Knots and variance ordering of sequential Monte Carlo algorithms

Joshua J Bon*
School of Mathematical Science, Adelaide University
and
Anthony Lee
School of Mathematics, University of Bristol

October 29, 2025

Abstract

Sequential Monte Carlo algorithms, or particle filters, are widely used for approximating intractable integrals, particularly those arising in Bayesian inference and state-space models. We introduce a new variance reduction technique, the knot operator, which improves the efficiency of particle filters by incorporating potential function information into part, or all, of a transition kernel. The knot operator induces a partial ordering of Feynman–Kac models that implies an order on the asymptotic variance of particle filters, offering a new approach to algorithm design. We discuss connections to existing strategies for designing efficient particle filters, including model marginalisation. Our theory generalises such techniques and provides quantitative asymptotic variance ordering results. We revisit the fully-adapted (auxiliary) particle filter using our theory of knots to show how a small modification guarantees an asymptotic variance ordering for all relevant test functions.

Keywords: Particle filters, Feynman–Kac models, variance ordering, Rao–Blackwellisation

*JJB was supported by the European Union under the GA 101071601, through the 2023-2029 ERC Synergy grant OCEAN. AL was supported by the Engineering and Physical Sciences Research Council (EP/Y028783/1).

Contents

1 Introduction

Sequential Monte Carlo (SMC) algorithms, or particle filters, are malleable tools for estimating intractable integrals. These algorithms generate particle approximations for a sequence of probability measures on a path space, typically specified as a discrete-time Feynman–Kac model for which the normalising constant is intractable. Particle filters construct such approximations using Monte Carlo simulation, importance weighting, and resampling to propagate particles through this sequence of probability measures.

SMC algorithms are used in diverse areas including signal processing (Gustafsson et al., 2002; Doucet and Wang, 2005), object tracking (Mihaylova et al., 2014; Wang et al., 2017), robotics (Thrun, 2002; Stachniss and Burgard, 2014), econometrics (Lopes and Tsay, 2011; Creal, 2012; Kantas et al., 2015), weather forecasting (Fearnhead and Künsch, 2018; Van Leeuwen et al., 2019), epidemiology (Endo et al., 2019; Temfack and Wyse, 2024), and industrial prognostics (Jouin et al., 2016). SMC algorithms are also used extensively in Bayesian posterior sampling (i.e. SMC samplers, Del Moral et al., 2006; Dai et al., 2022) and in other difficult statistical tasks, such as rare event estimation (Cérou et al., 2012). Recently too, SMC algorithms have been used in areas of machine learning such as reinforcement learning (Lazaric et al., 2007; Lioutas et al., 2023; Macfarlane et al., 2024) and denoising diffusion models (Cardoso et al., 2024; Phillips et al., 2024) for example.

The extensive use of SMC algorithms across sciences and their ubiquity in computational statistics can be explained by the generality of their specification. SMC can be used in many different statistical problems and there are often several types of particle filters for a specific case. As SMC is fundamentally related to importance sampling, it is typical that several components of the algorithm can be altered, and accommodated for, by weighting without affecting the target probability measure. SMC samplers have further degrees of freedom as the path of distributions targeted can also be selected. Given this malleability, the design of efficient SMC algorithms remains an active area of research.

In the canonical SMC algorithm, the bootstrap particle filter (Gordon et al., 1993), information is incorporated into the particle system according to the time of observation. Methods such as the auxiliary particle filter (Pitt and Shephard, 1999; Johansen and Doucet, 2008), look-ahead particle filters (Lin et al., 2013), and model twisting methods (Guarniero et al., 2017; Heng et al., 2020), define new particle filters that incorporate varying degrees of future information into the current time step. Idealised versions of these algorithms reduce or eliminate variance in the particle system, producing more accurate particle approximations at the cost of increased computation. Typically, these idealised filters are not computationally tractable or are prohibitively expensive to calculate, so practical methods frequently employ approximations to an ideal filter.

In the simplest case, locally (i.e. conditional) optimal proposal distributions are used to define new particle filters where Monte Carlo simulation is adapted to the current potential

information (Pitt and Shephard, 1999; Doucet et al., 2000). The locally optimal proposal ensures the variance of the importance weights, conditional on a current particle state, is zero. The so-called fully-adapted (auxiliary) particle filter may reduce the overall variance in the particle system and has demonstrated good empirical performance. However, Johansen and Doucet (2008) note that such strategies do not guarantee an asymptotic variance reduction for a given test function.

When it is not possible to implement locally optimal proposals exactly, it may still be possible to find a proposal which reduces the variance of the importance weights. Rao–Blackwellisation adapts a subset of the state space to current potential information, and has freedom in the choice of subset (Chen and Liu, 2000; Doucet et al., 2002; Andrieu and Doucet, 2002; Schön et al., 2005). Furthermore, heuristic approximations to these optimal proposals, or indeed Rao–Blackwellisation strategies, are also possible and often have good empirical performance (Doucet et al., 2000).

Extensions of adaptation using potential information beyond the current time have been explored in look-ahead methods (Lin et al., 2013) and typically employ approximations as the exact schemes are intractable. Recently, methods for iterative approximations to the optimal change of measure for the entire path space have seen interest. These rely on twisted Feynman–Kac models which generalise locally optimal proposals and look-ahead methods (Guarniero et al., 2017; Heng et al., 2020). In theory, applying a particle filter to an optimally twisted Feynman–Kac model results in a zero variance estimate of the model's normalising constant. Whilst in practice, iteratively estimating these models has been shown to dramatically reduce the variance for various test functions.

SMC samplers can also benefit from the aforementioned adaptation strategies but also have other degrees of freedom that are more prevalent in the literature on variance reduction (Del Moral et al., 2006). Moreover, it is often more difficult to implement any exact adaptation strategies in SMC samplers as the weights take a more complex form, though Dau and Chopin (2022) show this is possible in certain settings. Despite this, twisted Feynman–Kac models have been used successfully in SMC samplers (Heng et al., 2020).

A major limitation in the study of optimal proposals, exact adaptation, and Rao–Blackwellisation is their theoretical underpinning. These methods are typically justified by appealing to minimisation of the conditional variance of the importance weights, or reducing variance of the joint weight over the entire path space. They do not give a theoretical guarantee on variance reductions for particle approximations of a test function, arising from a particle filter with resampling. Approximations and heuristics motivated by these methods are also subject to this unsatisfactory understanding. Further, methods using twisted Feynman–Kac models are only optimal for the normalising constant estimate of the model in the idealised version. Achieving this in practice is not guaranteed, though empirical performance can be strong at the cost of additional computation.

This paper contributes knots for Feynman–Kac models, a technique to reduce the asymptotic variance for particle approximations. We generalise and unify 'full' adaptation and Rao–Blackwellisation, improving the theoretical underpinning for these methods, whilst providing a highly flexible strategy for designing new algorithms with guaranteed asymptotic variance reduction. Further, we resolve the discrepancy between the theoretical analysis of fully-adapted auxiliary particle filters and their demonstrated empirical performance. We demonstrate that a small change to particle filters with 'full' adaptation can guarantee an asymptotic variance reduction, resolving the counterexample in Johansen and Doucet (2008).

Our techniques lead naturally to a partial ordering on general Feynman–Kac models whose order implies superiority in terms of the asymptotic variance of particle approximations — a first for variance ordering of SMC algorithms by their underlying Feynman–Kac model. Further, we determine optimal knots and optimal sequences of knots to assist with SMC algorithm design.

The paper is structured as follows. We first review the background of SMC algorithms, including Feynman–Kac models and the asymptotic variance of particle approximations in Section 2. Knots are introduced in Section 3 where we discuss their properties as well as invariances and equivalences of models before and after the application of a knot. Section 4 contains our main variance reduction results for knots, whereas Section 5 discusses the optimality of knots. Section 6 provides variance and optimality results for terminal knots which require special treatment. Lastly, Section 7 contains examples including 'full' adaptation and Rao–Blackwellisation as special cases of knots, and an illustrative example that is a hybrid between these two cases. Proofs of results are deferred to the appendix.

2 Background

Feynman–Kac models are path measures that can represent the evolution of particles generated by sequential Monte Carlo algorithms. A Feynman–Kac model can be constructed by weighting a Markov process. We consider algorithms for discrete-time models and hence restrict focus to a process specified by an initial distribution, a sequence of non-homogeneous Markov kernels, and potential functions for weights. Before describing these discrete-time Feynman–Kac models, we first introduce our notation.

