Abstract

For a class of partially observed diffusions, sufficient conditions are given for the map from the
initial condition of the signal to filtering distribution to be contractive with respect to Wasserstein
distances, with rate which has no dependence on the dimension of the state-space and is stable
under tensor products of the model. The main assumptions are that the signal has affine drift
and constant diffusion coefficient, and that the likelihood functions are log-concave. Contraction
estimates are obtained from an $h$-process representation of the transition probabilities of the signal
rewighted so as to condition on the observations.

1 Introduction

Let $(\theta_t)_{t \in \mathbb{R}_+}$, called the signal process, be the solution of the stochastic differential equation:
\begin{equation}
d\theta_t = (\alpha + \beta \theta_t)dt + \sigma dB_t, \tag{1.1}
\end{equation}
where $\alpha \in \mathbb{R}^p$ and $\beta$ is a $p \times p$ matrix of reals, $\sigma \geq 0$ is a scalar, and $(B_t)_{t \in \mathbb{R}_+}$ is $p$-dimensional
Brownian motion. Let observations $(Y_k)_{k \in \mathbb{N}_0}$ be each valued in a measurable space $(\mathcal{Y}, \mathcal{Y})$, conditionally
independent given $(\theta_t)_{t \in \mathbb{R}_+}$ and such that the conditional probability that $Y_k$ lies in $A \in \mathcal{Y}$ given $(\theta_t)_{t \in \mathbb{R}_+}$ is of the form $\int_A g_k(\theta_{k\Delta} \mid \chi (dy))$, for a measure $\chi$ on $\mathcal{Y}$, a function $g_k : \mathbb{R}^p \times \mathcal{Y} \to (0, \infty)$ and
a constant $\Delta > 0$.

The filtering distributions $(\pi_k^\mu)_{k \in \mathbb{N}_0}$ associated with a fixed sequence $(y_k)_{k \in \mathbb{N}_0}$ and a probability
measure $\mu$ on the Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^p)$ are defined by
\begin{equation}
\pi_k^\mu(A) := \frac{E_{\mu} \left[ 1_A(\theta_{k\Delta}) \prod_{j=0}^{k} g_j(\theta_{j\Delta}, y_j) \right]}{E_{\mu} \left[ \prod_{j=0}^{k} g_j(\theta_{j\Delta}, y_j) \right]}, \quad A \in \mathcal{B}(\mathbb{R}^p), \tag{1.2}
\end{equation}
where $E_{\mu}$ denotes expectation with respect to the law of the solution of (1.1) with $\theta_0 \sim \mu$. When
$(y_0, \ldots, y_k)$ are replaced in (1.2) by the random variables $(Y_0, \ldots, Y_k)$ distributed according to the
above prescription and with true initialization also $\theta_0 \sim \mu$, then $\pi_k^\mu$ is a version of the conditional
distribution of $\theta_{k\Delta}$ given $(Y_0, \ldots, Y_k)$. It shall be assumed throughout that whichever $(y_k)_{k \in \mathbb{N}_0}$ and $\mu$
we consider, the denominator in (1.2) is finite for each $k$, so that $(\pi_k^\mu)_{k \in \mathbb{N}_0}$ are well defined as probability
measures. When $\mu$ is $\delta_\theta$, the Dirac mass located at $\theta$, we shall write and $E_\theta$ and $\pi_k^\theta$ instead of $E_{\delta_\theta}$ and
$\pi_k^{\delta_\theta}$.

Our overall aim is to obtain bounds on Wasserstein distance between differently initialized filtering
distributions, say $\pi^\mu_k, \pi^\nu_k$ in terms of distance between $\mu$ and $\nu$, and find conditions under which the
former distance decays as $k \to \infty$ at a rate which does not depend on the dimension of the state-space
$\mathbb{R}^p$. The question of under what conditions the filtering distributions forget their initial condition
has been approached using a variety of techniques, see [2, Chap. 4] for an overview. The topic
of dependence on dimension has received attention only quite recently, motivated by the increasing
importance of inference problems involving high-dimensional stochastic processes.
Recent contributions such as [6, 3, 4] study the rate of forgetting in total variation distance and V-norm, and the rate estimates obtained there depend on the constants associated with minorization-type conditions for the signal process. However such constants, and therefore the rate estimates based upon them, typically degrade with the dimension of the state-space. Infinite-dimensional filtering is treated in [12], where stability results are obtained involving weak convergence and the notion of local ergodicity, which pertains to the mixing properties of finite-dimensional components of the infinite dimensional signal process, conditional on the observations. The results hold under mild conditions and do not quantify the rate of convergence. For signals with certain spatio-temporal mixing properties, [10] provides local, quantitative filter stability results which do not degrade with dimension as part of their particle filter analysis.

The approach taken here does not rely on spatial structure of the model, but is instead connected with contraction properties of gradient flows and convexity, and influenced by analyses of Markov processes using abstract ideas of curvature and underlying links to functional inequalities [1]. The proofs ultimately rely on a quite simple coupling technique and the pathwise stability properties of diffusions whose drifts involve the gradients of certain convex potentials. This convexity arises from a combination of two features of the model we consider: firstly log-concavity of the likelihood functions \( \theta \to g_k(\theta, y_k) \), which will be one of our main assumptions (stated precisely below), and secondly a log-concavity-preservation characteristic of the signal model (1.1).

Log-concave likelihoods appear, for example, in statistical regression models built around the exponential family of distributions, in particular in Generalized Linear Models [8], which are used to solve high-dimensional data analysis problems in disciplines such as neuroscience, genomics and internet traffic prediction.

In this setting \( y_k = (y^1_k, \ldots, y^n_k) \in \mathbb{R}^n =: Y \), and with known covariates \( x_k = (x^i_k) \), \( i = 1, \ldots, n \), \( j = 1, \ldots, p \), \( g_k(\theta, y_k) \) is of the form:

\[
g_k(\theta, y_k) = \exp \left\{ \sum_{i=1}^n \left\{ \sum_{j=1}^p y^i_k x^i_j \theta^j - \psi \left( \sum_{j=1}^p x^i_j \theta^j \right) + \log f(y^i_k) \right\} \right\},
\]

where \( \theta = (\theta^1, \ldots, \theta^p) \), \( \phi \) is a given function, and \( \psi \) is convex, the latter implying \( \theta \to g_k(\theta, y_k) \) is log-concave. The situation in which the regression parameter \( \theta \) is treated as time-varying is known as a Dynamic Generalized Linear Model [5]. Linear-Gaussian vector auto-regressions for \( (\theta_k)_{k \in \mathbb{N}_0} \) are a popular choice in practice and indeed the solution of (1.1) satisfies

\[
\theta_{(k+1)\Delta} = a + B \theta_{k\Delta} + \sigma \xi_{k+1},
\]

where \( \xi_{k+1} = e^{\Delta \beta} \int_{k \Delta}^{(k+1) \Delta} e^{-(t-k \Delta) \beta} dB_t \) is a Gaussian random variable and \( a = e^{\Delta \beta} \int_0^\Delta e^{-t \beta} \alpha dt \), \( B = e^{\Delta \beta} \).

The signal model (1.1) also has an important analytical property: it is known that the semigroup of transition operators \( (P_t)_{t \in \mathbb{R}_+} \) associated with (1.1) preserves log-concavity, meaning that for any log-concave function \( f \) and \( t > 0 \), \( P_t f \) is log-concave, see for example [7]. Combined with log-concavity of \( \theta \to g_k(\theta, y_k) \), the Markov property of \( (\theta_t)_{t \in \mathbb{R}_+} \) and the fact that a pointwise product of log-concave functions is log-concave, this implies that \( \theta \to \varphi_{j,k}(\theta) := E \theta \left[ \prod_{i=j}^k g_i(\theta_{(i-j)\Delta}, y_i) \right] \) is log-concave. Functions of the latter form play an important role in filter stability because they provide the re-weighting of transition probabilities which corresponds to conditioning on observations, and this is where the convex potentials alluded to earlier arise.

