Stable approximations of optimal filters

Joaquin Miguez

Department of Signal Theory & Communications,
Universidad Carlos III de Madrid.
E-mail: joaquin.miguez@uc3m.es

Joint work with Dan Crisan (Imperial College London) and Alberto López-Yela
(Universidad Carlos III de Madrid).

Bayesian Computation for High-Dimensional Statistical Models
National University of Singapore, September 5, 2018
Introduction
  Optimal filtering. Stability.

Truncated filters
  Approximation of optimal filters: truncating the likelihood and “reshaping” the kernel.

The normalisation sequence
  Bounding the normalisation constants away from zero (infinitely often).

Stability of the approximations
  Stability of truncated filters. A metric space of optimal filters.

Discussion
  Summary, an example and some hand-waving.
Index

Introduction
  Optimal filtering. Stability.

Truncated filters
  Approximation of optimal filters: truncating the likelihood and “reshaping” the kernel.

The normalisation sequence
  Bounding the normalisation constants away from zero (infinitely often).

Stability of the approximations
  Stability of truncated filters. A metric space of optimal filters.

Discussion
  Summary, an example and some hand-waving.
Introduction

- Optimal filtering in discrete time
- Stable filters – forgetting initial conditions
- Why is stability important?
- Can we approximate unstable filters by stable ones?
- Are there “many” stable filters?
- Work in progress – expect some loose ends!!
Introduction

• Optimal filtering in discrete time
• Stable filters – forgetting initial conditions
  • Why is stability important?
  • Can we approximate unstable filters by stable ones?
  • Are there “many” stable filters?
• Work in progress – expect some loose ends!!
Introduction

- Optimal filtering in discrete time
- Stable filters – forgetting initial conditions
- Why is stability important?
  - Can we approximate unstable filters by stable ones?
  - Are there “many” stable filters?
- Work in progress – expect some loose ends!!
Introduction

• Optimal filtering in discrete time
• Stable filters – forgetting initial conditions
• Why is stability important?
• Can we approximate unstable filters by stable ones?
  • Are there “many” stable filters?
  • Work in progress – expect some loose ends!!
Introduction

- Optimal filtering in discrete time
- Stable filters – forgetting initial conditions
- Why is stability important?
- Can we approximate unstable filters by stable ones?
- Are there “many” stable filters?
- Work in progress – expect some loose ends!!
Introduction

- Optimal filtering in discrete time
- Stable filters – forgetting initial conditions
- Why is stability important?
- Can we approximate unstable filters by stable ones?
- Are there “many” stable filters?
- Work in progress – expect some loose ends!!
State space Markov models

- Random sequences $X_{0:}\infty$ and $Y_{1:}\infty$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
- State-space Markov model

\[
\begin{align*}
X_0 &\sim \pi_0(dx) \quad \text{(prior)} \\
X_t &\sim \kappa_t(dx|x_{t-1}) \quad \text{(Markov kernel)} \\
Y_t &\sim g_t(y_t|x_t)dy_t \quad \text{(likelihood } \propto \text{ pdf of observation)}
\end{align*}
\]

- $X_t \in \mathcal{X}$, $t \geq 0$, is the state of the system.
- $Y_t \in \mathcal{Y}$, $t \geq 1$, is the observation.
- The observations are conditionally independent given the states.
  - Notation: likelihood function $g_t^{Y_t}(x) = g_t(y_t|x)$
Optimal filtering

- The model: \( X_0 \sim \kappa_0, \ X_t \sim \kappa_t(\cdot|x_{t-1}), \ Y_t \sim g_t(\cdot|x_t) \)
- The optimal filtering problem: recursively compute...
  - the “filters” \( \pi_t(A) := \mathbb{P}(X_t \in A|Y_{1:t} = y_{1:t}) \)
  - the predictive measures \( \xi_t(A') := \mathbb{P}(X_t \in A'|Y_{1:t-1} = y_{1:t-1}) \)
- Notation for integrals: \( f \) a test function, \( \mu \) a measure
  \[
  (f, \mu) = \int f(x)\mu(dx).
  \]
- **Prediction** and **update** equations
  \[
  \xi_t = \kappa_t \pi_{t-1} \quad \iff \quad (f, \xi_t) = \int \int f(x_t)\kappa_t(dx_t|x_{t-1})\pi_{t-1}(dx_{t-1})
  \]
  \[
  \pi_t = g_{\cdot t}^Y \cdot \xi_t \quad \iff \quad (f, \pi_t) = \frac{(fg_{\cdot t}^Y, \xi_t)}{(g_{\cdot t}^Y, \xi_t)}
  \]
Stability of the optimal filter

