STABILITY OF THE DISCRETE TIME FILTER IN TERMS OF THE TAILS OF NOISE DISTRIBUTIONS

D. CRISAN AND K. HEINE

Abstract

The stability of the discrete time filter has been a field of active research throughout the recent years. By stability we mean that the effect of the possibly erroneous initial distribution in the filter eventually vanishes as time increases. One of the motivations for the interest in the stability is its close relation to the convergence of various numerical filter approximation schemes, e.g. particle filters. In this paper, the main result states easily verifiable conditions that are sufficient for the filter stability. Essentially, the conditions state that the filter is stable, if the observation noise is sufficiently light tailed compared to the randomness in the signal process. Compactness of the state space or ergodicity of the signal are not required.

1. Introduction and the Main Result

There are numerous applications in engineering where the unknown current state of some dynamical system is to be estimated based on partial observations made until the current time. This estimation problem is known as the stochastic discrete time filtering problem, when the unobserved part is modelled as a discrete time stochastic process \((X_i)_{i \geq 0}\) called the signal process, and the observed part is modelled as a discrete time stochastic process \((Y_i)_{i > 0}\), called the observation process. The solution of this filtering problem is a probability measure valued stochastic process \((\pi^*_i)_{i \geq 0}\), called the filter process, where \(\pi^*_i\) is the conditional distribution of \(X_i\) given the \(\sigma\)-algebra generated by the observations \((Y_1, Y_2, \ldots, Y_i)\). Roughly speaking, this \(\sigma\)-algebra can be interpreted as all the information available at time \(i\) and the conditional distribution \(\pi^*_i\) can be interpreted to represent our knowledge of the state of the signal in the light of this information.

The filter process \((\pi^*_i)_{i \geq 0}\) can be expressed recursively, and the filter is said to be stable if the distance between two filter recursions with different initial distributions vanishes as time increases. Throughout the recent years, there has been active research on the stability of the filter. Perhaps the most important motivation for this interest in the stability is the analysis of time uniform convergence of various numerical approximation schemes of the generally intractable filter recursion. Intuitively, the connection between the stability and the uniform convergence is due to the recursive nature of the filter. Because of the recurrence, the approximation error that was made earlier will contribute to the computation of the approximation at current time. This propagation of error can be considered as an incorrect initialisation of the filter at the time when the error was made, that is, every step of the recursion. If the filter is stable enough, the effect of each incorrect initialisation will vanish fast enough to ensure that the combined effect of all the errors remains bounded. Thus the uniform convergence analysis can be

brought under the framework of filter stability with respect to the initial conditions. The remainder of this work is focused solely on the analysis of the stability and further considerations regarding the analysis of the uniform convergence are left beyond the scope of this work. More detailed discussion on the connection between the stability and the uniform convergence can be found e.g. in [7, 14, 10].

Let us then describe in more detail the filtering framework under consideration. Let $\mathcal{B}(\mathbb{R}^d)$ denote the Borel $\sigma$-algebra on $\mathbb{R}^d$ and let $\lambda_d$ denote the Lebesgue measure on $\mathcal{B}(\mathbb{R}^d)$. Moreover, let $\mathcal{P}(\mathcal{S})$ denote the set of probability measures on an arbitrary $\sigma$-algebra $\mathcal{S}$. We define the signal process $(X_i)_{i \geq 0}$ such that for all $i > 0$

$$X_i = f_i(X_{i-1}) + W_i,$$  \hspace{1cm} (1.1)

where $X_0$ is a random variable with distribution $P_0 \in \mathcal{P}(\mathcal{B}(\mathbb{R}^{d_m}))$, $f_i : \mathbb{R}^{d_m} \to \mathbb{R}^{d_m}$ is measurable and $W_i$ is a random variable with distribution $P_{W_i} \in \mathcal{P}(\mathcal{B}(\mathbb{R}^{d_m}))$. It is assumed that $X_0$ and $(W_i)_{i > 0}$ are independent and for all $i > 0$, $P_0$, $P_{W_i}$ and $\lambda_{d_m}$ are absolutely continuous with respect to each other. The density of $P_{W_i}$ with respect to $\lambda_{d_m}$ is denoted by $\rho_{W_i}$.

The observation process $(Y_i)_{i > 0}$, is defined for all $i > 0$ as

$$Y_i = h_i(X_i) + V_i,$$

where $h_i : \mathbb{R}^{d_m} \to \mathbb{R}^{d_m}$ is measurable and $V_i$ is a random variable with distribution $P_{V_i} \in \mathcal{P}(\mathcal{B}(\mathbb{R}^{d_m}))$. It is assumed that the sequence $(V_i)_{i > 0}$ is independent and for all $i > 0$, $P_{V_i}$ and $\lambda_{d_m}$ are absolutely continuous with respect to each other. The density of $P_{V_i}$ with respect to $\lambda_{d_m}$ is denoted by $\rho_{V_i}$. The stochastic discrete time filter is then the probability measure valued stochastic process $(\pi_i^*)_{i \geq 0}$ such that for all bounded and measurable functions $\varphi : \mathbb{R}^{d_m} \to \mathbb{R}$,

$$\int \varphi d\pi_i^* = E[\varphi(X_0)]$$

$$\int \varphi d\pi_i^* = E[\varphi(X_i) | \mathcal{Y}_i],$$

where $i > 0$ and $\mathcal{Y}_i$ is the $\sigma$-algebra generated by the observations $(Y_1, Y_2, \ldots, Y_i)$. The filter process can also be expressed as a recursion

$$\pi_i^* = Q_i(\pi_{i-1}^*),$$

where $\pi_0^* = P_0$ and $Q_i : \mathcal{P}(\mathcal{B}(\mathbb{R}^{d_m})) \to \mathcal{P}(\mathcal{B}(\mathbb{R}^{d_m}))$ is to be specified later in Section 3. We say that the filter $(\pi_i^*)_{i \geq 0}$ is stable if for all $\pi_0, \bar{\pi}_0 \in \mathcal{P}(\mathcal{B}(\mathbb{R}^{d_m}))$

$$\lim_{i \to \infty} \left\| Q_i \circ \cdots \circ Q_1(\pi_0) - Q_i \circ \cdots \circ Q_1(\bar{\pi}_0) \right\|_{TV} = 0, \hspace{0.5cm} \text{P-a.s.,}$$

(1.2)

where $\left\| \cdot \right\|_{TV}$ denotes the total variation norm. In order to analyse the stability of the filter, we will impose the following assumptions on the filter application:

(A1) Functions $f_i, i > 0$ are $\alpha$-Lipschitz, that is, for all $x, y \in \mathbb{R}^{d_m}$,

$$\|f_i(x) - f_i(y)\| \leq a \|x - y\|,$$

where $a \in [0, \infty)$, and $\| \cdot \|$ denotes the Euclidean norm.

(A2) Functions $h_i, i > 0$ are bijective such that $h_i$ and $h_i^{-1}$ are $b$-Lipschitz.

(A3) There are $m_1, M_1, \alpha_1, \beta_1 > 0$ such that for all $i > 0$, $\rho_{W_i}$ satisfies

$$m_1 \exp \left( -\alpha_1 \|x\|^{\beta_1} \right) \leq \rho_{W_i}(x) \leq M_1 \exp \left( -\alpha_1 \|x\|^{\beta_1} \right).$$
There are \( m_2, M_2, \alpha_2, \beta_2 > 0 \) such that for all \( i > 0 \), \( \rho V_i \) satisfies
\[
m_2 \exp\left(-\alpha_2 \|x\|^{\beta_2}\right) \leq \rho V_i(x) \leq M_2 \exp\left(-\alpha_2 \|x\|^{\beta_2}\right).
\]

In this case, we have the following result where the constant \( \kappa_0 \) will be specified later in Section 4 before the proof of the theorem.

**Theorem 1.1.** If one of the following two conditions holds:

(i) \( \beta_1 = \beta_2 \) and \( \alpha_2 > \kappa_0 \),
(ii) \( \beta_1 < \beta_2 \),

then \((\pi^*_i)_{i \geq 0}\) is stable.

Essentially, Theorem 1.1 states that under the assumptions (A1) . . . (A4) the filter is stable provided that the tails of the observation noise \((V_i)_{i > 0}\) are sufficiently light compared to the tails of the signal noise \((W_i)_{i > 0}\). Conditions such as signal ergodicity or mixing signal kernel that in practice are often found to be too restrictive, are not required. According to this interpretation, it is somewhat counterintuitive that the first inequality in (A3) and the second inequality in (A4) are not sufficient but we also need to assume the existence of the upper bound in (A3) and the lower bound in (A4). However, these assumptions are crucial to the analysis and cannot be omitted (see Proposition 4.2 and Proposition 4.3).