2.1 Notation

Let $y_{i:j}$ be the vector $(y_i, y_{i+1}, \ldots, y_j)$ when $i < j$ and $y_{k:k} = y_k$. For integers $n_0 < n_1$, we denote the set of integers as $[n_0 : n_1] = \{n_0, n_0 + 1, \ldots, n_1\}$ and write the set of natural numbers as $\mathbb{N}_0 = \{0, 1, \ldots\}$ and $\mathbb{N}_1 = \{1, 2, \ldots\}$. The function mapping any input to unit value is denoted by $1(\cdot)$ and the indicator function for a set S is 1_S. If f and g are functions,

then $f \cdot g$ defines the map $x \mapsto f(x)g(x)$ whilst $f \otimes g = (x, y) \mapsto f(x)g(y)$. We denote the zero set for a function $f : \mathsf{X} \to \mathbb{R}$ as $S_0(f) = \{x \in \mathsf{X} : f(x) = 0\}$.

Let $(\mathsf{X}, \mathcal{X})$ be a measurable space. If μ is a measure on $(\mathsf{X}, \mathcal{X})$ and the function $\varphi : \mathsf{X} \to \mathbb{R}$ then let $\mu(\varphi) = \int \varphi(x)\mu(\mathrm{d}x)$ and if $S \in \mathcal{X}$ then let $\mu(S) = \mu(1_S)$. We use $\mathcal{L}(\mu)$ to denote the class of functions that are integrable with respect to a measure μ.

In addition to $(\mathsf{X}, \mathcal{X})$, let $(\mathsf{Y}, \mathcal{Y})$ be a measurable space. When referring to a kernel we consider non-negative kernels, say $K : (\mathsf{X}, \mathcal{Y}) \to [0, \infty)$. We define $K(\varphi)(\cdot) = \int K(\cdot, \mathrm{d}y)\varphi(y)$ for a function $\varphi : \mathsf{Y} \to \mathbb{R}$, $\mu K(\cdot) = \int \mu(\mathrm{d}x)K(x, \cdot)$ for a measure μ on $(\mathsf{X}, \mathcal{X})$, and the tensor product as $(\mu \otimes K)(\mathrm{d}[x, y]) = \mu(\mathrm{d}x)K(x, \mathrm{d}y)$. The composition of two non-negative kernels, say $L : (\mathsf{W}, \mathcal{X}) \to [0, \infty)$ and K, is defined as $LK(w, \cdot) = \int L(w, \mathrm{d}x)K(x, \cdot)$ and is a right-associative operator, whilst the tensor product is $(L \otimes K)(w, \mathrm{d}[x, y]) = L(w, \mathrm{d}x)K(x, \mathrm{d}y)$. A Markov kernel is a non-negative kernel K such that $K(x, \mathsf{Y}) = 1$ for all $x \in \mathsf{X}$. We denote the identity kernel by Id and the degenerate probability measure at x as δ_x.

We make use of twisted Markov kernels and will use the following superscript notation when describing these. If $K : (\mathsf{X}, \mathcal{Y}) \to [0, 1]$ is a Markov kernel and $H : \mathsf{Y} \to [0, \infty)$ is integrable w.r.t. $K(x, \cdot)$ for all $x \in \mathsf{X}$, let $K^H : (\mathsf{X}, \mathcal{Y}) \to [0, 1]$ be defined such that

$$K^H(x, \mathrm{d}y) = \begin{cases} \frac{K(x, \mathrm{d}y)H(y)}{K(H)(x)} & \text{if } K(H)(x) > 0, \\ K(x, \mathrm{d}y) & \text{if } K(H)(x) = 0. \end{cases}$$

The choice to use K when $K(H)(x) = 0$ is somewhat arbitrary, but useful for formalising our mathematical arguments. Particle filter approximations do not depend on this choice. Similarly, if μ is a probability measure on $(\mathsf{Y}, \mathcal{Y})$ then μ^H will be defined as $\mu^H(\mathrm{d}y) = \frac{\mu(\mathrm{d}y)H(y)}{\mu(H)}$ if $\mu(H) > 0$ and $\mu^H(\mathrm{d}y) = \mu(\mathrm{d}y)$ if $\mu(H) = 0$. If K is the identity kernel or degenerate probability measure, we take $K^H = K$ by convention.

The categorical distribution is denoted by $\mathcal{C}(q_1, q_2 \ldots, q_m)$ defined with positive weights $(q_1, q_2 \ldots, q_m)$ on support $[1 : m]$ with probabilities $p_i = \left(\sum_{j=1}^m q_j\right)^{-1} q_i$ for $i \in [1 : m]$.

2.2 Discrete-time Feynman–Kac models

To define a Discrete-time Feynman–Kac model, we require a notion of compatible kernels, we refer to as composability. Composability is also used when we define knots in Section 3.1.

Definition 2.1 (Composability)**.** *Let Y_p be a space and $(\mathsf{X}_p, \mathcal{X}_p)$ be a measurable space for $p \in \{1, 2\}$. Let $M_1 : (\mathsf{Y}_1, \mathcal{X}_1) \to [0, 1]$ be a non-negative kernel (or measure, $M_1 : \mathcal{X}_1 \to [0, 1]$) and $M_2 : (\mathsf{Y}_2, \mathcal{X}_2) \to [0, 1]$ be a non-negative kernel. If $\mathsf{X}_1 \subseteq \mathsf{Y}_2$ then $M_1 M_2$ is well-defined and we say that M_1 and M_2 are composable.*

Consider measurable spaces $(\mathsf{X}_p, \mathcal{X}_p)$ for $p \in [0 : n]$, then let $M_0 : \mathcal{X}_0 \to [0, 1]$ be a probability measure, $M_p : (\mathsf{X}_{p-1}, \mathcal{X}_p) \to [0, 1]$ for $p \in [1 : n]$ be Markov kernels, and consider

potential functions $G_p : \mathsf{X}_p \to [0, \infty)$ that are integrable with respect to $M_p(x, \cdot)$ for all $x \in \mathsf{X}_{p-1}$ and $p \in [0:n]$.

Definition 2.2 (Discrete-time Feynman–Kac model)**.** *A predictive Feynman–Kac model with horizon $n \in \mathbb{N}_0$ consists of an initial distribution M_0, Markov kernels M_p for $p \in [1:n]$, and potential functions G_p for $p \in [0:n-1]$ such that M_{p-1} and M_p are composable for $p \in [1:n]$. In addition to the above, an updated Feynman–Kac model includes a potential G_n at the terminal time.*

We will refer to a generic updated Feynman–Kac model with calligraphic notation $\mathcal{M}$ or specifically the collection $(M_{0:n}, G_{0:n})$. This specification includes both predictive and updated Feynman–Kac models, by taking $G_n = 1$ for the former. We will use $\mathscr{M}_n$ to denote the class of discrete-time Feynman–Kac models with horizon n.

The initial measure, kernels, and potentials define a sequence of predictive measures starting with $\gamma_0 = M_0$, and evolving by

$$\gamma_{p+1} = \int \gamma_p(\mathrm{d}x_p) G_p(x_p) M_{p+1}(x_p, \cdot) \tag{1}$$

for $p \in [0:n-1]$. The terminal measure can be thought of as the expectation over a path space, that is

$$\gamma_n(\varphi) = \mathbb{E}\left[\varphi(X_n) \prod_{p=1}^{n-1} G_p(X_p)\right] \tag{2}$$

with respect to a non-homogeneous Markov chain, specified by $X_0 \sim M_0$ and $X_p \sim M_p(X_{p-1}, \cdot)$ for $p \in [1:n]$.

In comparison, the time-p updated measures use potential information at time p and are defined by $\hat{\gamma}_p(\mathrm{d}x_p) = \gamma_p(\mathrm{d}x_p) G_p(x_p)$ for $p \in [0:n]$. The predictive and updated measures have normalised counterparts

$$\eta_p(\mathrm{d}x_p) = \frac{\gamma_p(\mathrm{d}x_p)}{\gamma_p(1)}, \quad \hat{\eta}_p(\mathrm{d}x_p) = \frac{\hat{\gamma}_p(\mathrm{d}x_p)}{\hat{\gamma}_p(1)},$$

which are probability measures for $p \in [0:n]$. The path space representation of the updated terminal measure can be expressed by considering $\hat{\gamma}_n(\varphi) = \gamma_n(G_n \cdot \varphi)$ using (2).