- Let $\mathcal{P}(\mathcal{X})$ denote the set of probability measures over $\mathcal{X}$.
- The state space Markov model can be represented by
  - a prior distribution $\pi_0$ and
  - an operator on measures $\Phi_t : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$, defined as

\[
\Phi_t(\mu) := g_t^y \cdot \kappa_t \mu \leftrightarrow (f, \Phi_t(\mu)) := \frac{(fg_t^y, \kappa_t \mu)}{(g_t^y, \kappa_t \mu)}
\]

- If we denote $\Phi_{t|k}(\mu) = (\Phi_t \circ \Phi_{t-1} \circ \cdots \circ \Phi_{k+1})(\mu)$ then

\[
\pi_t = \Phi_t(\pi_{t-1}), \quad \pi_t = \Phi_{t|k}(\pi_k), \quad \text{and, in particular,} \quad \pi_t = \Phi_{t|0}(\pi_0).
\]

- We refer to $\Phi_t$ as a prediction-update (PU) operator.
Stability of the optimal filter

**Definition (Stable filter)**

The sequence of optimal filters generated by the PU operators \( \{\Phi_t\}_{t \geq 1} \) is stable iff

\[
\lim_{t \to \infty} |(f, \Phi_{t|0}(\mu)) - (f, \Phi_{t|0}(\nu))| = 0
\]

for any pair of initial measures \( \mu, \nu \in \mathcal{P}(\mathcal{X}) \).

- We say that the sequence of optimal filters \( \pi_t = \Phi_t(\pi_{t-1}) \) is stable, ...
- ... or that the operator \( \Phi_{t|0} \) is stable...
- ... or simply that the optimal filter is stable.
- Various more detailed forms of stability can be defined (\( \mathbb{P}\text{-a.s.}, \) for \( \mu << \nu \), with specific rates, etc.)
Stability of the optimal filter

- The analysis of filter stability is not an easy task...

- Some (relatively) recent references:
  - (Chigansky & Liptser, 2006) – stability for a restricted class of test functions (defined in terms of \( g_t \) and \( \kappa_t \)); martingale convergence.
  - (Kleptsyna & Veretennikov, 2008; Douc et al., 2009; Gerber & Whiteley, 2017) – mixing state process; assumptions not straightforward to check.
  - (Van Handel, 2009) – continuous-time filters; links stability with observability.
  - (Heine & Crisan, 2008) – simple conditions (tails of noise distributions); restricted class of state space models.

- To this day, no attempt to assess whether stable filters are “many” or “few” for a given class of models.

- We adopt (and adapt) conditions from (Del Moral & Guionnet, 2001) and (Del Moral, 2004):
  - sufficient mixing,
  - can be tested on the kernel \( \kappa_t \).
Introduction

Optimal filtering. Stability.

Truncated filters

Approximation of optimal filters: truncating the likelihood and “reshaping” the kernel.

The normalisation sequence

Bounding the normalisation constants away from zero (infinitely often).

Stability of the approximations

Stability of truncated filters. A metric space of optimal filters.

Discussion

Summary, an example and some hand-waving.
Approximate SSM’s = Approximate filters

- We describe a state space model (SSM) as
  \[ S = (\pi_0, \kappa_t, g^y_t) = (\pi_0, \Phi_t) \]
  and it generates the sequence of filters
  \[ \pi_t = \Phi_{t|0}(\pi_0). \]

- We aim at constructing modified models, say
  \[ S^c = (\pi_0, \kappa^c_t, g^{y_t,c}) = (\pi_0, \Phi^c_t) \]
  in such a way that
  \[ \pi^c_t = \Phi^c_{t|0}(\pi_0) \quad \text{and} \quad |(f, \pi_t) - (f, \pi^c_t)| < \epsilon \quad \text{uniformly over time}. \]

- We keep the same \( \pi_0 \) for the two models. Why??
Truncation scheme

1. Choose a sequence of compact subsets $c = \{ C_t \subset \mathcal{X} \}_{t \geq 1}$.
2. Construct a sequence of truncated likelihoods
   \[
   g_{t}^{y_{t}, c}(x) := \mathbb{1}_{C_{t}}(x) g_{t}^{y_{t}}(x) = \begin{cases} 
   g_{t}^{y_{t}}(x), & \text{if } x \in C_{t} \\
   0, & \text{otherwise}
   \end{cases}
   \]
3. Construct a sequence of “reshaped” kernels
   \[
   \kappa_{1}^{c}(dx_{1}|x_{0}) := \kappa_{1}(dx_{1}|x_{0})
   \]
   \[
   \rho_{t}(dx_{t}) := \int_{\mathcal{X} \setminus C_{t-1}} \kappa_{t}(dx_{t}|x_{t-1}) \pi_{t-1}(dx_{t-1})
   \]
   \[
   \kappa_{t}^{c}(dx_{t}|x_{t-1}) := \pi_{t-1}(C_{t-1}) \kappa_{t}(dx_{t}|x_{t}) + \rho_{t}(dx_{t}).
   \]
4. Let $S^{c} := (\pi_{0}, \kappa_{t}^{c}, g_{t}^{y_{t}, c}) = (\pi_{0}, \Phi_{t}^{c})$, where
   \[
   \Phi_{t}^{c}(\mu) := g_{t}^{y_{t}, c} \cdot \kappa_{t}^{c} \mu \quad \text{and} \quad \pi_{t}^{c} = \Phi_{t|0}^{c}(\pi_{0}).
   \]
A picture of the reshaped kernel $\kappa^c_t$
An approximation theorem