Moreover, it appears that the analysis can be extended for the signal model
\[
X_i = f_i(X_{i-1}) + b_i(X_{i-1}) W_i,
\] (1.3)
where \( b_i : \mathbb{R}^d_i \to \mathbb{R} \) such that \( \sup_{i \geq 0, x \in \mathbb{R}_i} |b_i(x)| < \infty \) and \( \inf_{i \geq 0, x \in \mathbb{R}_i} |b_i(x)| > 0 \).

In this case, \( b_i(X_i) W_i \) satisfies an assumption similar to (A3) provided that \( W_i \) satisfies (A3) and therefore Theorem 1.1 should hold. Potentially, the result can be further extended for even more general signal models, such as (1.3) where \( b_i \) is only assumed Lipschitz. This claim, however, requires more detailed considerations and is therefore left for future research.

Theorem 1.1 is a refinement of the stability result given in [14] and the proof is similarly based on approximating the exact filter by a truncated filter with a compact domain as proposed in [11, 14]. Filters with compact domains are exponentially stable under mild conditions [1, 7]. The main contribution of this work is due to the fact that to a large extent the analysis is done in the almost sure sense and therefore it is substantially different from the approach described in [14]. Several benefits follow from this difference. Firstly, the extension of the stability theorem for more general signal and observation noise distributions is straightforward because almost sure bounds for the signal and observation noise terms are easily obtained in the almost sure sense. Secondly, the analysis provides convergence rates directly in the almost sure sense, which implies convergence in the mean sense by the dominated convergence theorem. Thirdly, the analysis provides almost sure nonuniform bounds for the error of the truncated filter approximation, which implies the almost sure convergence of the truncated filter approximation. Finally, unlike several other works [see, e.g. 2, 1, 12, 3, 14], the Hilbert metric is not used at any point of the analysis. The result in [14] holds only for filters whose initial distribution is comparable to the exact initial distribution in the sense of
Hilbert metric. This rather restrictive assumption is avoided by the approach used in this work\(^\dagger\).

Let us briefly review some existing works related to the stability of the discrete time filters. Atar and Zeitouni [2] use the Hilbert metric and the Birkhoff contraction coefficient to prove the exponential stability of finite state filters for ergodic signal processes. In [1], the Hilbert metric is used for extending the exponential stability for ergodic signals in infinite state spaces with mixing signal kernel. Under the additional assumption of compact state space it is shown that rate at which the error converges to zero increases without bound when the observations are made more accurate. LeGland and Oudjane [12] have also used the Hilbert metric and they relax the mixing property of the signal kernel by imposing the mixing condition on the kernel of the unnormalised filter process. Moreover, they study the effect of erroneous computation of the filter update in addition to the error due to an incorrect initialisation. Another account on the effect of model misspecifications, other than incorrect initial distribution, can be found in [3] where the cases of mixing signal kernel and bounded observation noise are considered by using the Hilbert metric. Del Moral and Guionnet [7] avoid using the Hilbert metric, and prove exponential stability of the filter in total variation under certain mixing assumptions imposed on the signal kernels.

Roughly speaking, the mixing property means that the signal itself is assumed to behave “well enough”. Recently, stability results have been obtained by relaxing the mixing assumption and instead assuming that the observations are sufficiently accurate. For an approach quite opposite to this, see [6], where the exponential stability is obtained even when the mixing assumption is relaxed and no assumptions on the observations are made. The signal is nevertheless assumed to be ergodic. Budhiraja and Ocone [4] prove exponential stability without signal ergodicity when the observations are bounded. In [5] also the boundedness of the observations is relaxed, and it is shown that the error due to incorrect initialisation converges exponentially fast to zero, provided that the observation error enters the observations with a sufficiently small coefficient. Similarly as in [14], this paper refines the results of [5] by giving explicit conditions on the observation noise coefficients and by showing that in certain cases the stability follows regardless of the observation noise coefficient. Moreover, the conditions in this work are somewhat different from those in [5]. LeGland and Oudjane [11] propose a method where the exact filter is approximated by a truncated filter. By using the Hilbert metric, this robust filter is shown to be stable under less stringent mixing assumptions for sufficiently accurate observations. Because the robust filter approximates the exact filter uniformly, the exact filter is then shown to inherit the stability of the robust filter. Oudjane and Rubenthaler [14] also use the method of truncation but avoid the use of Hilbert metric by adopting the method of the Dobrushin ergodic coefficient as proposed in [7].

The remainder of this work is organised as follows. Section 2 introduces some general notational conventions. In Section 3, we specify the approximation of the

\(^\dagger\)The reason for the popularity of the Hilbert metric is its scaling invariance which enables one to ignore the normalisation term in the filter recursions. However, for two measures to be comparable in Hilbert metric, it is required that the measures are absolutely continuous with respect to each other with bounded Radon-Nikodým derivatives. This requirement is rather restrictive and rules out many practical problems, e.g., those that are involved with Gaussian distributions.
discrete time filter by a truncated filter in a general setting where the assumptions (A1) . . . (A4) are not required to hold. Moreover, the section provides almost sure upper bounds for the approximation errors due to truncation and for the error between two truncated filters with different initial distributions. In Section 4, we impose the assumptions (A1) . . . (A4) on the filter application and give the proof for Theorem 1.1.

2. Notations and Preliminaries

Let \((S, \mathcal{S})\) be an arbitrary measurable space. The space of bounded, measurable functions \(\varphi : S \rightarrow \mathbb{R}\) is denoted by \(B(S)\) and it is endowed with the distance induced by the supremum norm
\[
\|\varphi\|_{\infty} \triangleq \sup_{x \in S} |\varphi(x)|.
\]
The space \(\mathbb{R}^d\) is endowed with the distance induced by the Euclidean norm denoted by \(\|\cdot\|\). For all \(\varphi \in B(S)\) and for all measures \(\mu\) on \(S\) we define the notation
\[
\mu(\varphi) \triangleq \int \varphi \, d\mu.
\]
When confusion will not arise, the parentheses can be omitted in the notation. The conventional notation \(\int \varphi(x) \, dx\) is always regarded as the Lebesgue integral with respect to the Lebesgue measure.

The space \(\mathcal{P}(S)\) of probability measures is endowed with the total variation distance which is defined for all \(\mu, \nu \in \mathcal{P}(S)\) as
\[
\|\mu - \nu\|_{\text{TV}} \triangleq \sup_{A \in \mathcal{S}} |\mu(A) - \nu(A)|
\]
\[
= \sup_{\|\varphi\|_{\infty} \leq 1} \|\mu \varphi - \nu \varphi\| = \frac{1}{2} \sup_{\|\varphi\|_{\infty} \leq 1} \|\mu \varphi - \nu \varphi\|.
\]
Note that in some references the total variation distance is defined to be twice this quantity.

A mapping \(K : S \times S \rightarrow [0, 1]\) is called a Markov kernel, if it satisfies the following two conditions:

(1) for all \(x \in S\), \(K(x, \cdot) \in \mathcal{P}(S)\),

(2) for all \(A \in \mathcal{S}\), \(K(\cdot, A)\) is measurable.

For all Markov kernels \(K\) we define a mapping \(\varphi \in B(S) \mapsto K(\varphi) \in B(S)\) as
\[
K(\varphi)(x) \triangleq \int \varphi(y) K(x, dy),
\]
and a mapping \(\mu \in \mathcal{P}(S) \mapsto \mu K \in \mathcal{P}(S)\) such that for all \(\varphi \in B(S)\),
\[
(\mu K)(\varphi) = \left[ \int \varphi(y) K(x, dy) \right] \mu(dx).
\]
The Dobrushin ergodic coefficient \([8]\) of a kernel \(K : S \times S \rightarrow [0, 1]\) is denoted by \(\alpha_S(K)\) and it is defined as
\[
\alpha_S(K) \triangleq 1 - \sup_{x,y \in S} \{ K(x, A) - K(y, A) \},
\]
By the definition of $\alpha_S(K)$, the quantity $1 - \alpha_S(K)$ is the contraction coefficient of the mapping $\mu \mapsto \mu K$ in total variation distance.

For all $\mu$-integrable functions $\psi \geq 0$ such that $\mu(\psi) > 0$, we define the mapping $\mu \in \mathcal{P}(S) \mapsto \psi \cdot \mu \in \mathcal{P}(S)$ such that for all $\varphi \in B(S)$,

$$(\psi \cdot \mu)(\varphi) = \frac{\mu(\psi \varphi)}{\mu(\psi)}.$$ 

This mapping is also known as the projective product [12]. For the projective product, we have the following elementary result for which a slightly different proof can be found in [13, page 40].