Lastly, an important quantity for the asymptotic variance calculation are the $Q_{p,n}$ kernels are defined as follows. Consider kernels $Q_p(x_{p-1}, \mathrm{d}x_p) = G_{p-1}(x_{p-1}) M_p(x_{p-1}, \mathrm{d}x_p)$ for $p \in [1:n]$ then let $Q_{n,n} = \mathrm{Id}$, $Q_{n-1,n} = Q_n$, and continue with $Q_{p,n} = Q_{p+1} \cdots Q_n = Q_{p+1} Q_{p+1,n}$ for $p \in [0:n-2]$. In contrast to the expectation presented in (2), the $Q_{p,n}$ kernels are conditional expectations on the same path space, that is for $p \in [0:n-1]$

$$Q_{p,n}(\varphi)(x_p) = \mathbb{E}\left[\varphi(X_n) \prod_{t=p}^{n} G_t(X_t) \;\middle|\; X_p = x_p\right].$$

In other words, at time p, the $Q_{p,n}$ complete the model in the sense that $\gamma_p Q_{p,n} = \gamma_n$.

2.3 SMC and particle filters

Sequential Monte Carlo algorithms, and in particular particle filters, approximate Feynman–Kac models by iteratively generating collections of points, denoted by ζ_p^i for $i \in [1:N]$, to approximate the sequence of probability measures η_p for $p \in [0:n]$. We consider the bootstrap particle filter (Gordon et al., 1993) to approximate the terminal measure η_n or its updated counterpart, which we simply refer to as a particle filter and describe in Algorithm 1. Different particle filters can be achieved by varying the underlying Feynman–Kac model whilst preserving the targets of the particle approximations of interest. After running

Algorithm 1 A Particle Filter

Input: Feynman–Kac model $\mathcal{M} = (M_{0:n}, G_{0:n})$

1. Sample initial $\zeta_0^i \overset{\text{iid}}{\sim} M_0$ for $i \in [1:N]$
2. For each time $p \in [1:n]$
   a. Sample ancestors $A_{p-1}^i \sim \mathcal{C}\left(G_{p-1}(\zeta_{p-1}^1), \ldots, G_{p-1}(\zeta_{p-1}^N)\right)$ for $i \in [1:N]$
   b. Sample prediction $\zeta_p^i \sim M_p(\zeta_{p-1}^{A_{p-1}^i}, \cdot)$ for $i \in [1:N]$

Output: Terminal particles $\zeta_n^{1:N}$

a particle filter the approximate terminal predictive measures are

$$\eta_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\zeta_n^i}, \quad \gamma_n^N = \left\{\prod_{t=0}^{n-1} \eta_t^N(G_t)\right\} \eta_n^N.$$

Similarly, the updated terminal measures are

$$\hat{\eta}_n^N = \sum_{i=1}^N W_n^i \delta_{\zeta_p^i}, \quad \hat{\gamma}_n^N = \left\{\prod_{t=0}^{n} \eta_t^N(G_t)\right\} \hat{\eta}_n^N,$$

where $W_n^i = \frac{G_n(\zeta_n^i)}{\sum_{j=1}^N G_n(\zeta_n^j)}$.

2.4 Asymptotic variance of particle approximations

The canonical asymptotic variance map $\sigma^2 : \mathcal{L}(\eta_n) \to [0, \infty]$ is defined as

$$\sigma^2(\varphi) = \sum_{p=0}^n v_{p,n}(\varphi), \quad v_{p,n}(\varphi) = \frac{\gamma_p(1)\gamma_p(Q_{p,n}(\varphi)^2)}{\gamma_n(1)^2} - \eta_n(\varphi)^2, \tag{3}$$

for a particle filter following Algorithm 1. Under various conditions, particle approximations relate to this variance by way of Central Limit Theorems (CLT, Del Moral, 2004). CLTs establish that the terminal predictive measure, $\gamma_n^N(\varphi) \to \gamma_n(\varphi)$ in probability as $N \to \infty$, and

$$\sqrt{N}\left(\frac{\gamma_n^N(\varphi) - \gamma_n(\varphi)}{\gamma_n(1)}\right) \to \mathcal{N}\left(0, \sigma^2(\varphi)\right),$$

in distribution. Whilst for the terminal predictive probability measure, $\eta_n^N(\varphi) \to \eta_n(\varphi)$ in probability as $N \to \infty$, and $\sqrt{N}\left(\eta_n^N(\varphi) - \eta_n(\varphi)\right) \to \mathcal{N}\left(0, \sigma^2(\varphi - \eta_n(\varphi))\right)$ in distribution. It is also possible to establish convergence in moments (Lee and Whiteley, 2018). In this case, $\gamma_n^N(\varphi) \to \gamma_n(\varphi)$ and $\eta_n^N(\varphi) \to \eta_n(\varphi)$ almost surely, whilst

$$N \operatorname{Var}\left\{\gamma_n^N(\varphi)/\gamma_n(1)\right\} \to \sigma^2(\varphi), \quad N\,\mathbb{E}\left[\left\{\eta_n^N(\varphi) - \eta_n(\varphi)\right\}^2\right] \to \sigma^2(\varphi - \eta_n(\varphi)).$$

When $\sigma^2(\varphi)$ is finite, CLTs and convergence statements motivate the use of the asymptotic variance map to characterise the theoretical performance of a particle filter.

We analyse the asymptotic variance map directly, assuming that $\sigma^2(\varphi)$ is finite, without additional assumptions. As such, for the predictive measure of a given Feynman–Kac model $\mathcal{M}$, we will consider functions $\varphi \in \mathcal{F}(\mathcal{M})$ such that $\mathcal{F}(\mathcal{M}) = \{\varphi \in \mathcal{L}(\eta_n) : \sigma^2(\varphi) < \infty\}$. To interpret our results, we will require a CLT to hold for the model $\mathcal{M}$ in question. For general Feynman–Kac models, for example, bounded potential functions and a bounded test function would suffice, but this need not be the case.

When applicable, analogous CLTs also hold for the updated marginal (probability) measures by using $\hat{\sigma}^2(\varphi)$ and $\hat{\sigma}^2(\varphi - \hat{\eta}_n(\varphi))$ in place of $\sigma^2(\varphi)$ and $\sigma^2(\varphi - \eta_n(\varphi))$ respectively, where

$$\hat{\sigma}^2(\varphi) = \sum_{p=0}^{n} \hat{v}_{p,n}(\varphi), \quad \hat{v}_{p,n}(\varphi) = \frac{v_{p,n}(G_n \cdot \varphi)}{\eta_n(G_n)^2}. \tag{4}$$

For updated measures, our analysis then considers the class of test functions $\hat{\mathcal{F}}(\mathcal{M}) = \{\varphi \in \mathcal{L}(\hat{\eta}_n) : \hat{\sigma}^2(\varphi) < \infty\}$. In our discussions, we will refer to functions in the classes $\mathcal{F}(\mathcal{M})$ and $\hat{\mathcal{F}}(\mathcal{M})$ as relevant test functions.

3 Tying knots in Feynman–Kac models

In order to present our procedure for reducing the variance of SMC algorithms, we must define how the underlying Feynman–Kac model is modified. To this end we define a *knot*, encoding the details of the modification, and the *knot operator* which describes how a knot is applied to a Feynman–Kac model.

A knot is specified by a time, t, and composable Markov kernels, R and K, which can be used to modify suitable Feynman–Kac models whilst preserving the terminal measure. In one view, a t-knot modifies a Feynman–Kac model by partially adapting the Markov kernel M_t to potential information at time t, though repeated applications of knots allow adaptation to potential information beyond just the next time.

We will introduce knots formally in Section 3.1 and a convenient notion for their simultaneous application, *knotsets*, in Section 3.2. Knots at time n are considered later, in Section 6, as *terminal knots* require special treatment and have a smaller scope compared to regular knots (time $t < n$).

3.1 Knots

We begin with a formal definition of a knot and what it means for a knot to be compatible with a Feynman–Kac model.

Definition 3.1 (Knot)**.** *A knot is a triple $\mathcal{K} = (t, R, K)$, consisting of a time $t \in \mathbb{N}_0$, and an ordered pair of composable Markov kernels R and K. Note that when $t = 0$, R is a probability distribution.*

For compactness we will often refer to knots abstractly as $\mathcal{K}$, meaning $\mathcal{K} = (t, R, K)$ for some t, R and K, and when emphasis on the time component of the knot is required, we will refer to $\mathcal{K}$ a t-knot.