It is important to note that log-concavity of \( \varphi_{j,k} \) cannot be expected in much greater generality. It was established in [7] that among all diffusions of the form:

\[
d\theta_t = b(\theta_t) dt + \sigma(\theta_t) dB_t,
\]

with \( b(\cdot) \), \( \sigma(\cdot) \) satisfying some mild regularity conditions, it is only in the case that \( b(\cdot) \) is affine and \( \sigma(\cdot) \) is a constant that \( P_t f \) is log-concave for all log-concave \( f \). This motivates our focus on signal processes of the form (1.1).
Notation and conventions

A function \( f : \mathbb{R}^p \to (0, \infty) \) is called log-concave if
\[
\log f(cu + (1-c)v) \geq c \log f(u) + (1-c) \log f(v), \quad \forall \ u, v \in \mathbb{R}^p, \ c \in [0,1],
\]
and strongly log-concave if there exists a log-concave function \( \tilde{f} \) and a constant \( \lambda_f \in (0, \infty) \) such that \( f(u) = \exp(-\frac{\lambda_f}{2} u^T \tilde{f}(u)) \). For a measure \( \mu \), function \( f \) and integral kernel \( K \), we shall write \( \mu f = \int f(u) \mu(du) \), \( \mu K(\cdot) = \int \mu(du)K(u, \cdot) \), \( Kf(u) = \int f(v)K(u, dv) \). For a nonnegative function \( f \), \( \mu \cdot f \) denotes the measure \( \mu(du)f(u) \). The gradient and Laplace operators with respect to \( \theta \) are denoted \( \nabla_\theta \) and \( \nabla^2_\theta \). The indicator function on a set \( A \) is denoted \( 1_A \). The class of real-valued and twice continuously differentiable functions with on \( \mathbb{R}^p \) is denoted \( C^2 \).

The order-\( q \) Wasserstein distance between probability measures on \( \mathcal{B}(\mathbb{R}^p) \) is:
\[
W_q(\mu, \nu) := \left( \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{\mathbb{R}^p \times \mathbb{R}^p} \|u-v\|^q \gamma(du, dv) \right)^{1/q},
\]
where \( \Gamma(\mu, \nu) \) is the set of all couplings of \( \mu \) and \( \nu \), and \( \| \cdot \| \) is the Euclidean norm.

2 Wasserstein distance between filtering distributions initialized at points

2.1 Main result

**Assumption 1.** For each \( k \in \mathbb{N}_0 \), \( \theta \mapsto g_k(\theta, y_k) \) is strictly positive, a member of \( C^2 \), and there exists a constant \( \lambda_g(k) \in [0, \infty) \) and a log-concave function \( \tilde{g}_k : \mathbb{R}^p \to (0, \infty) \) such that \( g_k(\theta, y_k) = \exp \left[ -\frac{\lambda_g(k)}{2} \theta^T \theta \right] \tilde{g}_k(\theta) \).

**Theorem 1.** If Assumptions 1 holds, then for any \( q \geq 1 \),
\[
W_q(\pi_k^\theta, \pi_k^\vartheta) \leq \exp \left[ -\sum_{j=1}^{k} \int_0^\Delta \lambda(j, t)dt \right] \|\theta - \vartheta\|, \quad \forall k \geq 1, \ \theta, \vartheta \in \mathbb{R}^p,
\]
where
\[
\lambda(j, t) := \lambda_{\text{sig}} + \frac{\sigma^2 \lambda_g(j) \lambda_{\text{min}}^\beta(\Delta - t)}{1 + \sigma^2 \lambda_g(j) \int_0^\Delta \lambda_{\text{max}}^\beta(\Delta - s)ds},
\]
\( \lambda_{\text{sig}} \in \mathbb{R} \) is the smallest eigenvalue of \(-((\beta + \beta^T)/2 \) and \( \lambda_{\text{min}}^\beta(t), \lambda_{\text{max}}^\beta(t) \in (0, \infty) \) are respectively the smallest and largest eigenvalues of \( e^{\beta t}(e^{\beta t})^T \).

2.1.1 Some specific cases

If Assumption 1 is satisfied with \( \lambda_g(k) = 0 \) for all \( k \), so that \( \theta \mapsto g_k(\theta, y_k) \) is log-concave, but not necessarily strongly log-concave, then (2.1) becomes:
\[
W_q(\pi_k^\theta, \pi_k^\vartheta) \leq \exp \left[ -k \Delta \lambda_{\text{sig}} \right] \|\theta - \vartheta\|.
\]
Note that the right hand side of this bound has no dependence on the observations \( (y_k)_{k \in \mathbb{N}_0} \). Since \( \lambda_g(k) = 0 \) allows \( \theta \mapsto g_k(\theta, y_k) \) to be a constant, in which case \( \pi_k^\theta(\cdot) = P_k \Delta(\theta, \cdot) \), Theorem 1 implies that \( \lambda_{\text{sig}} \), if it is positive, is the exponential rate of Wasserstein contraction of \( (P_t \Delta)_{t \in \mathbb{R}_+} \). In summary, assuming \( \theta \mapsto g_k(\theta, y_k) \) is log-concave and \( \lambda_{\text{sig}} > 0 \), the exponential rate of Wasserstein contraction of the filters \( (\pi_k^\theta)_{k \in \mathbb{N}_0} \) is positive and at least that of the \( (P_k \Delta)_{k \in \mathbb{N}_0} \).

3
As soon as $\sigma^2 > 0$, the observations can help achieve contraction of the filters without contraction of $(P_t)_{t \in \mathbb{R}_+}$. For example, with $\beta = -\lambda_{\text{sig}}I$ for any $\lambda_{\text{sig}} \in \mathbb{R}$, we have $\lambda_{\beta}^{\text{min}}(t) = \lambda_{\beta}^{\text{max}}(t) = e^{-2\lambda_{\text{sig}}}$, and it is straightforward to check that (2.1) becomes:

$$W_q(\pi_k^0, \pi_k^0) \leq \prod_{j=1}^{k} \frac{\exp(-\lambda_{\text{sig}} \Delta)}{1 + \sigma^2 \lambda_{\beta}(j) \int_0^\Delta e^{-2\lambda_{\text{sig}} t} dt} \|\theta - \tilde{\theta}\|,$$

so that if $\lambda_{\text{sig}} \leq 0$ and $\Delta$ is fixed, contraction can be achieved if the products $\sigma^2 \lambda_{\beta}(j), j \in \mathbb{N}$, are sufficiently large. A notable case is when $\lambda_{\text{sig}} = 0$,

$$W_q(\pi_k^0, \pi_k^0) \leq \left( \prod_{j=1}^{k} \frac{1}{1 + \sigma^2 \lambda_{\beta}(j) \Delta} \right) \|\theta - \tilde{\theta}\|.$$

### 2.1.2 Dimension-free nature of the contraction rate

The quantities $(\lambda_{\beta}(k))_{k \in \mathbb{N}_0}$, $\lambda_{\text{sig}}$, $\lambda_{\beta}^{\text{min}}(t)$, $\lambda_{\beta}^{\text{max}}(t)$ and $\sigma^2$ appearing in (2.1) do not necessarily have any dependence on the dimension of the state space, $\mathbb{R}^p$, and are stable under tensor products of the model described in section 1, in the sense that