**Lemma 2.1.** For all $\mu, \nu \in \mathcal{P}(S)$ and for all $\mu$- and $\nu$-integrable functions $\psi : S \to [0, \infty)$ such that $\mu(\psi) > 0$,

$$\|\psi \cdot \mu - \psi \cdot \nu\|_{TV} \leq \frac{\|\psi\|_{\infty}}{\mu(\psi)} \|\mu - \nu\|_{TV}.$$ 

**Proof.** Let $A = \{\varphi \in B(S) | \|\varphi\|_{\infty} \leq 1, \varphi \geq 0\}$. If $\varphi \in A$, then $1 - \varphi \in A$. Thus, for all $\mu, \nu \in \mathcal{P}(S)$,

$$\|\mu - \nu\|_{TV} = \sup_{\varphi \in A} (\mu \varphi - \nu \varphi) = \sup_{\varphi \in A} (\nu \varphi - \mu \varphi).$$

If $\mu(\psi)/\nu(\psi) \geq 1$, then

$$\|\psi \cdot \mu - \psi \cdot \nu\|_{TV} = \frac{\|\psi\|_{\infty}}{\mu(\psi)} \sup_{\varphi \in A} \left( \frac{\mu(\psi \varphi)}{\|\psi\|_{\infty}} \frac{\nu(\psi \varphi)}{\|\psi\|_{\infty}} \right) \leq \frac{\|\psi\|_{\infty}}{\mu(\psi)} \sup_{\varphi \in A} \left( \frac{\mu(\psi \varphi)}{\|\psi\|_{\infty}} - \frac{\nu(\psi \varphi)}{\|\psi\|_{\infty}} \right) = \frac{\|\psi\|_{\infty}}{\mu(\psi)} \|\mu - \nu\|_{TV}.$$ 

The case $\mu(\psi)/\nu(\psi) \leq 1$ follows from symmetry. \qed

3. **Filter Approximation by Truncation**

In this section, we will specify in detail what is meant by the truncated filter and provide almost sure upper bounds for the error due to truncation and for the error between two truncated filters with different initial distributions. It should be noted that the discussion in this section applies to the general filter framework, that is, the assumptions (A1)…(A4) are not required to hold.

Let $K_i : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ where $i > 0$, be the Markov kernels of $(X_i)_{i \geq 0}$ such that for all $A \in \mathcal{B}(\mathbb{R}^d)$, $i > 0$

$$K_i(X_{i-1}, A) = P(X_i \in A | X_{i-1}), \quad \text{P-a.s.},$$

and let $k_i(x_{i-1}, \cdot)$ denote the density of $K_i(x_{i-1}, \cdot)$ with respect to $\lambda_d$. In this case, $k_i(x_{i-1}, x_i) = \rho_{W_i}(x_i - f_i(x_{i-1}))$. We can then define a recursion

$$\pi_i \triangleq g_i \cdot \pi_{i-1} K_i,$$

where $i > 0$, $\pi_0 \in \mathcal{P}(\mathcal{B}(\mathbb{R}^d))$ and $g_i(x) \triangleq \rho_{Y_i}(Y_i - h_i(x))$. If $\pi_0 = \mathcal{P}_0$, then $(\pi_i)_{i \geq 0} = (\pi_i^*)_{i \geq 0}$, i.e. the recursion yields the filter process. In the stability analysis, however, we are interested in the filter process with arbitrary initial distributions and therefore we allow the slight abuse of terminology and refer to $(\pi_i)_{i \geq 0}$ with any initial distribution $\pi_0 \in \mathcal{P}(\mathcal{B}(\mathbb{R}^d))$ as the exact filter.
The truncated approximation $(\pi^\Delta_i)_{i \geq 0}$ of the exact filter $(\pi_i)_{i \geq 0}$ is defined by the recursion

$$
\pi^\Delta_i \triangleq g_i^\Delta \cdot \pi^\Delta_{i-1} K_i, 
$$

where $i > 0$, $\pi^\Delta_0 = \pi_0$, $g_i^\Delta \triangleq 1_{C_i(\Delta)} g_i$ and for all $\Delta > 0$,

$$
C_i(\Delta) \triangleq \begin{cases} 
\{ x \in \mathbb{R}^d \mid \| Y_i - h_i(x) \| \leq \Delta \}, & i > 0 \\
\{ x \in \mathbb{R}^d \mid \| X_0 - x \| \leq b\Delta \}, & i = 0.
\end{cases}
$$

By defining $C_i(\infty) \triangleq \mathbb{R}^d$, for all $i \geq 0$, we observe that when $\Delta = \infty$, (3.1) yields the exact filter. For all $i > 0$, the mapping $\pi^\Delta_{i-1} \mapsto \pi^\Delta_i$ is denoted by $Q^\Delta_i$ and also we define $Q_i \triangleq Q^\Delta_i$. Moreover, we define for all $i \geq j > 0$, $Q^\Delta_{j,i} \triangleq Q^\Delta_{j,i} \circ \cdots \circ Q^\Delta_{i,i}$ and if $i < j$, the mapping $Q^\Delta_{j,i}$ is defined to be identity. Then we can define for all $i \geq j \geq 0$

$$
\pi^\Delta_{i,j} \triangleq Q^\Delta_{i+1,j}(Q_{1,j}(\pi_0)).
$$

Accordingly, $\pi^\Delta_{i,0} = \pi^\Delta_i$ and $\pi^\Delta_{j,j} = \pi_j$.

Let us also define for all $\Delta \in (0, \infty]$ and $i \geq j > 0$ the mapping $S^\Delta_{j,i} : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ as

$$
S^\Delta_{j,i}(x, A) \triangleq \frac{K_j(g^\Delta_{j+1} K_{j+1}(g^\Delta_i) \cdots K_i(g^\Delta_0))(x)}{K_j(g^\Delta_{j+1} K_{j+1}(g^\Delta_i) \cdots K_i(g^\Delta_0))}(x)
$$

and the mapping $\psi^\Delta_{j,i} : \mathbb{R}^d \to \mathbb{R}$ as

$$
\psi^\Delta_{j,i}(x) \triangleq \begin{cases} 
K_j(g^\Delta_{j+1} K_{j+1}(g^\Delta_i) \cdots K_i(g^\Delta_0))(x) & i \geq j \\
1 & i < j.
\end{cases}
$$

Also we define for all $i > 0$

$$
\tilde{\alpha}_i(\Delta) \triangleq \| k_i \|_\infty^{-1} \inf_{y \in C_i(\Delta)} \sup_{x \in C_{i-1}(\Delta)} k_i(x, y).
$$

According to these definitions, we can now obtain an upper bound for the error between two exact or truncated filters initialised with different distributions according to the following proposition. The principle of the proof was proposed in [14] and it is based on [7]. Note that throughout the remainder of this work, products over empty sets are taken to be equal to one.

**Lemma 3.1.** For all $\Delta \in (0, \infty]$, $i \geq j > 0$ and $\pi, \bar{\pi} \in \mathcal{P}(\mathcal{B}(\mathbb{R}^d))$

$$
\| Q^\Delta_{j,i}(\pi) - Q^\Delta_{j,i}(\bar{\pi}) \|_{TV} \leq \prod_{n=j+1}^i (1 - \tilde{\alpha}_n(\Delta)) \|\psi^\Delta_{j,i} \cdot \pi - \psi^\Delta_{j,i} \cdot \bar{\pi}\|_{TV}.
$$