Lemma (Error reset)
Let $S = (\pi_0, \kappa_t, g_t)$ be a state space model and let $c = \{C_t\}_{t \geq 1}$ be a sequence of compact subsets of $\mathcal{X}$. The truncated state space model $S^c = (\pi_0, \kappa^c_t, g^y_t, c)$ yields sequences of predictive and filtering probability measures $(\xi^c_t$ and $\pi^c_t$, respectively) such that, for every $f \in B(\mathcal{X})$ and every $t \geq 1$,

\[
(\mathbb{1}_{C_t} f, \xi_t) = (\mathbb{1}_{C_t} f, \xi^c_t), \quad \text{and} \\
(\mathbb{1}_{C_t} f, \pi_t) = (f, \pi^c_t) \pi_t(C_t).
\]

Theorem (Uniform approximation)
Let $S = (\pi_0, \kappa_t, g_t)$ be a state space model and let $c = \{C_t\}_{t \geq 1}$ be a sequence of compact subsets of $\mathcal{X}$. Assume every $C_t$ satisfies $(\mathbb{1}_{\bar{C}_t}, \pi_t) < \frac{\epsilon}{2}$ for some prescribed $\epsilon \in (0, 1)$. Then, the truncated model $S^c = (\pi_0, \kappa^c_t, g^y_t, c)$ yields a sequence of filters $\pi^c_t = \Phi^c_{t|0}(\pi_0)$ such that, for every $f \in B(\mathcal{X})$,

\[
\sup_{t \geq 1} |(f, \pi_t) - (f, \pi^c_t)| < \|f\|_{\infty} \epsilon.
\]
Stability lemma

- The approximate model $S^c$ is useful only if the approximate filters $\pi_t^c = \Phi_{t|0}(\pi_0)$ are stable...

Lemma (Adapted from (Del Moral & Guionnet, 2001))

Let $c = \{C_t \subseteq \mathcal{X}\}_{t>0}$ be a sequence of compact subsets of the state space $\mathcal{X}$ and let $\Phi_t^c$ be the truncated PU operator induced by the pair $(\kappa_t, g_t^c)$. If the Markov kernels $\kappa_t$ have positive probability densities $k_t$ with respect to a reference measure $\lambda$,

$$k_t(\cdot|x_{t-1}) = \frac{d\kappa_t(\cdot|x_{t-1})}{d\lambda},$$

and

$$\sum_{t=1}^{\infty} \inf_{(x_{t-1},x_t) \in C_{t-1} \times C_t} \sup_{(x_{t-1},x_t) \in C_{t-1} \times C_t} k_t(x_t|x_{t-1}) = \infty,$$

then the PU operator $\Phi_t^c$ is stable, i.e.,

$$\lim_{t \to \infty} |(f_0, \Phi_{t|0}(\mu)) - (f_0, \Phi_{t|0}(\mu'))| = 0$$

for every $\mu, \mu' \in \mathcal{P}(\mathcal{X})$ and every $f \in \mathcal{B}(\mathcal{X})$. 
Index

Introduction
  Optimal filtering. Stability.

Truncated filters
  Approximation of optimal filters: truncating the likelihood and “reshaping” the kernel.

The normalisation sequence
  Bounding the normalisation constants away from zero (infinitely often).

Stability of the approximations
  Stability of truncated filters. A metric space of optimal filters.

Discussion
  Summary, an example and some hand-waving.
Random observations & random measures