Definition 3.2 (Knot compatibility)**.** *For $t < n$, a knot $\mathcal{K} = (t, R, K)$ is compatible with a Feynman–Kac model $\mathcal{M} = (M_{0:n}, G_{0:n})$ if $M_t = RK$.*

To describe how knots act on Feynman–Kac models, we first consider the domain of the requisite operator. Recall the set of all Feynman–Kac models with horizon n as $\mathscr{M}_n$. If we let $\mathscr{K}$ be the set of all possible knots, we can define the set of all compatible knots and Feynman–Kac models as

$$\mathscr{D}_n = \{(\mathcal{K}, \mathcal{M}) \in \mathscr{K} \times \mathscr{M}_n : \mathcal{K} \text{ and } \mathcal{M} \text{ are compatible}\} \tag{5}$$

for $n \in \mathbb{N}_1$. We define the knot operator as a right-associative operator acting on elements of this set.

Definition 3.3 (Knot operator)**.** *The knot operator maps compatible knot-model pairs to the space of Feynman–Kac models for horizon $n \in \mathbb{N}_1$ and is denoted by $* : \mathscr{D}_n \to \mathscr{M}_n$. We use the infix notation $\mathcal{K} * \mathcal{M}$ for convenience. For a knot $\mathcal{K} = (t, R, K)$ and model $\mathcal{M} = (M_{0:n}, G_{0:n})$, the knot-model $\mathcal{K} * \mathcal{M} = (M^*_{0:n}, G^*_{0:n})$ where*

$$M^*_t = R, \quad G^*_t = K(G_t), \quad M^*_{t+1} = K^{G_t} M_{t+1}.$$

The remaining Markov kernels and potential functions are identical to the original model, that is $M^*_p = M_p$ for $p \notin \{t, t+1\}$ and $G^*_p = G_p$ for $p \neq t$.

The knot operator preserves the terminal predictive and updated measures of the Feynman–Kac model, the predictive and updated path measures (see Proposition 3.13). Besides preserving key measures, the knot operator preserves the horizon of the Feynman–Kac model it is applied to, which is crucial for our comparisons of the asymptotic variance of particle estimates with and without knots. Figure 1 illustrates the transformation of a Feynman–Kac model to a new model by a knot.

To motivate our consideration of knots, we state a simplified variance reduction theorem.

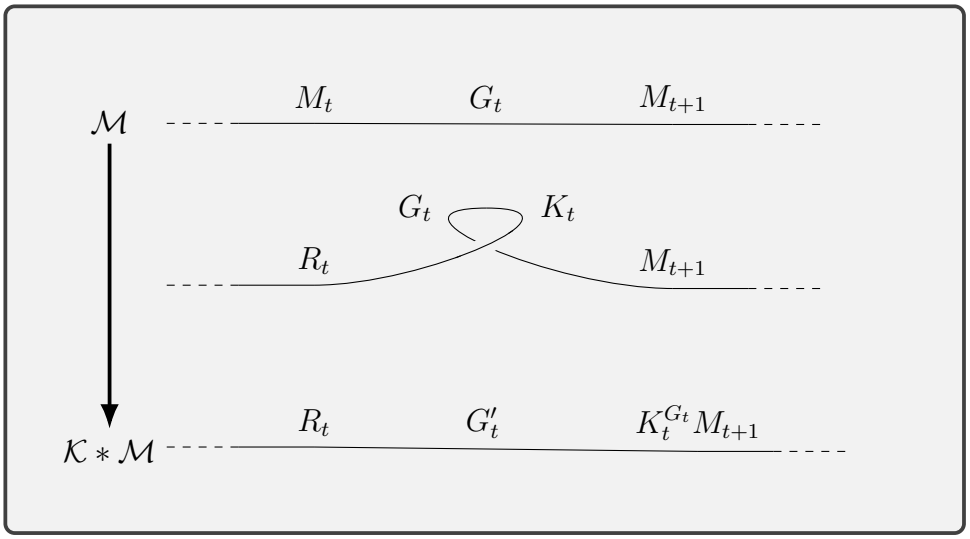

Figure 1: Effect of a knot $\mathcal{K} = (t, R_t, K_t)$ at time $t \in [0:n-1]$ on model $\mathcal{M}$ with $M_t = R_t K_t$. Note that $G'_t = K_t(G_t)$.

Theorem 3.4 (Variance reduction from a knot)**.** *Consider models $\mathcal{M}$ and $\mathcal{M}^* = \mathcal{K} * \mathcal{M}$ for a knot $\mathcal{K}$. Let $\mathcal{M}$ and $\mathcal{M}^*$ have terminal measures γ_n and γ_n^* with asymptotic variance maps σ^2 and σ_*^2 respectively. If $\varphi \in \mathcal{F}(\mathcal{M})$ then the terminal measures are equivalent, $\gamma_n(\varphi) = \gamma_n^*(\varphi)$, whilst the variances satisfy $\sigma_*^2(\varphi) \leq \sigma^2(\varphi)$.*

It is simple to show that Theorem 3.4 implies the same asymptotic variance inequality for the marginal updated measures as well as their normalised counterparts. As such, a model with a knot has terminal time particle approximations with better variance properties than the original model. We defer our proof to Theorem 4.1 which considers the general case with multiple knots.

The simplest possible knot is the trivial knot, in the sense that applying a trivial knot to a Feynman–Kac model does not change the model. The trivial knot is described in Example 3.5.

Example 3.5 (Trivial knot)**.** *Consider a model $\mathcal{M} = (M_{0:n}, G_{0:n})$ for $n \in \mathbb{N}_1$ and knot $\mathcal{K}$ at time $t \in [0:n-1]$. If $\mathcal{K} = (t, M_t, \mathrm{Id})$ then it is trivial in the sense that $\mathcal{K} * \mathcal{M} = \mathcal{M}$.*

Trivial knots do not change how the information at time t (the potential G_t) is incorporated into the Feynman–Kac model and do not change the asymptotic variance. On the other hand, we can define an adapted knot which fully adapts M_t to the information at time t. In fact, any knot can be thought of living on a spectrum between a trivial knot and an adapted knot. We discuss the optimality of adapted knots in Section 5.

Example 3.6 (Adapted knot)**.** *Consider a model $\mathcal{M} = (M_{0:n}, G_{0:n})$ for $n \in \mathbb{N}_1$ and knot $\mathcal{K}$ at time $t \in [1:n-1]$. If $\mathcal{K} = (t, \mathrm{Id}, M_t)$ we say it is an adapted knot for $\mathcal{M}$. The model $\mathcal{K} * \mathcal{M}$*

has new kernels and potential function, $M_t^* = \mathrm{Id}$*,* $M_{t+1}^* = M_t^{G_t} M_{t+1}$*, and* $G_t^* = M_t(G_t)$*. For* $t = 0$*, an adapted knot has the form* $\mathcal{K} = (0, \delta_0, K_0)$ *where the kernel* K_0 *satisfies* $K_0(0, \cdot) = M_0$*, whilst* $M_0^* = \delta_0$*,* $M_1^* = M_0^{G_0} M_1$*, and* $G_0^* = M_0(G_0)$*.*

At time t, an adapted knot results in a kernel of the form $M_{t+1}^* = M_t^{G_t} M_{t+1}$ where $M_t^{G_t}$ is now adapted to the information in G_t. One might argue that a more natural representation of such an adaptation would use $M_t^{G_t}$ as the tth kernel of the new model, not as a component of the $(t+1)$th kernel. However, our definition of knots is precisely what allows for an ordering of the asymptotic variance terms.

3.2 Knotsets

A knot is the elementary operator we consider, in the sense that it is the minimal modification of a Feynman–Kac model for which we can prove a variance reduction. However, it is natural to consider a set of knots acting on many time points of a model. Further, knots can be applied sequentially, but it is convenient to consider a set of knots that can be applied simultaneously. As such, we will now define a generalisation of knots and their associated operator, the *knotset* and *knotset operator*.

Definition 3.7 (Knotsets and compatibility)**.** *A knotset* $\mathcal{K} = (R_{0:n-1}, K_{0:n-1})$ *is specified by* n *knots of the form* $\mathcal{K}_p = (p, R_p, K_p)$ *for* $p \in [0\,{:}\,n-1]$*. Such a knotset is compatible with* $\mathcal{M} \in \mathscr{M}_n$ *if each* (p, R_p, K_p)*-knot is compatible with* $\mathcal{M}$ *for* $p \in [0\,{:}\,n-1]$*.*

Definition 3.8 (Knotset operator)**.** *Let* $\mathcal{K} = (R_{0:n-1}, K_{0:n-1})$ *be a knotset compatible with* $\mathcal{M} \in \mathscr{M}_n$*. The knotset operation is defined as* $\mathcal{K} * \mathcal{M} = \mathcal{K}_0 * \mathcal{K}_1 * \cdots * \mathcal{K}_{n-1} * \mathcal{M}$ *where* $\mathcal{K}_p = (p, R_p, K_p)$ *for* $p \in [0\,{:}\,n-1]$*.*

The knotset operator is defined to apply n knots, with unique times, in descending order so that the compatibility condition for each knot does not change after each successive knot application. This design also allows us to frame knotsets as a simultaneously application of n knots to a model, which is presented next.