**Proof.** For all $i \geq j > 0$

$$
Q^\Delta_{j,i}(\pi)(\varphi) = \frac{\pi(K_j(g^\Delta_j \cdots K_i(g^\Delta_0)\varphi))}{\pi(K_j(g^\Delta_j \cdots K_i(g^\Delta_0)))}
= \frac{1}{\pi(\psi^\Delta_{j,i})} \pi(\psi^\Delta_{j,i} K_j(g^\Delta_j \psi^\Delta_{j+1,i} \cdots K_i(g^\Delta_i \psi^\Delta_{i+1,i} \varphi)))(3.2)
$$
The substitution of
\[ S_{j,i}^\Delta(\varphi)(x) = K_j(g_j^\Delta \psi_{j+1,i}^\Delta)(x)/\psi_{j,i}^\Delta(x), \]
into (3.2) yields
\[ Q_j^\Delta(\pi)(\varphi) = \frac{1}{\pi(\psi_{j,i}^\Delta)} \pi(\psi_{j,i}^\Delta S_j^\Delta_1(\cdots S_j^\Delta_{n-1}(\varphi) )))
= ((\psi_{j,i}^\Delta), \pi) S_j^\Delta_1 S_j^\Delta_2 \cdots S_j^\Delta_{n-1})(\varphi). \quad (3.3) \]
It is also observed that for all \( i \geq j > 0, x \in C_{j-1}(\Delta) \) and \( A \in \mathcal{B}(\mathbb{R}^d) \)
\[ S_{j,i}^\Delta(x, A) = \frac{K_j(g_j^\Delta \psi_{j+1,i}^\Delta, 1_A)(x)}{K_j(g_j^\Delta \psi_{j+1,i}^\Delta)(x)} \leq \lambda_{d_i}(g_j^\Delta \psi_{j+1,i}^\Delta) \lambda_{d_j}(g_j^\Delta \psi_{j+1,i}^\Delta) \inf_{x_{j-1} \in C_{j-1}(\Delta)} \sup_{x_{j-1} \in C_{j-1}(\Delta)} \frac{\lambda_{d_i}(g_j^\Delta \psi_{j+1,i}^\Delta)}{\lambda_{d_j}(g_j^\Delta \psi_{j+1,i}^\Delta)} \]
and thus
\[ \alpha_{C_{j-1}(\Delta)}(S_{j,i}^\Delta) = \inf \sum_{n=1}^M \min \{ S_{j,j}^\Delta(x, A_n), S_{j,j}^\Delta(y, A_n) \} \geq \tilde{\alpha}_j(\Delta), \quad (3.4) \]
where the first equality is equivalent to the definition of the Dobrushin ergodic coefficient (see Eq. (1.16) and Eq. (1.5) in [8] and Section 3.2 in [9], see also Eq. (6) in [7]). The infimum is taken over all \( x, y \in C_{j-1}(\Delta) \) and all \( M \) set partitions of \( \mathbb{R}^d \). It is observed that all the kernels \( S_{j,i}^n \), where \( j + 1 \leq n \leq i \) are applied to probability measures on the set \( C_{n-1}(\Delta) \). Therefore, the claim follows from (3.3) and (3.4), because (see [8])
\[ \|Q_j^\Delta(\pi) - Q_j^\Delta(\tilde{\pi})\|_{TV} = \|((\psi_{j,i}^\Delta, \pi) S_j^\Delta_1 \cdots S_j^\Delta_{n-1} - (\psi_{j,i}^\Delta, \tilde{\pi}) S_j^\Delta_1 \cdots S_j^\Delta_{n-1})\|_{TV} \]
\[ \leq (1 - \alpha_{Rn}(S_{j,i}^\Delta)) \prod_{n=j+1}^i (1 - \alpha_{C_{n-1}(\Delta)}(S_{j,i}^\Delta)) \|\psi_{j,i}^\Delta, \pi - \psi_{j,i}^\Delta, \tilde{\pi}\|_{TV} \]
\[ \leq \prod_{n=j+1}^i (1 - \tilde{\alpha}_n(\Delta)) \|\psi_{j,i}^\Delta, \pi - \psi_{j,i}^\Delta, \tilde{\pi}\|_{TV}. \]

The following result establishes an almost sure upper bound for the error of the truncated approximation in terms of the contractivity of the Markov kernels \( S_{j,i}^\Delta \), and the local errors \( \|\pi_{i+1}^\Delta - \pi_{i+1}^\Delta\|_{TV} \), i.e., the errors due to only one truncation. The proof follows closely the principles proposed in [14].

**Proposition 3.2.** For all \( \pi_0 \in \mathcal{P}(\mathcal{B}(\mathbb{R}^d)), \Delta > 0 \) and \( i > 0 \)
\[ \|\pi_i - \pi_i^\Delta\|_{TV} \leq \sum_{j=1}^i \prod_{n=j+1}^i (1 - \tilde{\alpha}_n(\Delta)) \min \left( 1, \frac{\|\pi_{i,j} - \pi_{i,j-1}^\Delta\|_{TV}}{\tilde{\alpha}_{j+1}(\Delta)} \right). \quad (3.5) \]
STABILITY OF THE DISCRETE TIME FILTER

Figure 1. Exact filter and its approximation by truncation.

Proof. By the triangle inequality (see Figure 1), Lemma 3.1, and Lemma 2.1,

\[
\|\pi_i - \pi_i^\Delta\|_{TV} \leq \sum_{j=1}^{i} \|\pi_{i,j} - \pi_{i,j-1}\|_{TV} \\
\leq \sum_{j=1}^{i} \prod_{n=j+2}^{i} (1 - \alpha_n(\Delta)) \|\psi_{j+1,i}^\Delta \cdot \pi_{j,j} - \psi_{j+1,i}^\Delta \cdot \pi_{j,j-1}\|_{TV} \\
\leq \sum_{j=1}^{i} \prod_{n=j+2}^{i} (1 - \alpha_n(\Delta)) \min \left(1, \frac{\|\psi_{j+1,i}^\Delta\|_{\infty}}{\pi_{j,j-1}(\psi_{j+1,i}^\Delta)} \right). \tag{3.6}
\]

In the last inequality, we have used the fact that total variation takes values in [0,1]. It is observed that \(\pi_{j,j-1} = 1_{C_j(\Delta)} \cdot \pi_{j,j}^\Delta\), and hence for all \(i \geq j > 0\)

\[
\psi_{j,j-1}(x) = \frac{K_{j+1}(g_{j+1}^\Delta \psi_{j+2,i}^\Delta)(x)}{\pi_{j,j-1}(1_{C_j(\Delta)} \cdot K_{j+1}(g_{j+1}^\Delta \psi_{j+2,i}^\Delta))} \\
\leq \inf_{x_{j+1} \in C_{j+1}(\Delta)} \frac{\lambda_{j+1}(g_{j+1}^\Delta \psi_{j+2,i}^\Delta)}{k_{j+1}(x_j, x_{j+1}) \pi_{j,j-1}(\lambda_{j+1}(g_{j+1}^\Delta \psi_{j+2,i}^\Delta))} \\
= \frac{1}{\alpha_{j+1}(\Delta)}. \tag{3.7}
\]

The substitution of this approximation into (3.6) then yields the claim. \(\square\)

From Proposition 3.2 we can obtain the following simpler upper bound for the error due to truncation. Although this upper bound is looser than (3.5), it will be sufficient for the proof of stability.

Corollary 3.3. For all \(\pi_0 \in \mathcal{P}(\mathcal{B}(\mathbb{R}^d)), \Delta > 0, i > 0\)

\[
\|\pi_i - \pi_i^\Delta\|_{TV} \leq \sum_{j=1}^{i} \frac{\|\psi_{j,j-1} - \pi_{j,j-1}\|_{TV}}{\alpha_{j+1}(\Delta)}. \tag{3.8}
\]

Proof. Follows directly from Proposition 3.2, since \((1 - \alpha_j(\Delta)) \in [0,1]\). \(\square\)
Let us conclude this section by the following corollary of Lemma 3.1 which establishes an almost sure upper bound for the distance between two truncated filters with different initial distributions.

**Corollary 3.4.** For all \( \pi_0, \bar{\pi}_0 \in \mathcal{P}(\mathcal{B}(\mathbb{R}^d)) \), \( \Delta > 0 \), and \( i \geq 0 \)

\[
\|\pi_i^\Delta - \bar{\pi}_i^\Delta\|_{TV} \leq \prod_{j=2}^{i} (1 - \alpha_j(\Delta)).
\] (3.9)

**Proof.** Follows directly from Lemma 3.1, since by definition \( \pi_i^\Delta = Q_{i,i}^\Delta(\pi_0) \) and \( \bar{\pi}_i^\Delta = Q_{i,i}^\Delta(\bar{\pi}_0) \), and \( \|\psi_{1,i}^\Delta \cdot \pi_0 - \psi_{1,i}^\Delta \cdot \bar{\pi}_0\|_{TV} \in [0,1] \). \( \square \)

It should be noted that all terms of the product in (3.9) are subunitary, hence the upper bound in (3.9) cannot increase over time.