1) $g_k(\theta, y_{k,i}) = \exp\left[-\frac{\lambda_{\beta}(k)}{2}(\theta^T \theta)\right] \tilde{g}_{k,i}(\theta), \quad i = 1, 2,$

$$\implies g_k(\theta, y_{k,1})g_k(\tilde{\theta}, y_{k,2}) = \exp\left[-\frac{\lambda_{\beta}(k)}{2}(\theta^T \theta + \tilde{\theta}^T \tilde{\theta})\right] \tilde{g}_{k,1}(\theta)\tilde{g}_{k,2}(\tilde{\theta}),$$

2) spectrum$\{(\beta + \beta^T)/2\} = $spectrum$\{(\beta^{\otimes 2} + (\beta^{\otimes 2})^T)/2\}$,

3) spectrum$\{e^{\beta t}(e^{\beta^T t})^T\} = $spectrum$\{e^{\beta^{\otimes 2} t}(e^{\beta^{\otimes 2}^T t})\}$,

where $\beta^{\otimes 2}$ denotes the Kronecker product $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \beta$. This amounts to saying that if one expands the model to state-space $\mathbb{R}^{2p}$ by defining the signal to be two independent copies of (1.1), with independent observations $y_k = (y_{k,1}, y_{k,2}) \in \mathbb{V}^2$ whose likelihood functions have common log-concavity parameter $\lambda_{\beta}(k)$, then there is no degradation of $\lambda(j, t)$ in (2.1).

### 2.1.3 Improvement and generalization

Under the assumptions of Theorem 1 the bound in (2.1) cannot be improved in general; for example if $\sigma = 0$ and $\beta = -\lambda_{\text{sig}} I$, then $W_q(\pi_k^0, \pi_k^0) = \exp(-k \Delta \lambda_{\text{sig}})\|\theta - \tilde{\theta}\|$. The case of initial distributions which are not necessarily Dirac measures is addressed in section 3.

### 2.2 Proof of Theorem 1

Our starting point is the well-known fact that the filtering distributions can be written in terms of the transition probabilities of the signal process re-weighted so as to condition on observations. Fix $k \geq 0$ and define

$$\varphi_{k,k}(\theta) := g_k(\theta, y_k),$$

$$\varphi_{j,k}(\theta) := g_j(\theta, y_j)P_{\Delta} \varphi_{j+1,k}(\theta) \quad 0 \leq j < k,$$

$$R_{j,k}(\theta, A) := \frac{\int_A P_{\Delta}(\theta, \tilde{\theta}) \varphi_{j,k}(\tilde{\theta}) d\tilde{\theta}}{P_{\Delta} \varphi_{j,k}(\theta)} \quad 1 \leq j \leq k,$$

We will need the following preliminary lemma.
Lemma 1. For every log-concave $f$ and $t > 0$, $P_t f$ is log-concave.

Proof. [7, proof of Prop. 1.3]

Lemma 2. We have

$$\pi_k^\theta(A) = R_{1,k}R_{2,k} \cdots R_{k,k}(\theta, A).$$

(2.5)

If Assumption 1 holds, then for each $j, k$ such that $0 \leq j \leq k$, there exists a log-concave function $\varphi_{j,k}$ such that:

$$\varphi_{j,k}(\theta) = \exp \left[ -\frac{\lambda_0(j)}{2} \theta^T \tilde{\varphi}_{j,k}(\theta) \right].$$

(2.6)

Proof. The expression for $\pi_k^\theta(A)$ follows from (1.2) and the Markov property of the signal. The second claim is established by repeated application to (2.3)–(2.4) of Lemma 1 and the fact that the pointwise product of log-concave functions is log-concave.

The Wasserstein bound in Theorem 1 is a consequence of contraction estimates for the kernels $R_{j,k}$ derived in sections 2.3 and 2.4. In particular, (2.1) is obtained by combining Proposition 1, which is based on a synchronous coupling of an $h$-process interpretation of $R_{j,k}$ where $h$ is a certain space-time harmonic function, with Proposition 2, which quantifies the log-concavity of $h$ inherited from that of $\varphi_{j,k}$ in (2.6).

2.3 A space-time $h$-transform of the signal process

Let $C([0, \Delta], \mathbb{R}^p \times [0, \Delta])$ be the space of $\mathbb{R}^p \times [0, \Delta]$-valued, continuous functions on $[0, \Delta]$ endowed with the supremum norm. Let $(\theta, t)_{t \in [0, \Delta]}$ be the associated space-time coordinate process and let $\mathcal{F} = (\mathcal{F}_t)_{t \in [0, \Delta]}$ be the filtration it generates. The extended generator (in the sense of [11, p. 285]) of the space-time process on $C([0, \Delta], \mathbb{R}^p \times [0, \Delta])$ under the law associated with (1.1) and acting on functions $f$ on $\mathbb{R}^p \times \mathbb{R}_+$ is:

$$L f(\theta, t) := \frac{\partial}{\partial t} f(\theta, t) + (\alpha + \beta \theta)^T \nabla_\theta f(\theta, t) + \frac{\sigma^2}{2} \nabla^2 \theta f(\theta, t).$$

Lemma 3. Let Assumption 1 hold, fix any $j, k$ such that $1 \leq j \leq k$ and define

$$h(\theta, t) := P_{\Delta-t}\varphi_{j,k}(\theta).$$

(2.7)

There exists a probability kernel $\mathbf{P}^h : \mathbb{R}^p \times \mathcal{F}_\Delta \to [0, 1]$ such that for any $\theta_0 \in \mathbb{R}^p$ and $A \in \mathcal{B}(\mathbb{R}^p)$, $R_{j,k}(\theta_0, A) = \mathbf{P}^h(\theta_0, \{\theta_\Delta \in A\})$, and under $\mathbf{P}^h(\theta_0, \cdot)$ the extended generator of the space-time process $(\theta, t)_{t \in [0, \Delta]}$ on $C([0, \Delta], \mathbb{R}^p \times [0, \Delta])$ is:

$$L^h f(\theta, t) := L f(\theta, t) + \sigma^2 \nabla \theta \log h(\theta, t)^T \nabla \theta f(\theta, t).$$

(2.8)

Proof. Let $\mathbf{P} : \mathbb{R}^p \times \mathcal{F}_\Delta \to [0, 1]$ be a probability kernel such that $\mathbf{P}(\theta_0, \cdot)$ is the law of the space-time process associated with (1.1) on the time horizon $[0, \Delta]$ initialized from the point $(\theta_0, 0)$.

Note the following properties of the functions $\varphi_{j,k}$. Under Assumption 1, for all $k \geq 0$, $\theta \mapsto g_k(\theta, y_k)$ is strictly positive and therefore so is $\varphi_{j,k}$ for all $j \leq k$. Also, it follows from the assumption that for all $k \geq 0$, $\theta \mapsto g_k(\theta, y_k)$ is a member of $C^2$, combined with (2.3)–(2.4) and (1.3) that $\varphi_{j,k} \in C^2$. By the log-concavity established in Lemma 2, there exists a constant $c$ such that $\varphi_{j,k}(\theta)$ grows no faster than $c^{||\theta||}$ as $||\theta|| \to \infty$.