- Assume the sequence of observations $Y_{1:\infty}$ is random.
- The sequences of optimal filters & predictive measures are hence random,
  \[ \xi^{Y_{1:t-1}}_t := \kappa_t \pi^{Y_{1:t-1}}_{t-1} \iff (f, \xi^{Y_{1:t-1}}_t) = ((f, \kappa_t), \pi^{Y_{1:t-1}}_t) \]
  \[ \pi^{Y_{1:t}}_t := g^Y_t \cdot \xi^{Y_{1:t-1}}_{t-1} \iff (f, \pi^{Y_{1:t}}_t) = \left( \frac{fg^{Y_{1:t}}_t \xi^{Y_{1:t-1}}_t}{g^Y_t \xi^{Y_{1:t-1}}_{t-1}} \right) \]
- Define the $\sigma$-algebra generated by $Y_{1:t}$, denoted $G_t = \sigma - (Y_{1:t})$:
  \[ \xi^{Y_{1:t-1}}_t \text{ is } G_{t-1}\text{-measurable}, \quad \pi^{Y_{1:t}}_t \text{ is } G_t\text{-measurable.} \]
- The $G_t$-measurable r.v.
  \[ (g^Y_t, \xi^{Y_{1:t-1}}_t) \]
  is the normalisation constant of the (random) filter $\pi^{Y_{1:t}}_t$.
- The $\mathcal{Y} \to [0, \infty)$ function
  \[ p_t(y) := (g^Y_t = y, \xi^{Y_{1:t-1}}_t) \]
  is the (predictive) pdf of the random observation $Y_t$ conditional on the $\sigma$-algebra $G_{t-1}$. 
Why is the normalisation sequence important?

• Let \( c = \{ C_t \}_{t \geq 1} \) be a sequence of compacts for truncation. The approximation theorem “works” when

\[
\pi_t^{Y_{1:t}} (\bar{C}_t) < \frac{\epsilon}{2}, \quad \text{where} \quad \bar{C}_t = \mathcal{X} \setminus C_t.
\]

This is a “control on the tails” of the filter \( \pi_t \).

• If we decompose this factor,

\[
\pi_t^{Y_{1:t}} (\bar{C}_t) = \frac{\left( \mathbb{1}_{\bar{C}_t} g_t^{Y_t}, \xi_t^{Y_{1:t-1}} \right)}{(g_t^{Y_t}, \xi_t^{Y_{1:t-1}})},
\]

truncation can help with the numerator, e.g.,

\[
\left( \mathbb{1}_{\bar{C}_t} g_t^{Y_t}, \xi_t^{Y_{1:t-1}} \right) \leq \sup_{\bar{C}_t} g_t^{Y_t}(x)
\]

• ...but not with the denominator. In general,

\[
\mathbb{P} \left( \inf_{t \geq 1} (g_t^{Y_t}, \xi_t^{Y_{1:t-1}}) = 0 \right) > 0
\]

hence \( \pi_t^{Y_{1:t}} (\bar{C}_t) \) is hard to upper bound.
Why is the normalisation sequence important?

• Let $c = \{C_t\}_{t \geq 1}$ be a sequence of compacts for truncation. The approximation theorem “works” when

$$\pi_t^{Y_{1:t}}(\bar{C}_t) < \frac{\epsilon}{2}, \quad \text{where} \quad \bar{C}_t = \mathcal{X} \setminus C_t.$$  

This is a “control on the tails” of the filter $\pi_t$.

• If we decompose this factor,

$$\pi_t^{Y_{1:t}}(\bar{C}_t) = \left( \frac{\mathbb{1}\bar{C}_t g_t^{Y_t}, \xi_t^{Y_{1:t-1}}}{(g_t^{Y_t}, \xi_t^{Y_{1:t-1}})} \right),$$

truncation can help with the numerator, e.g.,

$$(\mathbb{1}\bar{C}_t g_t^{Y_t}, \xi_t^{Y_{1:t-1}}) \leq \sup_{\bar{C}_t} g_t^{Y_t}(x)$$

• ...but not with the denominator. In general,

$$\mathbb{P}\left( \inf_{t \geq 1} (g_t^{Y_t}, \xi_t^{Y_{1:t-1}}) = 0 \right) > 0$$

hence $\pi_t^{Y_{1:t}}(\bar{C}_t)$ is hard to upper bound.
Why is the normalisation sequence important?

• Let $c = \{C_t\}_{t \geq 1}$ be a sequence of compacts for truncation. The approximation theorem “works” when

$$\pi_t^{Y_{1:t}}(\bar{C}_t) < \frac{\epsilon}{2}, \quad \text{where} \quad \bar{C}_t = \mathcal{X} \setminus C_t.$$ 

This is a “control on the tails” of the filter $\pi_t$.

• If we decompose this factor,

$$\pi_t^{Y_{1:t}}(\bar{C}_t) = \frac{(\mathbb{1}_{\bar{C}_t} g_t^{Y_{t}}, \xi_t^{Y_{1:t-1}})}{(g_t^{Y_{t}}, \xi_t^{Y_{1:t-1}})},$$

truncation can help with the numerator, e.g.,

$$(\mathbb{1}_{\bar{C}_t} g_t^{Y_{t}}, \xi_t^{Y_{1:t-1}}) \leq \sup_{\bar{C}_t} g_t^{Y_{t}}(x)$$

• ...but not with the denominator. In general,

$$\mathbb{P} \left( \inf_{t \geq 1} (g_t^{Y_{t}}, \xi_t^{Y_{1:t-1}}) = 0 \right) > 0$$

hence $\pi_t^{Y_{1:t}}(\bar{C}_t)$ is hard to upper bound.
Bounding the normalisation sequence from below

- So, we need to bound \((g_t^{Y_t}, \xi_t)\) away from zero... at least sometimes.

