002/bies.202100255).
- [DDJ06] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. “Sequential Monte Carlo samplers”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 68.3 (2006), pp. 411–436. ISSN: 13697412. DOI: [10.1111/j.1467-9868.2006.00553.x](https://doi.org/10.1111/j.1467-9868.2006.00553.x). arXiv: [0212648](https://arxiv.org/abs/0212648) [cond-mat].
- [DG05] Pierre Del Moral and Josselin Garnier. “Genealogical particle analysis of rare events”. In: *Annals of Applied Probability* 15.4 (2005), pp. 2496–2534. ISSN: 10505164. DOI: [10.1214/105051605000000566](https://doi.org/10.1214/105051605000000566).
- [DSH20] Le Duc, Kazuo Saito, and Daisuke Hotta. “Analysis and design of covariance inflation methods using inflation functions. Part 1: Theoretical framework”. In: *Quarterly Journal of the Royal Meteorological Society* 146.733 (2020), pp. 3638–3660. ISSN: 1477870X. DOI: [10.1002/qj.3864](https://doi.org/10.1002/qj.3864).
- [EHN96] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*. Vol. 375. Springer Science & Business Media, 1996.
- [FA95] Akio Fujiwara and Shun-ichi Amari. “Gradient systems in view of information geometry”. In: *Physica D: Nonlinear Phenomena* 80.3 (1995), pp. 317–327.
- [Gil+17] Marie-Eve Gil et al. “Mathematical Properties of a Class of Integro-differential Models from Population Genetics”. en. In: *SIAM Journal on Applied Mathematics* 77.4 (Jan. 2017), pp. 1536–1561. ISSN: 0036-1399, 1095-712X. DOI: [10.1137/16M1108224](https://doi.org/10.1137/16M1108224). (Visited on 01/24/2024).
- [GM98] Andrew Gelman and Xiao Li Meng. “Simulating normalizing constants: From importance sampling to bridge sampling to path sampling”. In: *Statistical Science* 13.2 (1998), pp. 163–185. ISSN: 08834237. DOI: [10.1214/ss/1028905934](https://doi.org/10.1214/ss/1028905934).
- [Har09a] Marc Harper. “Information Geometry and Evolutionary Game Theory”. In: (2009), pp. 1–13. arXiv: [0911.1383](https://arxiv.org/abs/0911.1383). URL: <http://arxiv.org/abs/0911.1383>.
- [Har09b] Marc Harper. “The replicator equation as an inference dynamic”. In: *arXiv preprint arXiv:0911.1763* (2009).
- [HKX02] Y. Hu, G. Kallianpur, and J. Xiong. “An Approximation for the Zakai Equation”. en. In: *Applied Mathematics & Optimization* 45.1 (Jan. 2002), pp. 23–44. ISSN: 0095-4616, 1432-0606. DOI: [10.1007/s00245-001-0024-8](https://doi.org/10.1007/s00245-001-0024-8). (Visited on 01/04/2024).
- [HM97] P.L. Houtekamer and Herschel L. Mitchell. “Data assimilation using an ensemble kalman filter technique”. In: *Monthly Weather Review* 126 (1997). ISSN: 00270644. DOI: [https://doi.org/10.1175/1520-0493\(1998\)126<0796:DAUAEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2).
- [Hof85] Josef Hofbauer. “The selection mutation equation”. In: *Journal of mathematical biology* 23 (1985), pp. 41–53.

- [HS98] Josef Hofbauer and Karl Sigmund. *Evolutionary games and population dynamics*. Cambridge university press, 1998.
- [Kim24] Motoo Kimura. “Diffusion Models in Population Genetics”. en. In: (2024).
- [Kim65] M Kimura. “A stochastic model concerning the maintenance of genetic variability in quantitative characters.” en. In: *Proceedings of the National Academy of Sciences* 54.3 (Sept. 1965), pp. 731–736. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.54.3.731](https://doi.org/10.1073/pnas.54.3.731). URL: <https://pnas.org/doi/full/10.1073/pnas.54.3.731> (visited on 01/24/2024).
- [KM14] Michael Kopp and Sebastian Matuszewski. “Rapid evolution of quantitative traits: theoretical perspectives”. en. In: *Evolutionary Applications* 7.1 (2014). \_eprint: <https://onlinelibrary.wiley.com/doi/abs/10.1111/eva.12127> (visited on 01/07/2024).
- [KM16] David Kelly and Ian Melbourne. “Smooth approximation of stochastic differential equations”. en. In: *The Annals of Probability* 44.1 (Jan. 2016). ISSN: 0091-1798. DOI: [10.1214/14-AOP979](https://doi.org/10.1214/14-AOP979). (Visited on 11/18/2024).
- [KNP18] Michael Kopp, Elma Nassar, and Etienne Pardoux. “Phenotypic lag and population extinction in the moving-optimum model: insights from a small-jumps limit”. In: *Journal of Mathematical Biology* 77 (2018), pp. 1431–1458.