Now fix $j, k$ as in the statement. Then $\theta \mapsto h(\theta, t)$ is strictly positive, log-concave by Lemma 2 and Lemma 1, and a member of $C^2$ because of (2.7) and $\varphi_{j,k} \in C^2$. With:

$$D_t := \frac{h(\theta_t, t)}{h(\theta_0, 0)}.$$
\( (D_t)_{t \in [0, \Delta]} \) is a \((\mathcal{F}_t, \mathbb{P}(\theta_0, \cdot))\)-continuous martingale, and the expected value of \( D_t \) under \( \mathbb{P}(\theta_0, \cdot) \) is 1. Now define the probability kernel \( \mathbb{P}^h(\theta, \cdot) := D_\Delta \cdot \mathbb{P}(\theta, \cdot) \). Under \( \mathbb{P}^h(\theta_0, \cdot) \), \((\mathbb{P}_t)_{t \in [0, \Delta]} \) is an inhomogeneous Markov process with transition probabilities:

\[
P^h_{s,t}(\theta, d\vartheta) := \frac{P_{t-s}(\theta, d\vartheta)h(\vartheta, t)}{h(\theta, s)},
\]

and \( R_{j,k}(\theta, A) = P^h_{0,\Delta}(\theta, A) = \mathbb{P}^h(\theta, \{\theta \in A\}) \). By \([11, \text{Prop. 3.9, p.357}]\), the extended generator of the space-time process under \( \mathbb{P}^h(\theta_0, \cdot) \) is \( L^h f = h^{-1}L(h f) \), which is equal to the r.h.s. of \((2.8)\) because \( \int P_{t}(\theta, d\vartheta)h(\vartheta, t + s) = h(\theta, t) \) and hence \( L(h) = 0 \).

**Proposition 1.** Fix any \( j,k \) such that \( 1 \leq j \leq k \). If there exists a continuous function \( \lambda_h : [0, \Delta] \to [0, \infty) \) and a function \( \bar{h} : \mathbb{R}^p \times [0, \Delta] \to (0, \infty) \) such that for each \( t, \theta \mapsto \bar{h}(\theta, t) \) is log-concave and \( h \) as in Lemma 3 satisfies \( \bar{h}(\theta, t) = \exp \left[ -\frac{\lambda_h(t)(\theta^T \theta)}{2} \right] h(\theta, t) \), then for any \( q \geq 1 \),

\[
W_q(R_{j,k}(\theta, \cdot), R_{j,k}(\theta, \cdot)) \leq \exp \left[ -\lambda_{\text{sig}} \Delta - \sigma^2 \int_0^\Delta \lambda_h(t) \, dt \right] \| \theta - \vartheta \|.
\]

**Proof.** Consider the synchronous coupling

\[
\theta_t = \theta_0 + \int_0^t \alpha + \beta \theta_s + \sigma^2 \nabla_\theta \log h(\theta_s, s) \, ds + \sigma B_t, \\
\vartheta_t = \vartheta_0 + \int_0^t \alpha + \beta \vartheta_s + \sigma^2 \nabla_\theta \log h(\vartheta_s, s) \, ds + \sigma B_t.
\]

By Ito’s formula, for any continuous function \( \zeta : [0, \Delta] \to \mathbb{R} \),

\[
\| \theta_t - \vartheta_t \|^2 e^{2 \int_0^t \zeta(s) \, ds} = \| \theta_0 - \vartheta_0 \|^2 + 2 \int_0^t \zeta(s) \| \theta_s - \vartheta_s \|^2 + (\theta_s - \vartheta_s)^T \beta(\theta_s - \vartheta_s) \, e^{2 \int_0^s \zeta(u) \, du} \, ds \\
+ 2 \int_0^t \sigma^2 (\nabla_\theta \log h(\theta_s, s) - \nabla_\theta \log h(\vartheta_s, s))^T (\theta_s - \vartheta_s) \, e^{2 \int_0^s \zeta(u) \, du} \, ds \tag{2.9}
\]

Now set \( \zeta(s) = \lambda_{\text{sig}} + \lambda_h(s) \). For any skew-symmetric matrix, say \( A \), and any \( u \in \mathbb{R}^p, \ u^T Au = (Au)^T u = u^T A^T u = -u^T Au \), hence \( u^T A u = 0 \), so

\[
u^T \beta u = \frac{1}{2} u^T (\beta + \beta^T) u \leq -\lambda_{\text{sig}} \| u \|^2, \quad \forall u \in \mathbb{R}^p. \tag{2.10}
\]

The assumption on \( h \) implies

\[
(\nabla_\theta \log h(\theta, s) - \nabla_\theta \log h(\vartheta, s))^T (\theta - \vartheta) \leq -\lambda_h(s) \| \theta - \vartheta \|^2, \quad \theta, \vartheta \in \mathbb{R}^p. \tag{2.11}
\]

Applying \((2.10)\) and \((2.11)\) to \((2.9)\) gives:

\[
\| \theta_\Delta - \vartheta_\Delta \| \leq \exp \left( -\int_0^\Delta \lambda_{\text{sig}} + \lambda_h(t) \, dt \right) \| \theta_0 - \vartheta_0 \|.
\]

The proof is completed by taking expectations and applying Lemma 3. \(\square\)
2.4 Quantifying log-concavity of $\theta \mapsto h(\theta, t)$

The main result of this section is Proposition 2, which complements Lemma 1 by quantifying the influence on the log-concavity of $\theta \mapsto h(\theta, t)$ of the parameters of the signal process and the log-concavity of the likelihood functions, and provides verification of the hypotheses of Proposition 1.

**Proposition 2.** Let Assumption 1 hold, fix $j, k$ such that $1 \leq j \leq k$ and let $h$ be as in Lemma 3. Then there exists a function $h : \mathbb{R}^p \times [0, \Delta] \to (0, \infty)$ such that $\theta \mapsto h(\theta, t)$ is log-concave and

$$h(\theta, t) = \exp \left[ -\frac{\lambda_h(t)}{2} \theta^T \theta \right] \tilde{h}(\theta, t),$$

where

$$\lambda_h(t) := \frac{\lambda_\beta(j) \lambda_\beta^{\min}(\Delta - t)}{1 + \sigma^2 \lambda_\beta(j) \int_0^t \lambda_\beta^{\max}(\Delta - s) ds},$$

and $\lambda_\beta^{\min}(t), \lambda_\beta^{\max}(t)$ are respectively the smallest and largest eigenvalues of $e^{\beta t}(e^{\beta t})^T$.

We shall make use of the following well-known lemma [9, Thm. 6].

**Lemma 4.** For every function $(u, v) \mapsto f(u, v)$ on $\mathbb{R}^p \times \mathbb{R}^q$ which is log-concave in $(u, v)$, the integral $\int f(u, v) dv$ is a log-concave function of $u$.

Lemma 5 and Lemma 6 are technical results used in the proof of Proposition 2.

**Lemma 5.** Let $F, S$ be real, square, symmetric matrices such that $F + S$ is invertible. Then

$$v^TFv + (u-v)^T S (u-v) = u^TCu + z^T (F + S) z$$

where $C := (F + S)^{-1}$ and $z := v - (F + S)^{-1} Su$.