Proposition 3.9 (Knot-model)**.** *If* $\mathcal{K} = (R_{0:n-1}, K_{0:n-1})$ *is a knotset compatible with model* $\mathcal{M} = (M_{0:n}, G_{0:n})$*, then* $\mathcal{K} * \mathcal{M} = (M_{0:n}^*, G_{0:n}^*)$ *satisfies*

$$\begin{aligned} M_0^* &= R_0, & G_0^* &= K_0(G_0), \\ M_p^* &= K_{p-1}^{G_{p-1}} R_p, & G_p^* &= K_p(G_p), \quad p \in [1\,{:}\,n-1] \\ M_n^* &= K_{n-1}^{G_{n-1}} M_n, & G_n^* &= G_n. \end{aligned}$$

We refer to $\mathcal{M}^* = \mathcal{K} * \mathcal{M}$ for a knotset $\mathcal{K}$ as a knot-model and provide the form of $\mathcal{M}^*$ in Proposition 3.9. The proof is trivial due to the descending time-ordering of

knot applications specified in Definition 3.8. The knotset operator also inherits right-associativity from knots.

We illustrate two examples, trivial and adapted knotsets, that extend Example 3.5 and 3.6 respectively. The trivial knotset consists of n trivial knots, as such it does not change the Feynman–Kac model nor alter the asymptotic variance.

Example 3.10 (Trivial knotset)**.** *Consider a knotset* $\mathcal{K} = (M_{0:n-1}, K_{0:n-1})$ *where* $K_p = \mathrm{Id}$ *for* $p \in [0:n-1]$ *applied to model* $\mathcal{M} = (M_{0:n}, G_{0:n})$*. The resulting model is* $\mathcal{K} * \mathcal{M} = \mathcal{M}$*.*

We can also use n adapted knots to form an adapted knotset, which we describe in Example 3.6.

Example 3.11 (Adapted knotset)**.** *Consider a knotset* $\mathcal{K} = (R_{0:n-1}, K_{0:n-1})$ *such that each* (p, R_p, K_p)*-knot is an adapted knot for* $\mathcal{M} = (M_{0:n}, G_{0:n})$*. The adapted model is* $\mathcal{K} * \mathcal{M} = (M^*_{0:n}, G^*_{0:n})$ *where* $M^*_0 = \delta_0$*,* $M^*_1(0, \cdot) = M_0^{G_0}$*,* $M^*_p = M_{p-1}^{G_{p-1}}$ *for* $p \in [2:n-1]$*, and* $M^*_n = M_{n-1}^{G_{n-1}} M_n$*. Whilst the potentials satisfy* $G^*_p = M_p(G_p)$ *for* $[0:n-1]$*, and* $G^*_n = G_n$*.*

Adapted knotsets are related to fully-adapted auxiliary particle filters (Pitt and Shephard, 1999; Johansen and Doucet, 2008) but differ subtly. We discuss this class of knots and its relation to existing particle filters in Section 7.1.

The model in Example 3.11 has redundancy since $M^*_0 = \delta_0$ and $M^*_1(0, \cdot) = M_0^{G_0}$ and $G^*_0 = M_0(G_0)$ is a constant. This is an artifact of the knot operator which preserves the time horizon of the model and is essential for comparing the asymptotic variance terms. Using such the adapted knot-model in practice, one would ignore the initial transitions and begin the particle filter at time $p = 1$. If required, the constant potential function at time $p = 0$ can be accounted for including it in the potential function at time $p = 1$.

Though knotsets can change the Feynman–Kac model they are applied to, some quantities remain unchanged, whilst others can be expressed in terms of the original model. We demonstrate these invariances and equivalences now. In order to distinguish quantities related to the original model, $\mathcal{M}$, and to the relevant knot-model, $\mathcal{M}^*$, we embellish the latter with the same superscript. For example, the terminal measure of $\mathcal{M}$ will be denoted by γ_n, whilst we will use γ^*_n for $\mathcal{M}^*$.

Proposition 3.12 (Knot-model predictive measures)**.** *Let* $\mathcal{K}$ *be a* $(R_{0:n-1}, K_{0:n-1})$*-knotset and consider knot-model* $\mathcal{M}^* = \mathcal{K} * \mathcal{M}$*. For measurable* φ*, the knot-model* $\mathcal{M}^*$ *will have predictive marginal measures such that*

1. $\gamma^*_0(\varphi) = R_0(\varphi)$ *and* $\gamma^*_p(\varphi) = \hat{\gamma}_{p-1} R_p(\varphi)$ *for* $p \in [1:n-1]$*.*
2. $\gamma^*_p K_p(\varphi) = \gamma_p(\varphi)$ *for* $p \in [0:n-1]$*.*

Aside from establishing the connection between a model and its counterpart with knots, Part 1 of Proposition 3.12 is later used to compare the asymptotic variance of particle

approximations from the use of each Feynman–Kac model in an SMC algorithm. Part 2 indicates that if any marginal measures in the original model are of interest we can approximate these with one additional step, even when using the model with knots.

Proposition 3.13 (Knot-model invariants)**.** *Let $\mathcal{K}$ be a $(R_{0:n-1}, K_{0:n-1})$-knotset and consider knot-model $\mathcal{M}^* = \mathcal{K} * \mathcal{M}$. For measurable φ, the knot-model $\mathcal{M}^*$ will have the following invariants:*

1. *Terminal marginal measures, $\gamma_n^*(\varphi) = \gamma_n(\varphi)$ and $\hat{\gamma}_n^*(\varphi) = \hat{\gamma}_n(\varphi)$.*
2. *Terminal probability measures, $\eta_n^*(\varphi) = \eta_n(\varphi)$ and $\hat{\eta}_n^*(\varphi) = \hat{\eta}_n(\varphi)$.*
3. *Normalising constants, $\gamma_p^*(1) = \gamma_p(1)$ and $\hat{\gamma}_p^*(1) = \hat{\gamma}_p(1)$ for all $p \in [0:n]$.*

Proposition 3.13 establishes that the terminal measure is unchanged by knots, hence a model and its counterpart with knots can be used to estimate the same quantities. We will also make use of the invariants when making asymptotic variance comparisons.

In subsequent sections we will use the terms knots and knotset synonymously. We note that a knotset is a strict generalisation of a (t, R, K)-knot, which can be seen by taking the underlying knots $\mathcal{K}_p$ to be trivial for all $p \neq t$ in Proposition 3.9. As such, results for knotsets apply directly to knots. More practically, we can also use a knotset to describe only $m \in [0:n]$ knots by letting $n - m$ knots be trivial. This is useful if no suitable knot can be defined for one or more time points.

In Section 4 we show that terminal particle approximations using $\mathcal{K} * \mathcal{M}$ have lower asymptotic variance than their counterparts using $\mathcal{M}$. By the invariance property in Proposition 3.13 this indicates better particle approximations exist for the same quantities of interest when knots can be implemented.

4 Variance reduction and ordering from knots

Having established the equivalence of all terminal marginal measures in Proposition 3.13, we now consider the asymptotic variances of particle approximations to these quantities. Our main result is given in Theorem 4.1 where we state that applying knots to a Feynman–Kac model reduces the variance of particle approximations for all relevant functions. We will denote the predictive and updated asymptotic variance of the knot-model $\mathcal{M}^*$ by σ_*^2 and $\hat{\sigma}_*^2$ respectively.

Theorem 4.1 (Variance reduction with knots)**.** *Consider models $\mathcal{M} \in \mathscr{M}_n$ and $\mathcal{M}^* = \mathcal{K} * \mathcal{M}$ for knotset $\mathcal{K} = (R_{0:n-1}, K_{0:n-1})$. If $\varphi \in \mathcal{F}(\mathcal{M})$ then $\sigma_*^2(\varphi) \leq \sigma^2(\varphi)$ and the reduction in the variance is*

$$\sigma^2(\varphi) - \sigma_*^2(\varphi) = \sum_{p=0}^{n-1} \frac{\gamma_p(1)^2}{\gamma_n(1)^2} \nu_p \left\{ \mathrm{Var}_{K_p} \left[Q_{p,n}(\varphi) \right] \right\},$$

where $\nu_p = \hat{\eta}_{p-1} R_p$ *for* $p \in [1:n-1]$ *and* $\nu_0 = R_0$. *Moreover, the variance ordering is strict if there exists a time* $p \in [0:n-1]$ *such that* $\nu_p \left\{ \mathrm{Var}_{K_p} \left[Q_{p,n}(\varphi) \right] \right\} > 0$.