**Assumption (Predictability)**
There is some constant \(\gamma > 0\) such that \(E[(g_t^{Y_t}, \xi_t^{1:t-1})|G_{t-1}] \geq \gamma\) for every \(t \geq 1\).

**Theorem (Normalisation sequence)**
Choose a state space model \(S = (\pi_0, \kappa_t, g_t^{Y_t})\). If the predictability assumption holds and \(\|g\|_\infty < \infty\), then \(P\)-a.s. there exists an infinite sequence of positive integers \(\{t_n\}_{n \geq 1}\) such that

\[
\left( g_t^{Y_{t_n}}, \xi_t^{Y_{1:t_n-1}} \right) \wedge \left( g_t^{Y_{t_n+1}}, \xi_t^{Y_{1:t_n}} \right) > \frac{\gamma}{2} \quad \text{for every } n \geq 1
\]

and

\[
\lim_{T \to \infty} \frac{\sum_{t=1}^T \mathbb{1}_{\{t_n\}_{n \geq 1}}(t)}{T} \geq \epsilon_2 > 0
\]

for some \(\epsilon_2 > 0\).
Bounding the normalisation sequence from below

- So, we need to bound \((g_t^{Y_t}, \xi_t)\) away from zero... at least sometimes.

**Assumption (Predictability)**

There is some constant \(\gamma > 0\) such that \(\mathbb{E}[ (g_t^{Y_t}, \xi_t^{Y:1:t-1}) | G_{t-1}] \geq \gamma\) for every \(t \geq 1\). □

**Theorem (Normalisation sequence)**

Choose a state space model \(S = (\pi_0, \kappa_t, g_t^{Y_t})\). If the predictability assumption holds and \(\|g\|_\infty < \infty\), then \(\mathbb{P}\text{-a.s.}\) there exists an infinite sequence of positive integers \(\{t_n\}_{n \geq 1}\) such that

\[
\left( g_t^{Y_{t_n}}, \xi_t^{Y:1:t_n-1} \right) \land \left( g_t^{Y_{t_n+1}}, \xi_t^{Y:1:t_n} \right) > \frac{\gamma}{2} \quad \text{for every } n \geq 1
\]

and

\[
\lim_{T \to \infty} \frac{\sum_{t=1}^{T} \mathbb{1}_{\{t_n\}_{n \geq 1}}(t)}{T} \geq \epsilon_2 > 0
\]

for some \(\epsilon_2 > 0\).
When are the observations predictable?

- When are the observations predictable? Essentially, when the predictive pdf does not vanish with time,

\[ \lim_{t \to \infty} \sup_{y \in Y} p_t(y) > 0 \quad + \text{some continuity}. \]

- Alternative sufficient conditions can be stated for the predictability assumption.

- A simple example: for a linear-Gaussian system

\[
X_0 \sim \mathcal{N}(x_0; m_0, \sigma_0^2), \quad X_t \sim \mathcal{N}(x_t; aX_{t-1}, \sigma_u^2), \quad Y_t \sim \mathcal{N}(y_t; bX_t, \sigma_v^2),
\]

it is sufficient that \( \limsup_{t \to \infty} \text{Var}(Y_t | G_{t-1}) < \infty \).
When are the observations predictable?

- When are the observations predictable? Essentially, when the predictive pdf does not vanish with time,

\[ \lim_{t \to \infty} \sup_{y \in \mathcal{Y}} p_t(y) > 0 \quad + \text{some continuity.} \]

- Alternative sufficient conditions can be stated for the predictability assumption.

- A simple example: for a linear-Gaussian system

\[
X_0 \sim \mathcal{N}(x_0; m_0, \sigma_0^2), \quad X_t \sim \mathcal{N}(x_t; aX_{t-1}, \sigma_u^2), \quad Y_t \sim \mathcal{N}(y_t; bX_t, \sigma_v^2),
\]

it is sufficient that \( \limsup_{t \to \infty} \text{Var}(Y_t|G_{t-1}) < \infty \).
When are the observations predictable?

- A more elaborate one...