**Proof.** We have using the assumed symmetry of $F$ and $S$,

$$z^T (F+S)z = v^T (F+S) v - 2u^T Sv + u^T S (F+S)^{-1} Su.$$ 

Therefore

$$u^T Cu + z^T (F+S) z = u^T Su + v^T (F+S) v - 2u^T Sv = v^T Fv + (u-v)^T S(u-v).$$

**Lemma 6.** Let $f$ be any function of the form $f(u) : u \in \mathbb{R}^p \mapsto \exp(-\frac{1}{2} u^T F u) \tilde{f}(u)$ where $F$ is a real symmetric matrix and $\tilde{f}$ is log-concave, and let $S$ be a real symmetric matrix such that $F + S$ is invertible. Then for any $a \in \mathbb{R}^p$ and $p \times p$ real matrix $B$,

$$f(v) \exp \left[ \frac{1}{2} (v-a-Bu)^T S (v-a-Bu) \right] = \exp \left[ -\frac{1}{2} u^T B^T C Bu \right] \tilde{f}(v) \exp \left[ -\frac{1}{2} z^T (F+S) z \right],$$

where $C = (F+S)^{-1}$ and $z = v - (F+S)^{-1} S(a+Bu)$

**Proof.** Using Lemma 5 with $u$ there replaced by $a + Bu$,

$$f(v) \exp \left[ \frac{1}{2} (v-a-Bu)^T S (v-a-Bu) \right] = \tilde{f}(v) \exp \left[ -\frac{1}{2} \{ u^T Fv + (v-a-Bu)^T S (v-a-Bu) \} \right]$$

$$= \tilde{f}(v) \exp \left[ -\frac{1}{2} \{ (a+Bu)^T C (a+Bu) + z^T (F+S) z \} \right]$$

$$= \exp \left[ -\frac{1}{2} u^T B^T C Bu \right] \exp \left[ -\frac{1}{2} (a^T Ca + 2a^T CBu) \right] \tilde{f}(v) \exp \left[ -\frac{1}{2} z^T (F+S) z \right].$$

\[\square\]
Proof of Proposition 2. First note that for the signal process \((\theta)_{t \in \mathbb{R}^+}\) as per (1.1),
\[
m_t := E_{\theta_0}[\theta_t] = a_t + B_t \theta_0,
\[
\Sigma_t := E_{\theta_0}[(\theta_t - m_t)(\theta_t - m_t)^T] = \sigma^2 \int_0^t B_{t-s} B_{t-s}^T ds,
\]
where
\[
a_t := B_t \int_0^t s^{-1} ds, \quad B_t := e^{\beta t}.
\]
It follows that \(u^T \Sigma_t^{-1} u \geq \Lambda_t^{-1} u^T u\) for all \(u \in \mathbb{R}^p\) with the shorthand \(\Lambda_t := \sigma^2 \int_0^t \lambda_{\beta}^{\min}(s) ds\).

Applying Lemma 6 with \(a = a_t, B = B_t, S = I \Lambda_t^{-1}, f = \varphi_{j,k}, F = I \lambda_\varphi(j)\), and Lemma 2,
\[
\varphi_{j,k}(\theta) \exp \left[ -\frac{1}{2} \left( \theta - a_t - B_t \theta_0 \right)^T \Sigma_t^{-1} \left( \theta - a_t - B_t \theta_0 \right) \right] = \exp \left( -\frac{1}{2} \left( \frac{\lambda_{\varphi}(j) \lambda_{\beta}^{\min}(t)}{2 \lambda_{\varphi}(j) \lambda_t} \theta_0^T \right) \right)
\cdot \varphi_{j,k}(\theta) \exp \left[ -\frac{1}{2} \left( \lambda_{\varphi}(j) + \Lambda_t^{-1} \right) \right] \cdot \exp \left[ -\frac{1}{2} \left( \frac{\lambda_{\varphi}(j) \lambda_{\beta}^{\min}(t)}{2 \lambda_{\varphi}(j) \lambda_t} \theta_0^T \right) \right] (B_t^T B_t - I \lambda_{\beta}^{\min}(t)) \theta_0 \right] \cdot \exp \left[ -\frac{1}{2} \left( \theta - a_t - B_t \theta_0 \right)^T (\Sigma_t^{-1} - \Lambda_t^{-1} I) \left( \theta - a_t - B_t \theta_0 \right) \right], \quad (2.12)
\]
where \(C_t = I - \frac{\lambda_{\varphi}(j) \lambda_{\beta}^{\min}(t)}{1 + \lambda_{\varphi}(j) \lambda_t} (a_t + B_t \theta_0)\).

The product of the terms in (2.12)-(2.14) is jointly log-concave in \((\theta_0, \theta)\). Therefore by Lemma 4, there exists a function \(h\) such that \(\theta \mapsto h(\theta, t)\) is log-concave and
\[
h(\theta_0, t) = P^{\Delta t} \varphi_{j,k}(\theta_0)
\]
\[
= \int \varphi_{j,k}(\theta) \exp \left[ -\frac{1}{2} \left( \theta - a_{\Delta t} - B_{\Delta t} \theta_0 \right)^T \Sigma_{\Delta t}^{-1} \left( \theta - a_{\Delta t} - B_{\Delta t} \theta_0 \right) \right] d\theta
\]
\[
= \exp \left( -\frac{1}{2} \left( \frac{\lambda_{\varphi}(j) \lambda_{\beta}^{\min}(\Delta - t)}{2 \lambda_{\varphi}(j) \lambda_{\Delta - t}} \theta_0^T \right) \right) \tilde{h}(\theta_0, t),
\]
which completes the proof. \(\square\)

3 Smoothing distributions and a family of weighted Wasserstein distances

Obtaining a satisfactory generalization of Theorem 1 to allow for initial distributions \(\mu\) other than Dirac measures appears to be a non-trivial matter. The difficulty is that the corresponding generalization of (2.5) from which to start is:
\[
\pi_k^\mu(A) = \mu_{0,k} R_{1,k} R_{2,k} \cdots R_{k,k}(A), \quad \mu_{0,k} := \frac{\mu \cdot \varphi_{0,k}}{\mu \varphi_{0,k}},
\]
so a direct corollary of Theorem 1 is:
\[
W_\varphi(\pi_k^\mu, \pi_k^\nu) \leq \exp \left[ -\sum_{j=1}^k \int_0^\Delta \lambda(j, t) dt \right] W_\varphi(\mu_{0,k}, \nu_{0,k}). \quad (3.1)
\]
but it cannot immediately be deduced from (3.1) that \( \lim_k W_q(\pi_k^\mu, \pi_k^\nu) = 0 \), due to the dependence of \( W_q(\mu_{0,k}, \nu_{0,k}) \) on \( k \).

An alternative is to work with a certain family of weighted Wasserstein distances between filtering distributions. As we shall see, this is equivalent to establishing forgetting of the initial condition for smoothing distributions, which unlike filtering distributions condition on future as well as past and present observations. The starting point from which to describe this equivalence in more detail is the following lemma.

**Lemma 7.** Let \( d(\cdot, \cdot) \) be a metric on the set of probability measures on \( B(\mathbb{R}^p) \) and let \( \phi : \mathbb{R}^p \to (0, \infty) \). Then \( d^\phi(\cdot, \cdot) \) defined by:

\[
d^\phi : (\mu, \nu) \mapsto d\left( \frac{\mu \cdot \phi}{\mu \phi}, \frac{\nu \cdot \phi}{\nu \phi} \right)
\]

is a metric on the subset of probability measures \( \{ \mu \in B(\mathbb{R}^p) : \mu \phi < \infty \} \).