**4. Proof of the Main Result**

Before going into the detailed proof of Theorem 1.1, let us explain the intuition behind the reasoning. To prove the stability, we take the truncated approximations \( (\pi_i^\Delta)_{i \geq 0} \) and \( (\bar{\pi}_i^\Delta)_{i \geq 0} \) of two exact filters \( (\pi_i)_{i \geq 0} \) and \( (\bar{\pi}_i)_{i \geq 0} \) with different initial distributions \( \pi_0, \bar{\pi}_0 \in \mathcal{P}(\mathcal{B}(\mathbb{R}^d)) \). In this case, the error due to different initialisations can be written by the triangle inequality as

\[
\|\pi_i - \bar{\pi}_i\|_{TV} \leq \|\pi_i - \pi_i^\Delta\|_{TV} + \|\pi_i^\Delta - \bar{\pi}_i^\Delta\|_{TV} + \|\bar{\pi}_i^\Delta - \bar{\pi}_i\|_{TV}. \] (4.1)

It would be sufficient for the stability that all terms on the right hand side converge to zero as \( i \to \infty \). However, it follows from our analysis that the truncation does not yield a uniformly convergent approximation of the exact filter as \( \Delta \to \infty \) and therefore, the first and the last term on the right hand side of (4.1) do not converge to zero as \( i \to \infty \). Uniform, in this case, means uniform in time. For this reason, we write

\[
\|\pi_i - \bar{\pi}_i\|_{TV} \leq \|\pi_i - \pi_i^\Delta\|_{TV} + \|\pi_i^\Delta - \bar{\pi}_i^\Delta\|_{TV} + \|\bar{\pi}_i^\Delta - \bar{\pi}_i\|_{TV}, \] (4.2)

i.e. we write the upper bound for the error in terms of truncated filters with different truncation radii \( \Delta_i \). In this case, the first and the last terms on the right hand side can be made to converge to zero by letting \( (\Delta_i)_{i \geq 0} \) be a sufficiently fast increasing sequence. It should be emphasised that by definition, the truncated filter \( (\pi_i^\Delta)_{i \geq 0} \) satisfies

\[
\pi_i^\Delta = Q_{i,i}^\Delta(\cdots(Q_{2,2}^\Delta(\cdots(Q_{1,1}^\Delta(\pi_0))))),
\]

that is, the same truncation radius \( \Delta_i \) is used up to time \( i \). We do not allow a filter, where different radii would be used at consecutive steps, i.e.

\[
\pi_i^\Delta = Q_{i,i}^\Delta(\cdots(Q_{2,2}^\Delta(\cdots(Q_{1,1}^\Delta(\pi_0))))).
\]

The use of the upper bound in (4.2) has a downside as well. While the first and the last term converge to zero when \( (\Delta_i)_{i \geq 0} \) is a sufficiently fast increasing sequence, the convergence is not guaranteed for the middle term and it may well become divergent if \( (\Delta_i)_{i \geq 0} \) increases too rapidly. Therefore the fundamental issue in proving the stability is to show that there exists a rate for increasing \( (\Delta_i)_{i \geq 0} \) such that all three terms in (4.2) converge to zero as \( i \to \infty \).
Let us then turn into the details of the proof of Theorem 1.1. According to (3.4), for all \( \Delta > 0 \), and \( i \geq j > 0 \), we have \( \alpha_{C_{j-1}(\Delta)}(S^\Delta_{C_{j-1}(\Delta)}) \geq \alpha_j(\Delta) \). Under the assumptions (A1), (A2) and (A3), a more explicit lower bound for \( \tilde{\alpha}_j(\Delta) \), and thus for \( \alpha_{C_{j-1}(\Delta)}(S^\Delta_{C_{j-1}(\Delta)}) \), can be obtained. For this purpose, we define for all \( i \geq 0 \), \( \Delta > 0 \)

\[
\varepsilon_i(\Delta) \triangleq \frac{m_i}{M_1} \exp\left(-\alpha_1((ab + b)\Delta + \xi_i)^{\beta_i}\right),
\]

(4.3)

where

\[
\xi_i \triangleq \begin{cases} 
  b\|V_i\| + \|W_i\| + ab\|V_{i-1}\|, & i > 1 \\
  b\|V_i\| + \|W_i\|, & i = 1.
\end{cases}
\]

(4.4)

**Lemma 4.1.** For all \( \Delta > 0 \), \( i > 0 \), one has \( \tilde{\alpha}_i(\Delta) \geq \varepsilon_i(\Delta) \) \( \mathcal{P} \)-a.s.

**Proof.** From the definition of \( \tilde{\alpha}_i(\Delta) \) and (A3) it follows that

\[
\tilde{\alpha}_i(\Delta) = \|k_i\|_{\infty}^{-1} \inf_{W_i} \rho_{W_i}(x_i - f_i(x_{i-1})) \\
\geq \frac{m_i}{M_1} \inf \exp\left(-\alpha_1 \|x_i - f_i(x_{i-1})\|^{\beta_i}\right),
\]

(4.5)

where the infima are taken over all \( (x_{i-1}, x_i) \in C_{i-1}(\Delta) \times C_i(\Delta) \). In this set, as pointed out in [14], one has for all \( i > 1 \)

\[
\|x_i - f_i(x_{i-1})\| \leq \|x_i - h_{i-1}^{-1}(Y_i)\| + \|f_i(h_{i-1}^{-1}(Y_{i-1})) - f_i(x_{i-1})\| \\
+ \|h_{i-1}^{-1}(Y_{i-1}) - f_i(x_{i-1}) - W_i\| \\
\leq (ab + b)\Delta + \|h_{i-1}^{-1}(Y_i) - f_i(x_{i-1}) - W_i\| \\
+ \|f_i(X_{i-1}) + W_i - f_i(h_{i-1}^{-1}(Y_{i-1}))\| \\
\leq (ab + b)\Delta + \|h_{i-1}^{-1}(Y_{i-1}) - f_i(x_{i-1}) - W_i\| \\
+ \|f_i(h_{i-1}^{-1}(Y_{i-1}) - V_{i-1}) - f_i(h_{i-1}^{-1}(Y_{i-1}))\| + \|W_i\| \\
\leq (ab + b)\Delta + \xi_i.
\]

By substituting \( h_{i-1}^{-1}(Y_{i-1}) \) above with \( X_0 \), it can be shown similarly that

\[
\|x_i - f_i(x_{i-1})\| \leq (ab + b)\Delta + \xi_i
\]

holds also for \( i = 1 \) and hence for all \( i > 0 \). The claim then follows by substituting this approximation into (4.5).

As suggested by Corollary 3.3, bounding the error due to truncation can be done by bounding the local errors \( \|\pi^\Delta_{C_{j-1}} - \pi^\Delta_{C_{j-1}}\|_{TV} \) from above, and the ergodic coefficients \( \tilde{\alpha}_i(\Delta) \) from below. It is apparent from (4.3) and Lemma 4.1 that bounding the ergodic coefficients from below can be done by bounding the sequence \( (\xi_i)_{i > 0} \) from above. It also turns out that by bounding \( (\xi_i)_{i > 0} \), one can find an upper bound for the local errors and hence for the error of the truncated approximation. According to the definition of \( \xi_i \), a uniform upper bound for \( \xi_i \) does not exist. However, because of the assumptions (A3) and (A4), one can find an increasing almost sure upper bound for \( \xi_i \) as stated by the following proposition.

**Proposition 4.2.** For all \( \epsilon \in (0, \min(\alpha_1, \alpha_2)) \), there are positive random variables \( c_1 = c_1(\epsilon) \) and \( c_2 = c_2(\epsilon) \) such that for all \( i > 0 \), one has \( \xi_i \leq \xi_{i,\epsilon} \), \( \mathcal{P} \)-a.s.,
where

\[ q > 0 \quad \text{and} \quad \epsilon > 0 \text{ respectively.} \]

Hence, for all \( \tilde{c} \) random variable

This implies that for all \( \epsilon \leq 1 \), \( \tilde{c} \)

By setting \( q = (\alpha_1 - \epsilon^2)/(\alpha_1 - \epsilon) \), \( \epsilon = \epsilon/2 \) and \( c_1 = \ln c_3/((\alpha_1 - \epsilon^2)/2) \), we have for all \( \epsilon \in (0, \alpha_1) \)

The same reasoning applies to \( \| V_i \| \), and the claim then follows by substituting the resulting upper bounds of \( \| W_i \| \) and \( \| V_i \| \) into (4.4).