From Theorem 4.1 we can see that, loosely speaking, the variance is strictly reduced if $Q_{p,n}(\varphi)$ is not constant relative to K_p. As expected, degenerate K_p do not reduce the variance as we previously stated for the trivial knotset with $K_p = \mathrm{Id}$.

Note that the variance reduction excludes a contribution from time n due to the absence of a knot at the terminal time. We can also define a terminal time (n, R, K)-knot analogously to the knots discussed thus far. However, such a terminal knot will only guarantee a variance reduction of the normalising constant estimate, $\hat{\gamma}_n(1)$. We discuss terminal knots and how to achieve variance reductions for specific test functions in Section 6.

Theorem 4.1 is our main result, applying directly to predictive measures. The variance reduction result is extended to the remaining terminal measures by Corollary 4.2.

Corollary 4.2 (Knot variance reduction with knots)**.** *Under the conditions of Theorem 4.1, the following asymptotic variance inequalities hold.*

1. *Predictive terminal probability measure:* $\sigma_*^2(\varphi - \eta_n^*(\varphi)) \leq \sigma^2(\varphi - \eta_n(\varphi))$ *if* $\varphi \in \mathcal{F}(\mathcal{M})$.
2. *Updated terminal measure:* $\hat{\sigma}_*^2(\varphi) \leq \hat{\sigma}^2(\varphi)$ *if* $\varphi \in \hat{\mathcal{F}}(\mathcal{M})$.
3. *Updated terminal probability measure:* $\hat{\sigma}_*^2(\varphi - \hat{\eta}_n^*(\varphi)) \leq \hat{\sigma}^2(\varphi - \hat{\eta}_n(\varphi))$ *if* $\varphi \in \hat{\mathcal{F}}(\mathcal{M})$.

The inequalities are strict under the same conditions as Theorem 4.1.

The differences in the asymptotic variances stated in Corollary 4.2 are straightforward to derive using the quantitative result in Theorem 4.1 so we suppress them here.

Theorem 4.1 pertains to variance reduction from the application of one knotset to a Feynman–Kac model, however we can also consider multiple knotsets via iterative application. In doing so, we can establish a partial ordering of Feynman–Kac models induced by knots.

Definition 4.3 (A partial ordering of Feynman–Kac with knots)**.** *Consider two Feynman–Kac models,* $\mathcal{M}, \mathcal{M}^* \in \mathscr{M}_n$. *We say that* $\mathcal{M}^* \preccurlyeq \mathcal{M}$ *if there exists a sequence of knotsets* $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_m$ *such that* $\mathcal{M}^* = \mathcal{K}_m * \cdots * \mathcal{K}_1 * \mathcal{M}$ *for some* $m \in \mathbb{N}_1$.

From the above partial ordering we can state a general variance ordering results for sequential Monte Carlo algorithms. Note that each $\mathcal{K}_s$ in Definition 4.3 is required to be compatible with the knot-model resulting from $\mathcal{K}_{s-1} * \cdots * \mathcal{K}_1 * \mathcal{M}$.

Theorem 4.4 (Variance ordering with knots)**.** *Suppose* $\mathcal{M}^* \preccurlyeq \mathcal{M}$ *then* $\gamma_n^*(\varphi) = \gamma_n(\varphi)$, $\hat{\gamma}_n^*(\varphi) = \hat{\gamma}_n(\varphi)$, $\eta_n^*(\varphi) = \eta_n(\varphi)$, $\hat{\eta}_n^*(\varphi) = \hat{\eta}_n(\varphi)$, *and the following variance ordering results hold.*

1. *If* $\varphi \in \mathcal{F}(\mathcal{M})$ *then* $\sigma_*^2(\varphi) \leq \sigma^2(\varphi)$ *and* $\sigma_*^2(\varphi - \eta_n^*(\varphi)) \leq \sigma^2(\varphi - \eta_n(\varphi))$.

2. *If* $\varphi \in \hat{\mathcal{F}}(\mathcal{M})$ *then* $\hat{\sigma}_*^2(\varphi) \leq \hat{\sigma}^2(\varphi)$ *and* $\hat{\sigma}_*^2(\varphi - \hat{\eta}_n^*(\varphi)) \leq \hat{\sigma}^2(\varphi - \hat{\eta}_n(\varphi))$.

The inequalities are strict if at least one of the knotsets relating $\mathcal{M}^*$ *to* $\mathcal{M}$ *satisfy the conditions stated in Theorem 4.1.*

Our partial ordering result allows us to order the asymptotic variance of models related by multiple knots or knotsets. Such a result may be useful for some Feynman–Kac models in practice but will typically be more difficult to implement than a single knot or knotset. The partial ordering is, however, crucial to our exposition and proofs involving knot optimality in Section 5.

5 Optimality of adapted knots

Adapted knots and knotsets, introduced in Examples 3.6 and 3.11 respectively, possess optimality properties that distinguish them from other knots. Applying an adapted t-knot to a Feynman–Kac model results in the largest possible variance reduction of any single t-knot. Similarly, an adapted knotset will dominate any other knotset in terms of asymptotic variance reduction. This indicates that adapted knots should be appraised first before considering other types of knots that are compatible with the Feynman–Kac model at hand. The optimality of adapted knots and knotsets is expressed formally in Theorem 5.1.

Theorem 5.1 (Single adapted knot optimality)**.** *If* $\mathcal{K}$ *is a knotset (resp.* t*-knot) compatible with* $\mathcal{M}$*, and* $\mathcal{K}^\diamond$ *is the adapted knotset (resp. adapted* t*-knot) for* $\mathcal{M}$ *then* $\mathcal{K}^\diamond * \mathcal{M} \preccurlyeq \mathcal{K} * \mathcal{M}$.

Hence, in conjunction with Theorem 4.4, the asymptotic variance is lowest with adapted knots and knotsets. Further, from Proposition B.4 we can state that any sequence of t-knots applied to $\mathcal{M}$ is also dominated by the adapted t-knot applied to $\mathcal{M}$.

We can also deduce from Theorem 5.1 that the asymptotic variance can only be reduced beyond that of an adapted t-knot by using at least two non-trivial knots. For example, using the adapted t-knot followed by some other non-trivial knot at time $s \neq t$ would guarantee a further reduction in the variance.

Next we consider the case of multiple applications of knotsets by comparing models of the form $\mathcal{K}_m * \cdots * \mathcal{K}_1 * \mathcal{M}$ to the sequence of adapted knotsets applied to $\mathcal{M}$.

Theorem 5.2 (Multiple adapted knotset optimality)**.** *Let* $m \in \mathbb{N}_1$ *and consider two sequences of knotsets,* $\mathcal{K}_t$ *and* $\mathcal{K}_t^\star$*, over* $t \in [0\,{:}\,m-1]$*. Let* $\mathcal{M}_t = \mathcal{K}_{t-1} * \mathcal{M}_{t-1}$ *and* $\mathcal{M}_t^\star = \mathcal{K}_{t-1}^\star * \mathcal{M}_{t-1}^\star$ *for* $t \in [1\,{:}\,m]$ *and initial model* $\mathcal{M}_0 = \mathcal{M}_0^\star \in \mathscr{M}_n$*. If* $\mathcal{K}_t^\star$ *is the adapted knotset for* $\mathcal{M}_t^\star$ *for all* $t \in [0\,{:}\,m-1]$ *then* $\mathcal{M}_m^\star \preccurlyeq \mathcal{M}_m$.

Theorem 5.2 states that if $\mathcal{K}_t^\star$ are adapted knotsets for $t \in [1:m]$, then $\mathcal{K}_m^\star * \cdots * \mathcal{K}_1^\star * \mathcal{M} \preccurlyeq \mathcal{K}_m * \cdots * \mathcal{K}_1 * \mathcal{M}$, and hence a sequence of m adapted knotsets have a greater variance reduction than any other sequence of m knotsets, $\mathcal{K}_m$ for $t \in [1:m]$.

As adapted knotsets reduce the variance by the maximum amount of any knotset in each application, a natural question to ask is; to what extent can the asymptotic variance be reduced by repeated applications of adapted knotsets? Theorem 5.3 describes the minimal variance achievable by knots and/or knotsets.