**Proof.** It follows immediately from the assumption that \( d \) is a metric and \( \phi \) is strictly positive that on the given domain \( \{ \mu : \mu \phi < \infty \} \), \( d^\phi \) is nonnegative, symmetric, satisfies the triangle inequality and \( \mu = \nu \Rightarrow d^\phi(\mu, \nu) = 0 \). For the reverse implication, we have \( d^\phi(\mu, \nu) = 0 \Rightarrow \mu \phi = \nu \phi \), so since \( \phi \) is strictly positive, \( \mu \ll \nu \) and \( 1 = d\mu \phi / d\nu \phi = (d\mu / d\nu)(\nu \phi / \mu \phi), \nu \phi \)-a.e. and then also \( \nu \)-a.e. since \( \phi \) is strictly positive. Thus \( d\mu / d\nu = \text{const.}, \nu \)-a.e. and since \( \mu \) and \( \nu \) are probability measures, it follows that \( \mu = \nu \). \( \square \)

Introduce the nonnegative integral kernels

\[
Q_k(\theta, d\theta) := g_{k-1}(\theta, g_{k-1}) P_\Delta(\theta, d\theta), \quad k \geq 1, \quad Q_{j,k} := Q_{j+1} \cdots Q_k, \quad 0 \leq j < k.
\]

and the probability measures

\[
\eta_k^\mu(A) := \frac{\mu Q_{0,k}(1_A)}{\mu Q_{0,k}(\mathbb{1}_{\mathbb{R}^p})}, \quad k \geq 1, \quad \eta_0^\mu := \mu, \quad A \in B(\mathbb{R}^p),
\]

for any \( \mu \) such that the denominator is finite. Note from (1.1) that \( \eta_k^\mu = \pi_k^\mu P_\Delta \).

We shall use the functions appearing in the following assumption to define a family of weighted Wasserstein distances.

**Assumption 2.** There exists a probability measure \( \mu_0 \) such that for each \( k \in \mathbb{N}_0 \), the following pointwise limit exists:

\[
\phi_{k,\infty}(x) := \lim_{\ell \to \infty} \frac{\phi_{k,\ell}(x)}{\eta_k^\mu}(\phi_{k,\ell}^{-1}, \eta_k^\mu) \quad (3.3)
\]

\( \phi_{k,\infty}(\theta) \in (0, \infty) \) for all \( \theta \in \mathbb{R}^p \), and the functions \( (\phi_{k,\infty})_{k \in \mathbb{N}_0} \) so-defined belong to \( C^2 \) and satisfy

\[
Q_k \phi_{k,\infty} = \zeta_{k-1} \phi_{k-1,\infty}, \quad k \geq 1,
\]

where \( \zeta_k := \int \eta_k^\mu (du) g_k(u, y_k) \in (0, \infty) \).

Before discussing the interpretation of Assumption 2, consider the following lemma, which mirrors Lemma 2.

**Lemma 8.** If Assumption 2 holds, then for any \( \mu \) such that for all \( k \in \mathbb{N}_0 \), \( \pi_k^\mu P_\Delta \phi_{k+1,\infty} < \infty \), the probability measures \( (\pi_k^\mu)_{k \in \mathbb{N}_0} \) defined by:

\[
\pi_k^\mu(A) := \frac{\pi_k^\mu(\mathbb{1}_A P_\Delta \phi_{k+1,\infty})}{\pi_k^\mu P_\Delta \phi_{k+1,\infty}}, \quad A \in B(\mathbb{R}^p),
\]

satisfy

\[
\pi_k^\mu(A) = \pi_0^\mu R_{1,\infty} \cdots R_{k,\infty}(A),
\]
with the Markov kernels
\[
R_{k,\infty}(\theta, d\theta) := \frac{P_\Delta(\theta, d\theta)\phi_{k,\infty}(\theta)}{P_\Delta\phi_{k,\infty}(\theta)}.
\]

If additionally Assumption 1 holds, then for each \(k \in \mathbb{N}_0\), there exists a log-concave function \(\tilde{\phi}_{k,\infty}\) such that
\[
\phi_{k,\infty}(\theta) = \exp\left[\frac{-\lambda_g(k)}{2} \theta^T \theta\right] \tilde{\phi}_{k,\infty}(\theta).
\]

**Proof.** To establish (3.6) it suffices to show \(\pi_{k-1,\infty}^\mu R_{k,\infty} = \pi_{k,\infty}^\mu\). We have
\[
\pi_{k-1,\infty}^\mu R_{k,\infty}(A) = \frac{\pi_{k-1}^\mu(P_\Delta(\phi_{k,\infty})R_{k,\infty}(1_A))}{\pi_{k-1}^\mu P_\Delta \phi_{k,\infty}}
\]
\[
= \frac{\pi_{k-1}^\mu P_\Delta(1_A \phi_{k,\infty})}{\pi_{k-1}^\mu P_\Delta \phi_{k,\infty}}
\]
\[
= \frac{\pi_{k-1}^\mu P_\Delta(1_A Q_k \phi_{k+1,\infty})}{\pi_{k-1}^\mu P_\Delta Q_k \phi_{k+1,\infty}}
\]
\[
= \frac{\pi_{k-1}^\mu(1_A P_\Delta \phi_{k+1,\infty})}{\pi_{k-1}^\mu P_\Delta \phi_{k+1,\infty}} = \pi_{k,\infty}^\mu(A),
\]
where (3.4), (3.2) and the identity \(\pi_{k}^\mu(A) = \pi_{k-1}^\mu[P_\Delta(1_A Q_k 1_{ RP})]/\pi_{k-1}^\mu[P_\Delta( Q_k 1_{ RP})]\) have been used.

For the second claim, the fact that \(\phi_{j,\infty}\) is log-concave for every \(j \in \mathbb{N}_0\) follows from its definition as the pointwise limit in (3.3) and the log-concavity of \(\varphi_{j,k}\) established in Lemma 2. By Lemma 1, \(P_\Delta \phi_{k+1}\) is log-concave and since by Assumption 2, \(\phi_{k,\infty} = \varsigma_k^{-1} Q_{k+1} \phi_{k+1,\infty}\), we may take \(\tilde{\phi}_{k,\infty}(\theta) = \varsigma_k^{-1} g_k(\theta) P_\Delta \phi_{k+1,\infty}(\theta)\).

Recalling from section 1 the interpretation of \(\pi_{k}^\mu\) as the conditional distribution of \(\theta_{k,\infty}\) given \((y_0, \ldots, y_k)\), the measure \(\pi_{k}^\mu \cdot (P_\Delta \varphi_{k+1,\infty})/\pi_{k}^\mu P_\Delta \varphi_{k+1,\infty}\) is the smoothing distribution which conditions additionally on \((y_{k+1}, \ldots, y_{k+r})\). The interpretation of (3.3) is then that \(\phi_{k,\infty}\) is the function with which to re-weight \(\pi_{k}^\mu P_\Delta\) in order to condition on the infinite data record \((y_{k+1})_{k\in\mathbb{N}_0}\).

The question of whether there exists a well-behaved (in the sense of satisfying the other requirements of Assumption 2) function which achieves this conditioning is closely connected to the question of filter stability, see [14] for a discussion on doubly infinite time horizons. Indeed it is clear from (3.5) that Assumption 2 implies that the filtering and smoothing measures, \(\pi_{k}^\mu\) and \(\pi_{k,\infty}^\mu\), are equivalent, despite the fact that \(\pi_{k,\infty}^\mu\) conditions on an infinite number of observations. Various existing tools are available to verify Assumption 2, we shall illustrate some of them in an example below, it is an open question whether Assumption 2 can be deduced directly from Theorem 1.