**Proposition**

Assume that $\lambda$ is the Lebesgue measure and the three conditions below are satisfied:

(c1) The conditional pdf $g_t(y|x)$ is Lipschitz w.r.t. both $x$ and $y$; specifically, there exist constants $L_1, L_2 < \infty$ independent of $t$ such that, for all $t \geq 1$,

$$|g_t^y(x) - g_t^{y'}(x)| \leq L_1 \|y - y'\| \quad \text{and} \quad |g_t^y(x) - g_t^y(x')| \leq L_2 \|x - x'\|.$$

(c2) There exists $\bar{m} > 0$ independent of $t$ such that, for every $x \in \mathcal{X}$,

$$\sup_{y \in \mathcal{Y}} g_t^y(x) \geq \bar{m}.$$

(c3) There exists $\epsilon_0 > 0$, arbitrarily small, such that

$$\inf_{t \geq 1} \max_{x \in \mathcal{X}} \xi_{1:t}^1(B(x, \epsilon_0)) \geq C_{d_x} \epsilon_0^{d_x+1} \quad \mathbb{P}\text{-a.s.}$$

for some constant $C_{d_x} > 0$.

Then, there exists $\gamma > 0$ such that $\mathbb{E}[(g_t^{Y_1}, \xi_{t:1}^1|\mathcal{G}_{t-1}) > \gamma].$
Index

Introduction
   Optimal filtering. Stability.

Truncated filters
   Approximation of optimal filters: truncating the likelihood and “reshaping” the kernel.

The normalisation sequence
   Bounding the normalisation constants away from zero (infinitely often).

Stability of the approximations
   Stability of truncated filters. A metric space of optimal filters.

Discussion
   Summary, an example and some hand-waving.
State space model

- Prior: $\pi_0$
- Markov kernel: $\kappa_t$ has a pdf $k_t$,

$$\kappa_t(dx_t|x_{t-1}) = k_t(x_t|x_{t-1})\lambda(dx_t),$$

and there is a function $a_t : \mathcal{X} \to \mathcal{X}$ such that

$$a_t(x_{t-1}) = \int x_t\kappa_t(dx_t|x_{t-1}).$$

- Likelihood: there is some function $b_t : \mathcal{X} \to \mathcal{Y}$ such that

$$b_t(x_t) = \int y_t g_t^y(x_t)dy_t.$$

- Essentially, an additive model.
Model assumptions

Assumption (MLip)
Functions $a_t : \mathcal{X} \to \mathcal{X}$, $t = 1, 2, \ldots$, are uniformly Lipschitz, i.e., there exists a constant $L_a < \infty$ such that

$$
\|a_t(x) - a_t(x')\| < L_a \|x - x'\| \quad \text{for every pair } (x, x') \in \mathcal{X}^2 \text{ and every } t > 0.
$$

Assumption (M-)
For every $t \geq 1$ there exists and invertible decreasing function $s_t : [0, \infty) \to [0, \infty)$ such that

$$
k_t(x|x') \geq s_t(\|x - a_t(x')\|) \quad \text{for every pair } (x, x') \in \mathcal{X}^2,
$$

and $\lim_{r \to \infty} s_t(r) = 0$.

Assumption (M+)
The pdf’s $k_t$ are uniformly upper bounded. Specifically, there exists a constant $C_u < \infty$ such that

$$
\sup_{t \geq 1; (x,x') \in \mathcal{X}^2} k_t(x|x') < C_u < \infty
$$

and, $\lim_{\|x - a_t(x')\| \to \infty} k_t(x|x') = 0$. 
The sequence of compact sets

- Given a constant $M < \infty$, choose a sequence of closed balls
  \[ C^M_t := B(\ell_t, M r_t) = \{ x \in \mathcal{X} : \| x - \ell_t \| \leq M r_t \}, \quad t \geq 1 \]
  where $\ell_t \in \mathcal{X}$ and $r_t \in [0, \infty)$.  

- The centres satisfy
  \[ \| \ell_t - a_t(\ell_{t-1}) \| < M L r_t, \quad t = 1, 2, \ldots, \]
  for some constant $L < \infty$. …

- …and the radius $\{ r_t \}_{t \geq 1}$ is strictly increasing and satisfies
  \[ \lim_{t \to \infty} \frac{s^{-1}_t(\frac{1}{t})}{r_t} = \infty, \]
  for the functions $s_t$ in assumption (M-). Intuitively, $\{ r_t \}_{t \geq 1}$ increases at a sufficiently slow rate.

**Lemma**

For any given constant $M < \infty$, if assumptions (MLip) and (M-) hold then there exists $t'_M < \infty$ such that

\[ \inf_{(x, x') \in C^M_t \times C^M_{t-1}} k_t(x|x') > \frac{1}{t} \quad \text{for every } t > t'_M. \]
Finite horizon approximation

- Let $S = (\pi_0, \kappa_t, g_t^{y_t})$ be a target SSM satisfying assumptions (MLip), (M-) and (M+).
- Using the sequence $c = \{C_t^M\}_{t \geq 1}$, we construct the model $S_c^M = (\pi_0, \kappa_t^c, M, g_t^{y_t,c,M})$, where $g_t^{y_t,c,M} = 1_{C_t^M} g_t^{y_t}$ and
  
  $\kappa_t^c, M = \begin{cases} 
  \kappa_t^c, M, & \text{for } t = 1, \ldots, T \text{ (reshaped)} \\
  \kappa_t, & \text{for } t > T,
  \end{cases}$

  for some finite horizon $T < \infty$.
- The PU operator is $\Phi_t^{c,M}$ and $\pi_t^{c,M} = \Phi_t^{c,M}(\pi_0)$.