- [Kun81] Hiroshi Kunita. “Cauchy problem for stochastic partial differential equations arising in nonlinear filtering theory”. en. In: *Systems & Control Letters* 1.1 (July 1981), pp. 37–41. ISSN: 01676911. DOI: [10.1016/S0167-6911\(81\)80010-2](https://doi.org/10.1016/S0167-6911(81)80010-2). (Visited on 11/18/2024).
- [Kun82] Hiroshi Kunita. “Stochastic partial differential equations connected with nonlinear filtering”. In: *Nonlinear Filtering and Stochastic Control: eds. Mitter, SK. nd Moro, A.* Lecture Notes in mathematics 972. Springer-Verlag, 1982, pp. 100–169.
- [Lau+14] Richard S. Laugesen et al. *Poisson’s equation in nonlinear filtering*. en. arXiv:1412.5845 [math]. Dec. 2014. URL: <http://arxiv.org/abs/1412.5845> (visited on 01/04/2024).
- [Lee20] Peter Jan van Leeuwen. “A consistent interpretation of the stochastic version of the Ensemble Kalman Filter”. In: *Quarterly Journal of the Royal Meteorological Society* 146.731 (2020), pp. 2815–2825. ISSN: 1477870X. DOI: [10.1002/qj.3819](https://doi.org/10.1002/qj.3819).
- [LLN19] Yulong Lu, Jianfeng Lu, and James Nolen. *Accelerating Langevin Sampling with Birth-death*. en. arXiv:1905.09863 [stat]. May 2019. URL: <http://arxiv.org/abs/1905.09863> (visited on 11/18/2024).
- [LSW23] Yulong Lu, Dejan Slepčev, and Lihan Wang. “Birth–death dynamics for sampling: global convergence, approximations and their asymptotics”. en. In: *Nonlinearity* 36.11 (Nov. 2023), pp. 5731–5772. ISSN: 0951-7715, 1361-6544. DOI: [10.1088/1361-6544/acf988](https://doi.org/10.1088/1361-6544/acf988). (Visited on 11/18/2024).
- [MD14] Piere Del Moral and Arnaud Doucet. “Particle methods: An introduction with applications”. In: *ESAIM: Proceedings* 44. January (2014), pp. 1–46. DOI: [10.1051/proc/201444001](https://doi.org/10.1051/proc/201444001).
- [MH00] H.L. Mitchell and P.L. Houtekamer. “An Adaptive Ensemble Kalman Filter”. In: *Monthly Weather Review* 128.1994 (2000), pp. 416–433. DOI: [https://doi.org/10.1175/1520-0493\(2000\)128<0416:AAEF>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<0416:AAEF>2.0.CO;2).
- [MHK14] Sebastian Matuszewski, Joachim Hermisson, and Michael Kopp. “Fisher’s geometric model with a moving optimum”. In: *Evolution* 68.9 (2014), pp. 2571–2588. ISSN: 15585646. DOI: [10.1111/evo.12465](https://doi.org/10.1111/evo.12465).

- [MM24] Aimee Maurais and Youssef Marzouk. *Sampling in Unit Time with Kernel Fisher-Rao Flow*. en. arXiv:2401.03892 [stat]. June 2024. DOI: [10.48550/arXiv.2401.03892](https://doi.org/10.48550/arXiv.2401.03892). URL: <http://arxiv.org/abs/2401.03892> (visited on 11/24/2024).
- [Mor04] Pierre del Moral. *Feynman-Kac formulae: Genealogical and Interacting Particle Systems with Applications*. Probabilit. Springer, 2004. DOI: <https://doi.org/10.1007/978-1-4684-9393-1>.
- [Mor97] Pierre del Moral. “Nonlinear filtering: Interacting particle resolution”. In: *Comptes Rendus de l’Académie des Sciences - Series I - Mathematics* 325.1 (1997), pp. 653–658. DOI: [https://doi.org/10.1016/S0764-4442\(97\)84778-7](https://doi.org/10.1016/S0764-4442(97)84778-7).
- [MS18] Panayotis Mertikopoulos and William H. Sandholm. “Riemannian game dynamics”. In: *Journal of Economic Theory* 177.1991 (2018), pp. 315–364. ISSN: 10957235. DOI: [10.1016/j.jet.2018.06.002](https://doi.org/10.1016/j.jet.2018.06.002). arXiv: [1603.09173](https://arxiv.org/abs/1603.09173).
- [Nea01] Radford M. Neal. “Annealed importance sampling”. In: *Statistics and Computing* 11.2 (2001), pp. 125–139. ISSN: 09603174. DOI: [10.1023/A:1008923215028](https://doi.org/10.1023/A:1008923215028). arXiv: [9803008](https://arxiv.org/abs/9803008) [physics].
- [OR01] Jörg Oechssler and Frank Riedel. “Evolutionary dynamics on infinite strategy spaces”. In: *Economic theory* 17 (2001), pp. 141–162. DOI: [10.1007/PL00004092](https://doi.org/10.1007/PL00004092).
- [Pat24] Sahani Pathiraja. “ $L^2$  convergence of smooth approximations of stochastic differential equations with unbounded coefficients”. en. In: *Stochastic Analysis and Applications* 42.2 (Mar. 2024), pp. 354–369. ISSN: 0736-2994, 1532-9356. DOI: [10.1080/07362994.2023.2260863](https://doi.org/10.1080/07362994.2023.2260863). (Visited on 09/25/2024).