When Assumption 2 holds, we shall consider the family of weighted Wasserstein distances
\[
W_{q,k}(\mu, \nu) := W_q\left(\frac{\mu \cdot P_\Delta \phi_{k+1,\infty}}{\mu P_\Delta \phi_{k+1,\infty}}, \frac{\nu \cdot P_\Delta \phi_{k+1,\infty}}{\nu P_\Delta \phi_{k+1,\infty}}\right), \quad k \in \mathbb{N}_0,
\]
whenever \(\mu, \nu\) satisfy appropriate integrability conditions for this object to be well-defined. The interest in the distances \(W_{q,k}\) is in the identity:
\[
W_{q,k}(\pi_{k}^\mu, \pi_{k}^\nu) = W_q(\pi_{k,\infty}^\mu, \pi_{k,\infty}^\nu),
\]
which follows from (3.5). Thus \(W_{q,k}\) quantifies distance between \(\pi_{k}^\mu\) and \(\pi_{k}^\nu\) as the \(W_q\)-distance between the corresponding smoothing distributions \(\pi_{k,\infty}^\mu\) and \(\pi_{k,\infty}^\nu\).

We denote the set of probability measures
\[
\mathcal{P}_q := \left\{\mu \text{ on } B(\mathbb{R}^p) : \int (1 + \|u\|^q)\phi_0(u)\mu(du) < \infty \text{ and } \pi_{k}^\mu P_\Delta \phi_{k+1,\infty} < \infty, \forall k \in \mathbb{N}_0\right\}.
\]
Theorem 2. If Assumptions 1 and 2 hold, then for any \( q \geq 1 \),

\[
W_{q,k}(\pi_k^\mu, \pi_k^\nu) \leq \exp \left[ -\sum_{j=1}^{k} \int_0^\Delta \lambda(j, t) dt \right] W_{q,0}(\pi_0^\mu, \pi_0^\nu), \quad \forall k \geq 1, \mu, \nu \in \mathcal{P}_q,
\]

where \( \lambda(j, t) \) is as in Theorem 1.

Given the identities (3.6) and (3.7), the proof of Theorem 2 follows almost exactly the same programme as the proof of Theorem 1, except working with the kernels \( R_{k,\infty} \), the functions \( \phi_{k,\infty} \) and their log-concavity in Lemma 8, instead of \( R_{j,k}, \phi_{j,k} \) and their log-concavity in Lemma 2. The requirement \( \mu, \nu \in \mathcal{P}_q \) ensures that \( W_{q,0}(\mu, \nu) \) and \( \pi_{k,\infty}^\mu, \pi_{k,\infty}^\nu \) are well-defined.

Example - dynamic logistic regression

Consider the case: \( \sigma^2 > 0, \beta = -I\lambda_{\text{sig}} \) for some \( \lambda_{\text{sig}} > 0 \), and with \( \mathcal{Y} = \{0, 1\}^n \), the observations \( Y_k = (Y_k^1, \cdots, Y_k^n) \) are conditionally independent given \( \theta_{k,\Delta} \), with the conditional probability of \( Y_k^i = 1 \) being \( 1/(1 + e^{-\sum_{t=1}^{\Delta} \theta_{j,t} x_{k,j}^i}) \), where \( x_{k,j}^i \) are known covariates. The likelihood function is then:

\[
g_k(\theta, y_k) = \exp \left[ \sum_{i=1}^{n} \left\{ -\sum_{j=1}^{p} y_k^i x_{k,j}^i \theta_j \log \left( 1 + e^{\sum_{j=1}^{p} x_{k,j}^i \theta_j} \right) \right\} \right].
\]

For any \( (y_k)_{k \in \mathbb{N}_0} \), Assumption 1 is satisfied with \( \lambda_{\vartheta}(k) = 0 \), and therefore (2.2) holds by Theorem 1. Checking Assumption 2 is more involved, we shall use some results from [13].

Let us assume that the covariates satisfy

\[
\sup_{k \geq 0} \sum_{i,j} (x_{k,j}^i)^2 < \infty, \quad (3.8)
\]

and fix an arbitrarily sequence of observations \( (y_k)_{k \in \mathbb{N}_0} \).

The following properties of this model are easily checked (see [13, Sec. 3.1] for a similar example): there exists a constant \( c > 0 \) such that with

\[
V(\theta) := 1 + c\|\theta\|, \quad C_d := \{\theta \in \mathbb{R}^p : V(\theta) \leq d\},
\]

we have for some \( d \in [1, \infty) \) and all \( d \geq d \),

- \( \sup_k g_k(\theta, y_k) \leq 1 \), \( \forall \theta \in \mathbb{R}^p \), and there exist constants \( \delta \in (0, 1), b_d \in [0, \infty) \) such that

\[
P_{\Delta}(e^V) \leq \exp(V(1 - \delta) + b_d 1_{C_d}), \quad (3.10)
\]

- \( \inf_k g_k(\theta, y_k) P_{\Delta}(\theta, C_d) > 0 \), \( \forall \theta \in \mathbb{R}^p \),

- there exist constants \( \epsilon_{d}^{-} \), \( \epsilon_{d}^{+} \) such that \( \forall \theta \in C_d \) and \( k \in \mathbb{N}_0 \),

\[
\epsilon_{d}^{-} \nu_d(\partial \vartheta) 1_{C_d}(\vartheta) \leq g_k(\theta, y_k) P_{\Delta}(\theta, \partial \vartheta) 1_{C_d}(\vartheta) \leq \epsilon_{d}^{+} \nu_d(\partial \vartheta) 1_{C_d}(\vartheta),
\]

where the probability measure \( \nu_d \) is the normalized restriction of Lebesgue measure to \( C_d \).

Define the norm on functions \( f : \mathbb{R}^p \to \mathbb{R}, \|f\|_{e^V} := \sup_{\vartheta} |f(\vartheta)|/e^{V(\vartheta)} \).

Proposition 3. For any \( \mu_0(e^V) < \infty \), define \( \phi_{j,k}(\theta) := \varphi_{j,k}(\theta)/\pi_{j-1}^{\mu_0} P_{\Delta}\varphi_{j,k} \). Then:

1) \( \sup_{k \geq 0} \eta_k^{\mu_0}(e^V) < \infty \)

2) \( \sup_{0 \leq j \leq k} \|\phi_{j,k}\|_{e^V} < \infty \),

3) for all \( d \geq d \), \( \inf_{0 \leq j \leq k} \inf_{\theta \in C_d} \phi_{j,k}(\theta) > 0 \),
4) for all $0 < j \leq k$, $Q_j \phi_{j,k} = \varsigma_j \phi_{j-1,k}$, where $\varsigma_j = \int \eta_j(\theta)g_j(\theta, y_j)$.  
5) there exist constants $\rho < 1$ and $c_{\mu_0} < \infty$ such that for any $f : \mathbb{R}^p \to \mathbb{R}$ with $\|f\|_{e^V} < \infty$,

$$\left| \frac{Q_j f(\theta)}{\prod_{i=j}^k \varsigma_i} - \phi_{j,k-1}(\theta) \eta^{(0)}_k f \right| \leq \rho^{k-j} \|f\|_{e^V} c_{\mu_0} e^{V(\theta)} \mu_0(e^V), \quad \forall \theta \in \mathbb{R}^p, 0 \leq j < k$$

**Proof.** The properties identified immediately before the statement of proposition and the requirement $\mu_0(e^V) < \infty$ imply that conditions (H1)-(H4) of [13] are satisfied. Then 1) and 2) are established by [13, Prop. 1 and 2], 3) by [13, Lem. 10], 4) by [13, Lem. 1], and 5) by [13, Thm. 1].

The following proposition establishes that the conditions of Theorem 2 are satisfied. 