**Theorem (Stable approximation, finite horizon)**

Let assumptions (MLip), (M-) and (M+) hold. Then, for every pair of constants $\epsilon > 0$ and $T < \infty$, there exists $M_{\epsilon,T} < \infty$ such that

$$|(f, \pi_t) - (f, \pi_t^{c,M})| < \epsilon \|f\|_{\infty}, \quad 1 \leq t \leq T,$$

for every $f \in B(\mathcal{X})$ and every $M \geq M_{\epsilon,T}$. Moreover, the PU operator $\Phi_t^{c,M}$ is stable.
A topological interpretation

- Let $\mathcal{G}$ denote the set of state space models $S = (\pi_0, \kappa_t, g^{\gamma_t}_t)$ that satisfy assumptions (MLip), (M-) and (M+).
- We construct metric spaces $(\mathcal{G}, D_q)$, $q > 1$, where
  \[ D_q(S, S') := \sum_{t=0}^{\infty} \frac{1}{q^t} D_{tv}(\pi_t, \pi'_t) \]
  and
  \[ D_{tv}(\pi, \pi') := \sup_{A \in \mathcal{B}(\mathcal{X})} |\pi(A) - \pi'(A)| \leq 1 \]
  is the total variation distance between $\pi, \pi' \in \mathcal{P}(\mathcal{X})$.
- Stable truncated approximations are dense in the space $(\mathcal{G}, D_q)$...

**Theorem (Stable approximations are dense)**

The subset $\mathcal{G}_0 := \{ S \in \mathcal{G} : \Phi_t \text{ is stable} \}$ is dense in the metric space $(\mathcal{G}, D_q)$. 
Uniform & stable approximations

• Back to random observations & measures!

Assumption (G)
For any $\bar{\epsilon} > 0$, arbitrarily small, there is some $t_{\bar{\epsilon}}$ and some $M_{\bar{\epsilon}}$ such that for every $t \geq t_{\bar{\epsilon}}$ and every $M \geq M_{\bar{\epsilon}}$ we can find a sequence $\ell_{Y}^{1:t} \in \mathcal{X}^{t}$ satisfying the inequality

$$\sup_{x \in \mathcal{X} \setminus B(\ell_{t}^{Y1:t}, Mr_{t})} g_{t}^{Y_{t}}(x) < \bar{\epsilon} \quad \mathbb{P}\text{-a.s.}$$

Theorem (Uniform & stable approximation)
Let $S = (\pi_{0}, \kappa_{t}, g_{t}^{Y_{t}})$ be a state space model satisfying assumptions (MLip), (M-), (M+) and (G). If the predictability assumption also holds, then for every $\epsilon > 0$ it is possible to choose a sequence of compact sets $c = \{C_{t}\}_{t \geq 0}$ and a truncated state space model $S_{c} = (\pi_{0}, \kappa_{t}^{c}, g_{t}^{Y_{t}, c})$ such that the associated PU operator $\Phi_{t}^{c,Y_{t}}$ is $\mathbb{P}$-a.s. stable and

$$|(f, \pi_{t}) - (f, \pi_{t}^{Y_{1:t}, c})| \leq \epsilon \|f\|_{\infty} \quad \mathbb{P}\text{-a.s., for every } t \geq 1 \text{ and any } f \in B(\mathcal{X}).$$
Sketch of the proof: uniform & stable approx.

1. Under the predictability assumption, $\mathbb{P}$-a.s. there exists an infinite sequence $\{t_n\}_{n \geq 1}$ such that

$$
(g_{t_n}^{Y_{t_n}}, \xi_{t_n}^{Y_{1:t_n-1}}) \land (g_{t_n+1}^{Y_{t_n+1}}, \xi_{t_n}^{Y_{1:t_n}}) > \frac{\gamma}{2}.
$$

2. We construct the sequence of compacts $c = \{C_t\}$

$$
C_t = \begin{cases} 
B(\ell_t^{Y_{1:t}}, M' r_t), & \text{if } t \in \{t_n, t_n + 1\}_{n \geq 1}, \\
B(\ell_t^{Y_{1:t}}, R_t), & \text{if } t \notin \{t_n\}_{n \geq 1},
\end{cases}
$$

with $M'$ just “sufficiently large” (via assumption (G)) while $R_t$ is as large as needed to ensure that $\pi_t^{Y_{1:t}}(\mathcal{X} \setminus C_t) < \frac{\epsilon}{2}$.