Theorem 5.3 (Minimal variance from knots)**.** *For every model $\mathcal{M} \in \mathscr{M}_n$ there exists a sequence of knot(sets) $\mathcal{K}_n, \mathcal{K}_{n-1}, \ldots, \mathcal{K}_1$ such that the model*

$$\mathcal{M}^* = \mathcal{K}_n * \mathcal{K}_{n-1} * \cdots * \mathcal{K}_1 * \mathcal{M}$$

has asymptotic variance terms satisfying

$$v_{n,n}^*(\varphi) = v_{n,n}(\varphi), \quad v_{p,n}^*(\varphi) = 0, \text{ for } p \in [0:n-1]$$

for all $\varphi \in \mathcal{F}(\mathcal{M})$ where $v_{p,n}^*(\varphi)$ and $v_{p,n}(\varphi)$ are the asymptotic variance terms for $\mathcal{M}^*$ and $\mathcal{M}$ respectively.

Example 5.4 provides a (non-unique) construction of a sequence of knot-models, $\mathcal{M}_t$ for $t \in [1:n]$, where the corresponding $v_{p,n}(\varphi)$ terms are zero for all $p < t$. Hence the model $\mathcal{M}_n$ proves the existence of a sequence of knots in Theorem 5.3. In fact, $\mathcal{M}_n$ produces exact samples from the terminal predictive distribution, indicating that repeated applications of knotsets to a Feynman–Kac model can yield a perfect sampler. In an SMC algorithm, the exact samples will be independently and identically distributed only if adaptive resampling (Kong et al., 1994; Liu and Chen, 1995; Del Moral et al., 2012) is used.

Example 5.4 (A sequence of adapted knot-models)**.** *For $t \in [0:n-1]$, consider a sequence of t-knots $\mathcal{K}_t^\diamond$ and models $\mathcal{M}_{t+1} = \mathcal{K}_t^\diamond * \mathcal{M}_t$ with initial model $\mathcal{M}_0 = (M_{0:n}, G_{0:n})$. Let $\eta_{0:n}$ be the predictive probability measures for $\mathcal{M}_0$. If $\mathcal{K}_t^\diamond$ is an adapted t-knot of $\mathcal{M}_t$ for $t \in [0:n-1]$ then $\mathcal{M}_t = (M_{t,0:n}, G_{t,0:n})$ such that*

$$M_{t,0} = \delta_0, \quad M_{t,t}(0,\cdot) = \eta_t, \quad M_{t,p} = \begin{cases} \mathrm{Id} & \text{if } p \in [1:t-1] \text{ and } t \geq 2, \\ M_p & \text{if } p \in [t+1:n] \text{ and } t \leq n-1, \end{cases}$$

for $t \in [1:n]$. Whilst the potential functions are

$$G_{t,p} = \begin{cases} \eta_p(G_p) & \text{if } p \in [0:t-1], \\ G_p & \text{if } p \in [t:n]. \end{cases}$$

The sequence of models $\mathcal{M}_t$ in Example 5.4 accumulate zero asymptotic variance from times $p \in [0:t-1]$ without changing the asymptotic variance in later times $p \in [t:n]$. Applying a sequence of adapted knotsets would reduce the overall variance faster and yield the same $\mathcal{M}_n$ but is more complicated to describe and less informative as an example.

For the final model, $\mathcal{M}_n$, the remaining variance is only incurred at the terminal time. We can view the SMC algorithm on $\mathcal{M}_n$ with adaptive resampling as equivalent to an importance sampler using η_n as the importance distribution and weight function $G_n(x_n)\prod_{p=0}^{n-1}\eta_p(G_p) = G_n(x_n)\gamma_n(1)$. From the importance sampling view, we know that to reduce the remaining variance its minimal value, we will need to consider both the terminal potential and the test function of interest. Hence we note that terminal knots, introduced next, will need a treatment that reflect this, and is necessarily different from standard knots.

6 Tying terminal knots

The knots considered so far have only acted at times $p \in [0:n-1]$ and have led to a variance ordering for all terminal particle approximations of relevant test functions. When considering particle approximations of a fixed test function, a terminal knot can be used to (further) reduce the asymptotic variance. Compared to standard knots, terminal knots require special treatment to ensure that the resulting Feynman–Kac model retains the same horizon n and terminal measure. Naively adapting the knot procedure from Section 3.1 would result in a model with an $n+1$ horizon and asymptotic variance that may be difficult to compare. As such, our approach is to explicitly extend the state-space of the terminal time to prepare the Feynman–Kac model for use with terminal knots. We introduce such extended models in Section 6.1.

6.1 Extended models

Any Feynman–Kac model with terminal elements M_n and G_n can be trivially extended by replacing these terminal components with $M_n \otimes \mathrm{Id}$ and $G_n \otimes 1$ respectively. This replacement artificially extends the horizon and preserves the terminal measures, without inducing further resampling events. We generalise this notion in Definition 6.1, describing a ϕ-extension that is useful to characterise variance reduction and equivalence among models with terminal knots for specific test functions.

Definition 6.1 (ϕ-extended Feynman–Kac model)**.** *Let $\mathcal{M} = (M_{0:n}, G_{0:n})$ and $\phi \in \mathcal{L}(\hat{\gamma}_n)$ be a $\hat{\gamma}_n$-a.e. positive function. The ϕ-extended model of $\mathcal{M}$, $\mathcal{M}^\phi = (M^\phi_{0:n}, G^\phi_{0:n})$, has terminal Markov kernel $M^\phi_n = M_n \otimes \mathrm{Id}$ where the identity kernel is defined on $(\mathsf{X}_n, \mathcal{X}_n)$, and terminal potential function $G^\phi_n = (G_n \cdot \phi) \otimes \phi^{-1}$. The remaining kernels and potentials are unchanged.*

We will refer to $\mathcal{M}$ as the reference model for the ϕ-extension and to ϕ as the target function. As with the Markov kernels and potentials, marginal measures of $\mathcal{M}^\phi$ will be distinguished with a ϕ superscript. Note that the use of superscript ϕ will be reserved for extended models and should not be confused with twisted Markov kernels or measures. A ϕ-extended Feynman–Kac model can be thought of as a superficial change to the model, with several equivalences stated next. This construction ensures that the particle approximations are unchanged, and prepares the model for use with terminal knots. It is clear from Definition 6.1 that the non-terminal measures of a ϕ-extended model are equivalent to that of the reference model. We characterise the equivalences for terminal measures and the asymptotic variance in Proposition 6.2.

Proposition 6.2 (ϕ-extended model equivalences)**.** *Consider the ϕ-extended model $\mathcal{M}^\phi$ and reference model $\mathcal{M}$. Let γ_n and $\hat{\gamma}_n$ be marginal terminal measures of $\mathcal{M}$ and $\hat{\sigma}^2$ be the asymptotic variance map. If γ_n^ϕ and $\hat{\gamma}_n^\phi$ are the marginal terminal measures of $\mathcal{M}^\phi$ and $\hat{\sigma}_\phi^2$ is the asymptotic variance map then*

1. $\gamma_n^\phi(1 \otimes \varphi) = \gamma_n(\varphi)$ *for all* $\varphi \in \mathcal{L}(\gamma_n)$.
2. $\hat{\gamma}_n^\phi(1 \otimes \varphi) = \hat{\gamma}_n(\varphi)$ *for all* $\varphi \in \mathcal{L}(\hat{\gamma}_n)$.
3. $\hat{\sigma}_\phi^2(1 \otimes \varphi) = \hat{\sigma}^2(\varphi)$ *for all* $\varphi \in \hat{\mathcal{F}}(\mathcal{M})$.

The ϕ-extended model creates an additional pseudo-time step in the Feynman–Kac model which can then be manipulated by the terminal knot defined analogously to a standard knot. To work with terminal knots, we will replace reference models with their ϕ-extended counterpart.

6.2 Terminal knots

The definition of a terminal knot is essentially equivalent to that of standard knot in Definition 3.1. However, a knot (t, R, K) will only be terminal with respect to a model $\mathcal{M} \in \mathscr{M}_n$ when $t = n$. The key difference when applying a terminal knot is the additional compatibility condition on the model described in Definition 6.3. In essence, we require an additional time step at $n + 1$ for the terminal knot to operate analogously to standard knots, but we do not want an additional resampling step.