**Proposition 4.** For any sequence of observations $(y_k)_{k \in \mathbb{N}_0}$, the dynamic logistic regression model described above satisfies Assumption 2 with $\sup_{k \geq 0} \|\phi_{k,\infty}\|_{e^V} < \infty$, and for any $q \geq 1$,

$$W_{q,k}(\pi^\mu_0, \pi^\nu_0) \leq \exp(-k \Delta_{\text{sig}}) W_{q,0}(\pi^\mu_0, \pi^\nu_0), \quad (3.11)$$

for all $\mu, \nu$ in the set of probability measures $\{\mu \in \mathcal{B}(\mathbb{R}^p) : \int (1 + \|u\|^q) e^{c\|u\|} \mu(du) < \infty\}$ where $c$ is as in (3.9).

**Remark 1.** The constant $\rho < 1$ appearing in part 5) of Proposition 3 and obtained using the techniques of [13] may degrade with dimension of the state-space. Note however, that $\rho$ does not appear in (3.11), it only serves as an intermediate tool used to in the following proof to help establish that Assumption 2 holds.

**Proof of Proposition 4.** Choose any $\mu_0$ such that $\mu_0(e^V) < \infty$. Noting the identities $\pi^{\mu_0}_{k-1} P \Delta \varphi_{k,\ell} = \prod_{i=k}^j \varsigma_j$ and $\phi_{j,k} = Q_{j,k+1} 1_{\mathbb{R}^p} / \prod_{i=j}^k \varsigma_i$, we have for any $\ell \geq 1$,

$$\phi_{j,k} - \phi_{j,k+\ell} = \frac{Q_{j,k+1}}{\prod_{i=j}^k \varsigma_i} \left( 1 - \frac{Q_{k+1,k+\ell+1} 1_{\mathbb{R}^p}}{\prod_{i=k+1}^{k+\ell} \varsigma_i} \right).$$

Since $\prod_{i=k+1}^{k+\ell} \varsigma_i = \eta^{\mu_0}_{k+1} Q_{k+1,k+\ell+1} 1_{\mathbb{R}^p}$, we have $\eta^{\mu_0}_{k+1} (1 - \frac{Q_{k+1,k+\ell+1} 1_{\mathbb{R}^p}}{\prod_{i=k+1}^{k+\ell} \varsigma_i}) = 0$ and by part 2) of Proposition 3, $\sup_{j,k,\ell} \|\frac{Q_{k+1,k+\ell+1} 1_{\mathbb{R}^p}}{\prod_{i=k+1}^{k+\ell} \varsigma_i}\|_{e^V} = cQ < \infty$, so an application of part 5) of Proposition 3 gives:

$$\|\phi_{j,k} - \phi_{j,k+\ell}\|_{e^V} \leq \rho^{k+1-j} cQ c_{\mu_0} \mu_0(e^V), \quad \forall \ell \geq 1.$$

It follows for each $j$, $(\phi_{j,k})_{k \geq j}$ is a Cauchy sequence in the Banach space of functions $f : \mathbb{R}^p \to \mathbb{R}$ endowed with the norm $\|f\|_{e^V} < +\infty$. With the strong limit of $(\phi_{j,k})_{k \geq j}$ then denoted $\phi_{j,\infty}$, we have $\|\phi_{j,\infty}\|_{e^V} < \infty$ and $\phi_{j,\infty}(\theta) = \lim_{k \to \infty} \phi_{j,k}(\theta)$ pointwise.

From part 4) of Proposition 3,

$$Q_j \phi_{j,k} = Q_j \phi_{j,\infty} + Q_j (\phi_{j,k} - \phi_{j,\infty}) = \varsigma_j \phi_{j-1,\infty} + \varsigma_j (\phi_{j-1,k} - \phi_{j-1,\infty}) = \varsigma_j \phi_{j-1,k},$$

and since using (3.10), $\|Q_j (\phi_{j,k} - \phi_{j,\infty})\|_{e^V} \leq \|Q_j (\phi_{j,k} - \phi_{j,\infty})\|_{e^V} \to 0$, both as $k \to \infty$, we have $Q_j \phi_{j,\infty} = \varsigma_j \phi_{j-1,\infty}$. Since $g_j(\theta, y_j) \in (0, 1)$, we have $\varsigma_j \in (0, 1)$ and using part 3) of Proposition 3, $Q_j \phi_{j,\infty}(\theta) > 0$ for all $\theta$ hence $\phi_{j-1,\infty}(\theta) > 0$ for all $\theta$. Also $\|\phi_{j,\infty}\|_{e^V} < \infty$ implies $\phi_{j,\infty}(\theta) < \infty$ for all $\theta$. The membership $\phi_{j-1,\infty} \in C^2$ follows from $Q_j \phi_{j,\infty} = \varsigma_j \phi_{j-1,\infty}$ together with $\theta \mapsto g_{j-1}(\theta, y_{j-1}) \in C^2$ by Assumption 1 and the fact that $\Delta_\varphi$ is given by (1.3). That completes the verification of Assumption 2.

To complete the proof, observe that in order for $\mu \in \mathcal{P}_q$ it is sufficient that $\int (1 + \|\theta\|^q) e^{V(\theta)} \mu(d\theta) < \infty$, because using part 2) of Proposition 3, $\sup_{k \geq 0} \|\phi_{k,\infty}\|_{e^V} < \infty$, we have $\pi^{\mu_0}_{k-1} P \Delta = \eta^{\mu_0}_k$ and by part 1) of Proposition 3, $\sup_k \|\eta^{\mu_0}_k\|_{e^V} < \infty$.

**Acknowledgement.** The author thanks Anthony Lee for helpful comments.
References

[1] P. Cattiaux and A.Guillin. Semi log-concave Markov diffusions. In Séminaire de probabilités XLVI, pages 231–292. Springer, 2014.

[2] D. Crisan and B. Rozovskii. The Oxford handbook of nonlinear filtering. Oxford University Press, 2011.

[3] R. Douc, E. Gassiat, B. Landelle, and E. Moulines. Forgetting of the initial distribution for nonergodic hidden Markov chains. The Annals of applied probability, 20(5):1638–1662, 2010.

[4] M. Gerber and N. Whiteley. Stability with respect to initial conditions in V-norm for nonlinear filters with ergodic observations. Journal of Applied Probability, 54(1):118–133, 2017.

[5] J. Harrison and M. West. Bayesian forecasting & dynamic models, volume 1030. Springer New York City, 1999.

[6] M.L. Kleptsyna and A.Y. Veretennikov. On discrete time ergodic filters with wrong initial data. Probability Theory and Related Fields, 141(3-4):411–444, 2008.

[7] A.V. Kolesnikov. On diffusion semigroups preserving the log-concavity. Journal of Functional Analysis, 186(1):196–205, 2001.

[8] P. McCullagh and J.A Nelder. Generalized Linear Models, volume 37 of Monograph on Statistics and Applied Probability. Chapman & Hall,, 1989.

[9] A. Prékopa. On logarithmic concave measures and functions. Acta Scientiarum Mathematicarum, 34:334–343, 1973.

[10] P. Rebeschini and R. Van Handel. Can local particle filters beat the curse of dimensionality? The Annals of Applied Probability, 25(5):2809–2866, 2015.

[11] D. Revuz and M. Yor. Continuous martingales and Brownian motion. Springer, 3rd edition, 1999.

[12] X.T. Tong and R. van Handel. Conditional ergodicity in infinite dimension. The Annals of Probability, 42(6):2243–2313, 2014.

[13] N. Whiteley. Stability properties of some particle filters. The Annals of Applied Probability, 23(6):2500–2537, 2013.

[14] N. Whiteley and A. Lee. Twisted particle filters. The Annals of Statistics, 42(1):115–141, 2014.