- [PR23] Jakiw Pidstrigach and Sebastian Reich. “Affine-invariant ensemble transform methods for logistic regression”. In: *Foundations of Computational Mathematics* 23.2 (2023), pp. 675–708. DOI: [10.1007/s10208-022-09550-2](https://doi.org/10.1007/s10208-022-09550-2).
- [PRS21] Sahani Pathiraja, Sebastian Reich, and Wilhelm Stannat. “McKean-Vlasov SDEs in nonlinear filtering”. en. In: *SIAM Journal on Control and Optimization* 59.6 (Jan. 2021). arXiv:2007.12658 [cs, math], pp. 4188–4215. ISSN: 0363-0129, 1095-7138. DOI: [10.1137/20m1355197](https://doi.org/10.1137/20m1355197). (Visited on 01/04/2024).
- [Rei11] Sebastian Reich. “A dynamical systems framework for intermittent data assimilation”. In: *BIT Numerical Mathematics* 51 (2011), pp. 235–249. DOI: [10.1007/s10543-010-0302-4](https://doi.org/10.1007/s10543-010-0302-4).
- [Sha09] Cosma Rohilla Shalizi. “Dynamics of bayesian updating with dependent data and misspecified models”. In: *Electronic Journal of Statistics* 3 (2009), pp. 1039–1074. ISSN: 19357524. DOI: [10.1214/09-EJS485](https://doi.org/10.1214/09-EJS485). arXiv: [0901.1342](https://arxiv.org/abs/0901.1342).
- [SS83] Peter Schuster and Karl Sigmund. “Replicator dynamics”. In: *Journal of theoretical biology* 100.3 (1983), pp. 533–538. DOI: [https://doi.org/10.1016/0022-5193\(83\)90445-9](https://doi.org/10.1016/0022-5193(83)90445-9).
- [Stu10] Andrew M Stuart. “Inverse problems: a Bayesian perspective”. In: *Acta numerica* 19 (2010), pp. 451–559. DOI: [10.1017/S0962492910000061](https://doi.org/10.1017/S0962492910000061).
- [Tag+17] Amirhossein Taghvaei et al. “Kalman Filter and its Modern Extensions for the Continuous-time Nonlinear Filtering Problem”. In: *Journal of Dynamic Systems, Measurement and Control* 140.3 (2017). DOI: [10.1115/1.4037780](https://doi.org/10.1115/1.4037780).

- [TLK96] Lev S. Tsimring, Herbert Levine, and David A. Kessler. “RNA virus evolution via a fitness-space model”. In: *Physical Review Letters* 76.23 (1996), pp. 4440–4443. ISSN: 10797114. DOI: [10.1103/PhysRevLett.76.4440](https://doi.org/10.1103/PhysRevLett.76.4440).
- [TMK16] Xin T. Tong, Andrew J. Majda, and David Kelly. “Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation”. en. In: *Communications in Mathematical Sciences* 14.5 (2016), pp. 1283–1313. DOI: [10.4310/CMS.2016.v14.n5.a5](https://doi.org/10.4310/CMS.2016.v14.n5.a5). (Visited on 01/04/2024).
- [Vil21] Cédric Villani. *Topics in optimal transportation*. Vol. 58. American Mathematical Soc., 2021.
- [Vla20] Andrew Vlasic. *Markovian Switching of Mutation Rates in Evolutionary Network Dynamics*. en. arXiv:2005.00900 [math, q-bio]. May 2020. URL: <http://arxiv.org/abs/2005.00900> (visited on 01/24/2024).
- [WN24] L. Wang and N. Nüsken. *Measure transport with kernel mean embeddings*. en. arXiv:2401.12967 [math]. Sept. 2024. DOI: [10.48550/arXiv.2401.12967](https://doi.org/10.48550/arXiv.2401.12967). (Visited on 11/24/2024).
- [WRS18] Jana de Wiljes, Sebastian Reich, and Wilhelm Stannat. “Long-Time Stability and Accuracy of the Ensemble Kalman–Bucy Filter for Fully Observed Processes and Small Measurement Noise”. en. In: *SIAM Journal on Applied Dynamical Systems* 17.2 (Jan. 2018), pp. 1152–1181. ISSN: 1536-0040. DOI: [10.1137/17M1119056](https://doi.org/10.1137/17M1119056). (Visited on 10/01/2024).
- [Yan+12] Tao Yang et al. “Multivariable Feedback Particle Filter”. en. In: *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. arXiv:1303.1205 [math]. Dec. 2012, pp. 4063–4070. DOI: [10.1109/CDC.2012.6425937](https://doi.org/10.1109/CDC.2012.6425937). URL: <http://arxiv.org/abs/1303.1205> (visited on 01/04/2024).
- [Zha17] Chi Zhang. “Controlled particle systems for nonlinear filtering and global optimization”. PhD thesis. University of Illinois at Urbana-Champaign, 2017.