By using the upper bound for \( (\xi_i)_{i>0} \) we can now obtain an almost sure upper bound for the truncation error according to the following proposition. The proposition can be considered as a restatement of Corollary 3.3 under the assumptions (A3) and (A4). In the proof of the proposition we use the elementary property of total variation norm that for all \( \mu \in \mathcal{P}(\mathcal{S}) \) and \( D \in \mathcal{S} \), one has \( \| \mu - 1_D \cdot \mu \|_{TV} = \mu(D) \) where \( D \) is the complement of \( D \). Therefore, for all \( i > 0 \), one has

\[ \| \pi_{i,i} - \pi_{i,i-1} \|_{TV} = \pi_i(C_i(D)). \] (4.6)

**Proposition 4.3.** For all \( \epsilon \in (0, \min(\alpha_1, \alpha_2)) \) and \( \Delta^* > 0 \), there exists a random variable \( c_4 = c_4(\epsilon, \Delta^*) > 0 \), such that for all \( \pi_0 \in \mathcal{P}(\mathcal{B}(\mathbb{R}^{d_i})), \Delta \geq \Delta^* \) and \( i > 0 \)

\[ \| \pi_i - \pi_i^\Delta \|_{TV} \leq c_4 \sum_{j=1}^{i} \exp\left((\alpha_2 - \epsilon)\Delta_{j+1} + \frac{2\alpha_1((ab + b)\Delta + \xi_{j+1,i}^e + \xi_{j+1,i}^\Delta)\Delta_{j+1}}{1 - \pi_{j-1}(C_{j-1}(\Delta))}\right). \] (4.7)
Proof. Let $\tilde{B} = \{y \in \mathbb{R}^d \mid \|y\| > \Delta\}$. For all $i > 0$ and $t > 1$

$$
\pi_i(C_i(\Delta)) = \int \left[ \int_{C_i(\Delta)} g_i(x)k_i(x_{i-1}, x_i) \, dx_i \right] \pi_{i-1}(dx_{i-1})
$$

$$
\leq \frac{\int \left[ \int_{C_i(\Delta)} g_i(x)k_i(x_{i-1}, x_i) \, dx_i \right] \pi_{i-1}(dx_{i-1})}{\|k_i\|_{\infty} \int_{C_i(\Delta)} \rho_V(Y_i - h_i(x_i)) \, dx_i}
$$

$$
\leq \frac{d_i b \int_{\tilde{B}} \rho_V(x) \, dx}{\tilde{a}_i(\Delta) \inf_{x \in C_i(\Delta)} g_i(x) \lambda_d(C_i(\Delta/t)) \pi_{i-1}(C_i(\Delta))}
$$

$$
\leq \frac{M_2 d_i b \lambda_d \int_{\Delta} x^{d-1} \exp\left(-\alpha_2 x^{\beta_2}\right) \, dx}{\tilde{a}_i(\Delta) \inf_{x \in C_i(\Delta)} g_i(x) \lambda_d(C_i(\Delta/t)) \pi_{i-1}(C_i(\Delta))}, \tag{4.8}
$$

Let $B(x, r)$ denote a ball of radius $r$ centered at $x$ in $\mathbb{R}^d$. For all $x \in B(h_i^{-1}(Y_i), \Delta/b)$,

$$
\frac{1}{b} \|h_i(x) - Y_i\| \leq \|x - h_i^{-1}(Y_i)\| \leq \frac{\Delta}{b},
$$

where the first inequality follows from the assumption (A2). This implies that $B(h_i^{-1}(Y_i), \Delta/b) \subset C_i(\Delta)$. Therefore for all $i > 0$

$$
\lambda_d(B(0, \Delta^*/b)) \leq \lambda_d(B(h_i^{-1}(Y_i), \Delta/b)) \leq \lambda_d(C_i(\Delta)). \tag{4.9}
$$

It is also observed that for all $\tilde{\epsilon} \in (0, \alpha_2)$, there is $c_5 = c_5(\tilde{\epsilon}, \Delta^*) > 0$ such that

$$
\int_{\Delta} x^{d-1} \exp\left(-\alpha_2 x^{\beta_2}\right) \, dx \leq c_5 \exp\left(-\alpha_2 \tilde{\epsilon} \Delta^{\beta_2}\right). \tag{4.10}
$$

By setting $\tilde{\epsilon} = \epsilon/2$ and $t = (2\alpha_2/\epsilon)^{1/\beta_2}$, where $\epsilon \in (0, \alpha_2)$, we have, according to (4.8), (4.9), and (4.10)

$$
\pi_i(C_i(\Delta)) \leq \frac{M_1 M_2 d_i b \lambda_d \int_{\Delta} x^{d-1} \exp\left(-\alpha_2 x^{\beta_2}\right) \, dx}{m_1 m_2 \lambda_d(B(0, \Delta^*/b))(1 - \pi_{i-1}(C_{i-1}(\Delta)))}. \tag{4.11}
$$

According to (4.6), the claim then follows directly with

$$
c_4 = \frac{M_1^2 M_2 d_i b \lambda_d \int_{\Delta} x^{d-1} \exp\left(-\alpha_2 x^{\beta_2}\right) \, dx}{m_1^2 m_2 \lambda_d(B(0, \Delta^*/b))}.
$$

by substituting (4.3) and (4.11) into (3.8) and by substituting $\xi_i$ and $\xi_{i+1}$ with their common upper bound $\xi_{i+1, \tilde{\epsilon}}$. \qed

According to Proposition 4.3, it is sufficient for

$$
\lim_{i \to \infty} \|\pi_i - \pi_{i+1, \tilde{\epsilon}}\|_{TV} = 0 \tag{4.12}
$$

to choose $(\Delta_i)_{i \geq 0}$ such that the upper bound in (4.7) converges to zero as $i \to \infty$. It can be shown that for the convergence of this upper bound, it is necessary to have

$$
\lim_{i \to \infty} (-\alpha_2 + \epsilon)\Delta_i^{\beta_2} + 2\alpha_1 \Delta_i^{\beta_1} \left(ab + b + \frac{\xi_{i+1, \tilde{\epsilon}}}{\Delta_i}\right)^{\beta_1} = -\infty. \tag{4.13}
$$

Moreover, it can be shown that for (4.13) to hold when $\beta_2 \geq \beta_1$, it is sufficient that the ratio $\xi_{i+1, \tilde{\epsilon}}/\Delta_i$ is bounded, provided that $\alpha_2$ is sufficiently large\(^1\). Therefore it

\(^1\)In the case $\beta_1 = \beta_2$ the boundedness of the ratio is also necessary.
is observed that according to Proposition 4.2, if for all \( i > 1 \) we set \( \Delta_i^{\beta_1} = s \ln i \), where \( s > 0 \), then for all \( \epsilon \in (0, \min(\alpha_1, \alpha_2)) \)

\[
\lim_{i \to \infty} \frac{\xi_{i+1,\epsilon}}{\Delta_i} = \left\{ \begin{array}{ll}
\frac{ab + b}{s^{1/\beta_1}(\alpha_2 - \epsilon)^{1/\beta_2}} + \frac{1}{s^{1/\beta_1}(\alpha_1 - \epsilon)^{1/\beta_2}}, & \beta_2 = \beta_1 \\
\frac{1}{s^{1/\beta_1}(\alpha_1 - \epsilon)^{1/\beta_2}}, & \beta_2 > \beta_1,
\end{array} \right.
\](4.14)

and hence \( \xi_{i+1,\epsilon}/\Delta_i \) is bounded. Thus, in order to prove Theorem 1.1, it remains to prove that (4.13) is in fact a sufficient condition for (4.12) and that if \( \Delta_i^{\beta_1} = s \ln i \) for all \( i > 1 \), the middle term in (4.2) also converges to zero, i.e.

\[
\lim_{i \to \infty} \|\pi_i^{\Delta_i} - \#_i^{\Delta_i}\|_{TV} = 0.
\](4.15)

Let us first prove (4.15). For this purpose we choose \( d > 0 \) such that for all \( i > 1 \), \( P(\xi_i > d) < 1/4 \) and define

\[
\tilde{\epsilon}(\Delta) = \frac{m_1}{2M_1} \exp(-\alpha_1((ab + b)\Delta + d)^{\beta_1}).
\]

It can then be shown that for all \( \Delta, n, i > 0 \)

\[
E \left[ \prod_{j=1}^i (1 - \epsilon_{n+j}(\Delta)) \right] \leq (1 - \tilde{\epsilon}(\Delta))^{i-1},
\](4.16)

The proof of this inequality is given in [14] and for completeness, it is also included in Appendix A.