3. Using assumption (G) again, together with inequality (1),

$$
\pi_t^{Y_{1:t_n}}(\mathcal{X} \setminus B(\ell_{t_n}^{Y_{1:t_n}}, M' r_{t_n})) = \frac{(\mathbb{1}_{\mathcal{X} \setminus B(\ell_{t_n}^{Y_{1:t_n}}, M' r_{t_n})} g_{t_n}^{Y_{t_n}}, \xi_{t_n}^{Y_{1:t_n-1}})}{(g_{t_n}^{Y_{t_n}}, \xi_{t_n}^{Y_{1:t_n-1}})} < \frac{2\bar{\epsilon}}{\gamma} = \frac{\epsilon}{2},
$$

with $\epsilon = \frac{4\bar{\epsilon}}{\gamma}$.

4. Now the uniform approximation Theorem yields

$$
\sup_{t \geq 1} |(f, \pi_t) - (f, \pi_t^{Y_{1:t},c})| \leq \epsilon \|f\|_{\infty} \quad \mathbb{P}\text{-a.s., for every } f \in B(\mathcal{X}).
$$
Sketch of the proof: uniform & stable approx.

5. As for the stability of $\Phi_{t}^{c, Y_t}$, we note that

$$\inf_{(x,x') \in C_{tn+1} \times C_{tn}} k_{tn+1|tn}^{c} (x|x') \geq \pi_{tn}^{Y_{1:tn}} (C_{tn}) \inf_{(x,x') \in C_{tn+1} \times C_{tn}} k_{tn+1|tn} (x|x') > \frac{1 - \frac{\epsilon}{2}}{tn}$$

for $tn > t_{M'}$ (and $t_{M'} < \infty$).

6. Therefore,

$$\sum_{t \geq 1} \inf_{(x,x') \in C_{t+1} \times C_{t}} k_{t+1}^{c} (x|x') > \sum_{n:tn > t_{M'}} \pi_{tn}^{Y_{1:tn}} (C_{tn}) \inf_{(x,x') \in C_{tn+1} \times C_{tn}} k_{tn+1|tn} (x|x')$$

$$> \sum_{n:tn > t_{M'}} \frac{1 - \frac{\epsilon}{2}}{tn} \rightarrow \infty.$$

7. The latter inequality yields stability of $\Phi_{t}^{c}$ (via the stability Lemma). □
Index

Introduction
   Optimal filtering. Stability.

Truncated filters
   Approximation of optimal filters: truncating the likelihood and “reshaping” the kernel.

The normalisation sequence
   Bounding the normalisation constants away from zero (infinitely often).

Stability of the approximations
   Stability of truncated filters. A metric space of optimal filters.

Discussion
   Summary, an example and some hand-waving.
Summary

- An (ongoing) effort to construct stable approximations of optimal filters – for as general families as possible.

- Main results up to this time:
  - A uniform approximation theorem for truncated filters.
  - A probabilistic characterisation of the normalisation constants of optimal filters.
  - A class of metric spaces of optimal filters where stable filters are provably dense.
  - A (somewhat restrictive) class of particle filters which admit stable & uniform (over time) approximations.

- Current version of the manuscript:
  https://arxiv.org/abs/1809.00301
PS. An example & hand-waving

- We are seeking examples of models which yield unstable filters but can be approximated (for a fixed prior $\pi_0$) by stable truncated filters.
- Not so easy: proving either stability or instability is usually hard.
- Good candidates in the following class of state space models:
  - The state space can be partitioned as $\mathcal{X} = \bigcup_{i=1}^{m} \mathcal{X}_i$.
  - The Markov kernel does not transfer probability mass from $\mathcal{X}_i$ to $\mathcal{X}_j$:
    $$\kappa_t(A_i|x_j) = 0 \quad \text{whenever } A_i \subseteq \mathcal{X}_i \text{ and } x_j \in \mathcal{X}_j, \ i \neq j.$$
  - The restricted likelihoods $g_{i,t}(x) = \mathbb{1}_{\mathcal{X}_i}(x)g_{i,t}^{yt}(x)$ separately satisfy assumption (G).
  - A simple one (which can be worked out):

$$\begin{align*}
\mathcal{X}_0 &\sim \pi_0, \\
\mathcal{X}_t &= a\mathcal{X}_{t-1} + \frac{\sigma \mathcal{X}_t}{|\mathcal{X}_t|} |U_t|, \quad U_t \sim \mathcal{N}(0, 1), \\
Y_t &= |\mathcal{X}_t| + \sigma' V_t, \quad V_t \sim \mathcal{N}(0, 1),
\end{align*}$$