**Proposition 4.4.** For all \( i > 1 \), let \( \Delta_i^{\beta_1} = s \ln i \). If \( s < \alpha_1^{-1}(ab + b)^{-\beta_1} \), then for all \( \epsilon > 0 \), there is \( c_6 = c_6(s) > 0 \) and a positive random variable \( c_7 = c_7(\epsilon) \) such that for all \( i > 1 \)

\[
\|\pi_i^{\Delta_i} - \#_i^{\Delta_i}\|_{TV} \leq c_7\exp((-m_1/2M_1 + \epsilon)^{i}\epsilon^s), \quad \text{P-a.s.}
\](4.17)

**Proof.** By the inequality \( (1 - a) \leq \exp(-a) \), we have

\[
(1 - \tilde{\epsilon}(\Delta_i))^{i-2} \leq \exp\left(-(i - 2)\frac{m_1}{2M_1} \exp(-\alpha_1 s \ln i \left(ab + b + \frac{d}{\Delta_i}\right)^{\beta_1})\right) = \exp\left(-\frac{m_1}{2M_1}i^{1-\alpha_1 s(ab + b + \frac{d}{\Delta_i})^{\beta_1}} + \frac{m_1}{M_1}i^{1-\alpha_1 s(ab + b + \frac{d}{\Delta_i})^{\beta_1}}\right).
\](4.18)

Let us choose, \( c_6 = c_6(s) = 1/2 - \alpha_1 s(ab + b)^{\beta_1}/2 \) which is positive because \( s < \alpha_1^{-1}(ab + b)^{-\beta_1} \). In this case,

\[
\lim_{i \to \infty} 1 - \alpha_1 s \left(ab + b + \frac{d}{\Delta_i}\right)^{\beta_1} = 2c_6,
\](4.19)

where the convergence is from below. It is also observed that

\[
\lim_{i \to \infty} \exp\left(\frac{m_1}{M_1}i^{-\alpha_1 s(ab + b + \frac{d}{\Delta_i})^{\beta_1}}\right) = 1,
\](4.20)

where the convergence is from above. It then follows from (4.16), (4.18), (4.19) and (4.20), that there is \( c_8 = c_8(s) \) such that

\[
E \left[ \prod_{j=1}^{i-1} (1 - \epsilon_{j+1}(\Delta_i)) \right] \leq (1 - \tilde{\epsilon}(\Delta_i))^{i-2} \leq c_8\exp\left(-\frac{m_1}{2M_1}c_6\right),
\]
We can now give the proof of the main result. Thus, in turn, implies that there is an $I$ and $\pi$ such that for all $\epsilon > 0$ and $\alpha$ there is

$$\|\pi^{\Delta_i} - \pi^{\Delta_i}\|_\text{TV} \leq \prod_{j=2}^{i-1} (1 - \alpha_j(\Delta_i)) \leq \prod_{j=1}^{i-1} (1 - \epsilon_j(\Delta_i)),$$

which, according to (4.21), yields the claim.

Let us define for all $p \geq 0$, $\kappa_p \triangleq \gamma^{-1}(-p)$, where $\gamma : \mathbb{R}_+ \to \mathbb{R}$ is defined as

$$\gamma(x) \triangleq 2 \left( 2 + \frac{\alpha_1^1/\beta_1(ab + b)}{x^{1/\beta_1}} \right)^{\beta_1} + 1 - \frac{x}{\alpha_1(ab + b)^{\beta_1}}.$$

The constant $\kappa_p$ is well defined because $\gamma$ is strictly decreasing continuous function with

$$\lim_{x \to -0} \gamma(x) = -\lim_{x \to -\infty} \gamma(x) = \infty.$$

We can now give the proof of the main result.

**Proof of Theorem 1.1.** Let us first consider the case (i). Let $\beta = \beta_1 = \beta_2$ and

$$\theta_1(z) = \frac{-\alpha_2 + z}{(ab + b)^{\beta_1}(1 + z)} + 2 \left( \frac{1}{(1 + z)^{1/\beta_1}} + \frac{\alpha_1^{1/\beta_1}(ab + b)}{(\alpha_2 - z)^{1/\beta_1}} + \frac{\alpha_1^{1/\beta_1}}{(\alpha_1 - z)^{1/\beta_1}} \right)^{\beta_1}.$$

According to the assumption $\alpha_2 > \kappa_0$, there exists $p > 0$ such that $\alpha_2 > \kappa_p$ and thus

$$\lim_{z \to -0} \theta_1(z) = -\frac{\alpha_2}{\alpha_1(ab + b)^{\beta_1}} + 2 \left( \frac{2 + \alpha_1^{1/\beta_1}(ab + b)}{\alpha_2^{1/\beta_1}} \right)^{\beta_1} < -p - 1. \quad (4.22)$$

Therefore we can choose $\epsilon \in (0, \min(\alpha_1, \alpha_2))$ such that $\theta_1(\epsilon) < -p - 1$. Let us then set $s^{-1} = \alpha_1(ab + b)^{\beta_1}(1 + \epsilon)$ and $\Delta_i^0 = s \ln i$. It can then be shown that according to (4.14) and the definition of $\theta_1$, there exists $I_1 = I_1(\epsilon) \in \mathbb{N}$ such that for all $i \geq I_1$

$$\exp \left( (-\alpha_2 + \epsilon)\Delta_i^0 + 2\alpha_1\Delta_i^0 \left( ab + b + \frac{\xi_i + \epsilon}{\Delta_i} \right)^{\beta_1} \right) \leq x_{\beta_1}(\epsilon) \leq i^{-p - 1}. \quad (4.23)$$

This, in turn, implies that there is $I_2 = I_2(\epsilon, \pi_0) \in \mathbb{N}$ such that for all $i \geq I_2$

$$\frac{M_1M_2d_\epsilon b^{\delta_1}dr_{i,e,c} \exp \left( (-\alpha_2 + \epsilon)\Delta_i^0 + \alpha_1((ab + b)\Delta_i + \xi_i)^{\beta_1} \right)}{m_1m_2\lambda_{i,e}(B(0, \Delta_i^{+}/bt))} \leq \frac{\epsilon}{(1 + \epsilon)^2} \quad (4.24)$$

and $\pi_0(\bar{C}_j(\Delta_i)) < \epsilon/(1 + \epsilon)$. By substituting (4.24) into (4.11) and by using the fact that $\xi_i,e$ is increasing in $i$, it can be shown by induction that for all $i > I_2$ and $0 \leq j \leq i$ one has

$$\frac{1}{1 - \pi_j(\bar{C}_j(\Delta_i))} \leq 1 + \epsilon. \quad (4.25)$$
By substituting this approximation into (4.7) and by using (4.23), one has for sufficiently large \( i \)

\[
\|\pi_i - \pi_i^{\Delta}\|_{TV} \leq c_4(1 + \epsilon) \sum_{j=1}^{i} \exp\left((-\alpha_2 + \epsilon)\Delta_j^\beta + 2\alpha_1 \left((ab + b)\Delta_i + \xi_{i+1,\epsilon}\right)^\beta\right)
\]

\[
\leq c_4(1 + \epsilon) \sum_{j=1}^{i} \exp\left(-\alpha_2 + \epsilon\right)\Delta_j^\beta + 2\alpha_1 \left((ab + b)\Delta_i + \xi_{i+1,\epsilon}\right)^\beta
\]

\[
\leq c_4(1 + \epsilon)\epsilon^{-\alpha_2}.
\]

The same reasoning applies to \( \|\pi_i - \pi_i^{\Delta}\|_{TV} \) as well, with possibly different magnitude of \( i \) required for (4.26) to hold. This is because \( I_2 \) depends on the choice of the initial distribution \( \pi_0 \). Because \( s^{-1} > \alpha_1(ab + b)^{\beta_2} \), it follows from Proposition 4.4 that for all \( \epsilon > 0 \), there exists constants \( c_6 = c_6(s) > 0 \) and \( c_7 = c_7(\epsilon) > 0 \) such that

\[
\|\pi_i^{\Delta} - \pi_i^{\Delta}\|_{TV} \leq c_7\exp((-m_1/2M_1 + \epsilon_2)i^{\alpha_2}).
\]

In conclusion, it follows from (4.2), (4.26) and (4.27) that there exists a random constant \( c_9 = c_9(\pi_0, \pi_0, p) \) such that for all \( i > 0 \)

\[
\|\pi_i - \pi_i\|_{TV} \leq c_9\epsilon^{-p} + c_7\exp((-m_1/2M_1 + \epsilon_2)i^{\alpha_2}).
\]

Let us then consider the case (ii). Let \( \tilde{\alpha} = \lim_{\epsilon \to 0} \theta_2(z) \), where

\[
\theta_2(z) = \frac{\exp\left((-\alpha_2 + \epsilon)\Delta_z^\beta + 2\alpha_1 \left((ab + b)\Delta_i + \xi_{i+1,\epsilon}\right)^\beta\right)}{\alpha_2 - 2z^{\beta_2/\beta_1}(ab + b)^{\beta_2}(1 + z)^{\beta_2/\beta_1}}.
\]

Let \( \epsilon_1 > 0 \) be arbitrary and take \( \epsilon \in (0, \min(\alpha_1, \alpha_2)) \) such that \( \theta_2(\epsilon) > \tilde{\alpha} - \epsilon_1 \). Let \( s \) and \( \Delta_i \) be defined similarly as above. From (4.14) it then follows that there is \( I_3 = I_3(\epsilon) \in \mathbb{N} \) such that for all \( i \geq I_3 \)

\[
\exp\left(-\alpha_2 + \epsilon\right)\Delta_i^\beta + 2\alpha_1 \left((ab + b)\Delta_i + \xi_{i+1,\epsilon}\right)^\beta \leq \exp\left(-\theta_2(\epsilon)(\ln i)^{\beta_2/\beta_1}\right).
\]

Similarly as above, (4.25) can be shown to hold for all \( 0 \leq j \leq i \), where \( i \) is sufficiently large and thus

\[
\|\pi_i - \pi_i^{\Delta}\|_{TV} \leq c_4(1 + \epsilon)\epsilon^{-\alpha_2} \exp\left((-\tilde{\alpha} + \epsilon_1)(\ln i)^{\beta_2/\beta_1}\right),
\]

for sufficiently large \( i \). Again, according to Proposition 4.4, we conclude that for all \( \epsilon_1, \epsilon_2 > 0 \) there exist constants \( c_6 = c_6(s) \), \( c_7 = c_7(\epsilon_2) \) and \( c_{10} = c_{10}(\pi_0, \pi_0, \epsilon_1) \) such that for all \( i > 0 \)

\[
\|\pi_i - \pi_i\|_{TV} \leq c_{10}\exp\left((-\tilde{\alpha} + \epsilon_1)(\ln i)^{\beta_2/\beta_1}\right) + c_7\exp((-m_1/2M_1 + \epsilon_2)i^{\alpha_2}).
\]

Equations (4.28) and (4.29) give directly the following corollary which establishes lower bounds for the rates at which the initialisation error converges to zero.

**Corollary 4.5.**

(i) If \( \beta_1 = \beta_2 \), then for all \( \pi_0, \tilde{\pi}_0 \in \mathcal{P}(\mathbb{B}(\mathbb{R}^d)) \), \( p > 0 \) and \( \alpha_2 > \kappa_p \) there exists random variable \( c_{11} \) such that for all \( i > 0 \),

\[
\|\pi_i - \tilde{\pi}_i\|_{TV} \leq c_{11}i^{-p}, \quad \mathbb{P}\text{-a.s.}
\]
(ii) If $\beta_1 < \beta_2$, then for all $\pi_0, \bar{\pi}_0 \in \mathcal{P}(\mathbb{R}^d)$, there exist $c_{12} > 0$ and a positive random variable $c_{13}$ such that for all $i > 0$,

$$\|\pi_i - \bar{\pi}_i\|_{TV} \leq c_{13}{i^{\beta_2/\beta_1}}, \quad \text{P.-a.s.}$$

Appendix A

The following lemma and its corollary are taken from [14] and included here for completeness.

**Lemma A.1.** Let $\mathcal{F}_0 = \sigma(X_0)$ and $\mathcal{F}_n \triangleq \sigma(X_0, W_1, \ldots, W_n, V_1, \ldots, V_n)$, then for all $\Delta > 0$, $i > 0$ and $n \geq 0$

$$\mathbb{E}\left[ \prod_{j=1}^{i} (1 - \varepsilon_{n+j}(\Delta)) \bigg| \mathcal{F}_n \right] \leq (1 - \bar{\varepsilon}(\Delta))^{i-1}.$$  

**Proof.** For clarity, let us suppress the dependency on $\Delta$ in the notation $\varepsilon_j(\Delta)$ and let $\tau_j \triangleq (1 - \varepsilon_j)$. Since for all $i \geq n > 1$, $\tau_n$ is $\mathcal{F}_i$-measurable, we have for all $i > 1, n \geq 0$

$$\mathbb{E}\left[ \prod_{j=n+1}^{n+i} \tau_j \bigg| \mathcal{F}_n \right] = \mathbb{E}\left[ \mathbb{E}[\tau_{n+i} \tau_{n+i-1} | \mathcal{F}_{n+i-2}] \prod_{j=n+1}^{n+i-2} \tau_j \bigg| \mathcal{F}_n \right]. \quad (A.1)$$

Let us define

$$\tau(x, \Delta) \triangleq 1 - \frac{m_1}{M_1} \exp(-\alpha_1((ab+b)\Delta + x)^{\beta_1}).$$

Then for all $x, \Delta > 0$

$$\mathbb{E}[\tau_{n+i} \tau_{n+i-1} | \mathcal{F}_{n+i-2}] \leq \mathbb{E}[\tau(x, \Delta) \tau_{n+i-1} \{\tau(x, \Delta) \geq \tau_{n+i-1} \} + \tau_{n+i-1} \{\tau(x, \Delta) < \tau_{n+i-1} \} | \mathcal{F}_{n+i-2}]$$

$$\leq \mathbb{E}[\tau(x, \Delta) \tau_{n+i-1} + (1 - \tau(x, \Delta))1_{\{\tau(x, \Delta) < \tau_{n+i-1} \}} | \mathcal{F}_{n+i-2}]$$

$$\leq \tau(x, \Delta) \mathbb{E}[\tau_{n+i-1} | \mathcal{F}_{n+i-2}] + (1 - \tau(x, \Delta)) p_{n+1}(x), \quad (A.2)$$

where

$$p_{n+1}(x) \triangleq \mathbb{P}(\xi_{n+i} > x) = \mathbb{P}(\xi_{n+i} > x | \mathcal{F}_{n+i-2}) = \mathbb{P}(\tau_{n+i} > \tau(x, \Delta) | \mathcal{F}_{n+i-2}).$$

In the case $i = 2$, according to the definition of $\bar{\varepsilon}(\Delta)$ and (A.2), one has

$$\mathbb{E}[\tau_{n+1} \tau_{n+2} | \mathcal{F}_n] \leq \tau(d, \Delta) + \frac{1}{4}(1 - \tau(d, \Delta)) < 1 - \bar{\varepsilon}(\Delta),$$

were $d > 0$ is chosen such that $p_i(d) < 1/4$ for all $i > 0$. As the claim holds trivially for $i = 1$, it has now been proved for $i = 1, 2$. To complete the proof by induction, it is assumed that the claim holds for $0 < i < m$. Then, by setting $i = m + 1$, the substitution of (A.2) into (A.1) yields

$$\mathbb{E}\left[ \prod_{j=n+1}^{n+m+1} \tau_j \bigg| \mathcal{F}_n \right] \leq \tau(d, \Delta) \mathbb{E}\left[ \prod_{j=n+1}^{n+m} \tau_j \bigg| \mathcal{F}_n \right] + \frac{1}{4}(1 - \tau(d, \Delta)) \mathbb{E}\left[ \prod_{j=n+1}^{n+m-1} \tau_j \bigg| \mathcal{F}_n \right]$$

$$\leq \tau(d, \Delta)(1 - \bar{\varepsilon}(\Delta))^{m-1} + \frac{1}{4}(1 - \tau(d, \Delta))(1 - \bar{\varepsilon}(\Delta))^{m-2}$$

$$\leq (1 - \bar{\varepsilon}(\Delta))^m.$$
Let us point out the following simple corollary of Lemma A.1 which can also be found in \cite{14}.

**Corollary A.2.** For all $\pi_0, \bar{\pi}_0 \in \mathcal{P}(B(\mathbb{R}^d)), \Delta > 0$ and $i \geq 0$
\[
E\left[\|\pi_i^\Delta - \bar{\pi}_i^\Delta\|_{TV}\right] \leq (1 - \bar{\varepsilon}(\Delta))^{i-2}.
\]

**Proof.** From Corollary 3.4, Lemma 4.1 and Lemma A.1 it follows that
\[
E\left[\|\pi_i^\Delta - \bar{\pi}_{i-1}^\Delta\|_{TV}\right] \leq E\left[\prod_{j=1}^{i-1}(1 - \varepsilon_{j+1}(\Delta))\right] \leq (1 - \bar{\varepsilon}(\Delta))^{i-2}.
\]

Although this corollary is not required by the proof of stability it has been included because it essentially states that truncated filters are exponentially stable in mean, i.e. the expected value of the error due to initialisation decreases exponentially fast.

---

**References**

2003.


---

D. Crisan
Department of Mathematics
Huxley Building
Imperial College London
180 Queens Gate
London SW7 2BZ, UK
d.crisan@imperial.ac.uk

K. Heine
Department of Mathematics
Tampere University of Technology
Korkeakoulunkatu 1
P.O.Box 553 FI-33101, Tampere, Finland