Definition 6.3 (Terminal knot compatibility)**.** *Let $\mathcal{M}^\circ = (M_{0:n}^\circ, G_{0:n}^\circ)$ be a Feynman–Kac model and $\mathcal{K} = (n, R, K)$ be a terminal knot. The model $\mathcal{M}^\circ$ and terminal knot $\mathcal{K}$ are compatible if*

(i) The model satisfies $M_n^\circ = U \otimes V^{G_n \cdot \phi}$ and $G_n^\circ = V(G_n \cdot \phi) \otimes \phi^{-1}$, for some Markov kernels U and V, reference model $\mathcal{M}$ with terminal potential G_n, and target function ϕ.

(ii) The knot satisfies $U = RK$.

Overall, Definition 6.3 extends the notion of knot-model compatibility to the special case of terminal knots, and hence expands the compatible knot and model set, $\mathscr{D}_n$, given in (5). We will refer to compatibility condition (i) as *model compatibility*. Recall that the knot operator, $* : \mathscr{D}_n \to \mathscr{M}_n$, maps compatible knot-model pairs to the space of Feynman–Kac models for horizon $n \in \mathbb{N}_0$. Definition 6.4 extends this operation to terminal knots.

Definition 6.4 (Knot operator, terminal knots)**.** *Consider a compatible terminal knot $\mathcal{K} = (n, R, K)$ and model $\mathcal{M} = (M_{0:n}, G_{0:n})$. If $M_n = P_1 \otimes P_2$ and $G_n = H \otimes \phi^{-1}$ then the knot operation yields $\mathcal{K} * \mathcal{M} = (M^*_{0:n}, G^*_{0:n})$ where*

$$M^*_n = R \otimes K^H P_2, \quad G^*_n = K(H) \otimes \phi^{-1},$$

for some Markov kernels P_1 and P_2, and functions H and ϕ. The remaining Markov kernels and potential functions are identical to the original model, that is $M^*_p = M_p$ and $G^*_p = G_p$ for $p \in [0 : n-1]$.

Note that the existence of P_1, P_2, H, and ϕ in Definition 6.4 are guaranteed by compatibility condition (i), ensuring model compatibility. However, this still leaves the question of which models have the required form. Taking $U = M_n$ and $V = \mathrm{Id}$, demonstrates that a ϕ-extended model is a compatible model. Further, we can demonstrate the set of compatible models, i.e. those satisfying compatibility condition (i), are closed under application of knots.

Proposition 6.5 (Compatible models closed under knots)**.** *Consider a model $\mathcal{M}^\circ \in \mathscr{M}_n$ satisfying compatibility condition (i) in Definition 6.3. If $\mathcal{K}$ is a (terminal) knot compatible with $\mathcal{M}^\circ$ then $\mathcal{K} * \mathcal{M}^\circ$ will also satisfy the same compatibility condition.*

Proposition 6.5 demonstrates that compatibility condition (i) of Definition 6.3 is preserved by the applications of knots. Hence, models of the form $\mathcal{K}_m * \cdots * \mathcal{K}_1 * \mathcal{M}^\circ$ will satisfy model compatibility, where $\mathcal{K}_p$ are (terminal) knots or knotsets and $\mathcal{M}^\circ$ is a ϕ-extended model. Typically though, a terminal knot will be the first to be applied to a ϕ-extended model, which we demonstrate in Example 6.6.

Example 6.6 (Terminal knot for ϕ-extended model)**.** *If $\mathcal{M}^\phi$ is the ϕ-extended model of $\mathcal{M} = (M_{0:n}, G_{0:n})$ and $\mathcal{K} = (n, R, K)$ is a terminal knot then $\mathcal{K} * \mathcal{M}^\phi = (M^*_{0:n}, G^*_{0:n})$ where $M^*_n = R \otimes K^{G_n \cdot \phi}$ and $G^*_n = K(G_n \cdot \phi) \otimes \phi^{-1}$.*

Example 6.6 shows that applying a terminal knot to a ϕ-extended model incorporates information from the terminal potential function and the target function ϕ into the new model. The use of ϕ-extensions and specific target function can be motivated by drawing a comparison to optimal importance distributions (see discussion in Section 5). These

components are carefully constructed to preserve the marginal distributions, at least in some form, and facilitate our variance reduction results in Section 6.3.

Proposition 6.7 states the invariance of terminal measures when a terminal knot is applied.

Proposition 6.7 (Terminal knot-model invariants)**.** *Let $\mathcal{K} = (n, R, K)$ be a terminal knot and consider knot-model $\mathcal{M}^* = \mathcal{K} * \mathcal{M}^\circ$. For measurable function φ, the knot-model $\mathcal{M}^*$ will have the following invariants and equivalences:*

1. *$\gamma_p^*(\varphi) = \gamma_p(\varphi)$, $\eta_p^*(\varphi) = \eta_p(\varphi)$, $\hat{\gamma}_p^*(\varphi) = \hat{\gamma}_p(\varphi)$, and $\hat{\eta}_p^*(\varphi) = \hat{\eta}_p(\varphi)$ for all $p \in [0:n-1]$.*
2. *$\hat{\gamma}_n^*(1 \otimes \varphi) = \hat{\gamma}_n(1 \otimes \varphi)$ and $\hat{\eta}_n^*(1 \otimes \varphi) = \hat{\eta}_n(1 \otimes \varphi)$.*

Two types of special terminal knots are stated in state in Example 6.8 and 6.9 which are the terminal counterparts to trivial and adapted standard knots. Note that for compatibility the reference model will need to be ϕ-extended before these knots can be applied.

Example 6.8 (Trivial terminal knot)**.** *Consider a terminal knot $\mathcal{K} = (n, P_1, \mathrm{Id})$ and model $\mathcal{M} \in \mathscr{M}_n$ where the terminal kernel is $M_n = P_1 \otimes P_2$. The model resulting from $\mathcal{K} * \mathcal{M} = \mathcal{M}$.*

As in the standard case, the trivial terminal knot does not change the Feynman–Kac model. At the other extreme is the adapted terminal knot, for which we discuss optimality in Section 6.4.

Example 6.9 (Adapted terminal knot)**.** *Consider a terminal knot $\mathcal{K} = (n, \mathrm{Id}, P_1)$ and model $\mathcal{M} \in \mathscr{M}_n$ where the terminal Markov kernel and potential are $M_n = P_1 \otimes P_2$ and $G_n = H \otimes \phi^{-1}$ respectively. The model $\mathcal{M}^* = \mathcal{K} * \mathcal{M}$ has new terminal kernel and potential function,*

$$M_n^* = \mathrm{Id} \otimes P_1^H P_2, \quad G_n^* = P_1(H) \otimes \phi^{-1}.$$

Whilst the remaining kernels and potentials are unchanged. Further, if the initial model is a ϕ-extension, say $\mathcal{M}^\phi$ where $\mathcal{M} = (M_{0:n}, G_{0:n})$, then

$$M_n^* = \mathrm{Id} \otimes M_n^{G_n \cdot \phi}, \quad G_n^* = M_n(G_n \cdot \phi) \otimes \phi^{-1}.$$

To conclude our treatment of terminal knots, we now define *terminal knotsets*. Unlike standard knotsets, a terminal knotset includes a knot at the terminal time.

Definition 6.10 (Terminal knotsets and compatibility)**.** *A terminal knotset $\mathcal{K} = (R_{0:n}, K_{0:n})$ is specified by $n+1$ knots of the form $\mathcal{K}_p = (p, R_p, K_p)$ for $p \in [0:n]$. Such a knotset is compatible with $\mathcal{M} \in \mathscr{M}_n$ if each (p, R_p, K_p)-knot is compatible with $\mathcal{M}$ for $p \in [0:n]$.*

Note that the compatibility condition for the terminal knot in the set ensures that a ϕ-extension has occurred on a reference model as part of the construction of $\mathcal{M}$.

Definition 6.11 (Terminal knotset operator)**.** *Let $\mathcal{K} = (R_{0:n}, K_{0:n})$ be a terminal knotset compatible with $\mathcal{M} \in \mathscr{M}_n$. The terminal knotset operation is defined as $\mathcal{K} * \mathcal{M} = \mathcal{K}_0 * \mathcal{K}_1 * \cdots * \mathcal{K}_n * \mathcal{M}$ where $\mathcal{K}_p = (p, R_p, K_p)$ for $p \in [0:n]$.*

When considering the application of a series of knots to a model, it is not necessary to apply terminal knots first. However, as with standard knotsets, such an order ensures that the individual compatibility conditions correspond to the overall compatibility condition.

A special case of Definition 6.10 is an adapted terminal knotset, which consists of an adapted knotset extended to include an adapted terminal knot. We explore terminal knotsets further in Section 6.5 to reduce the variance of normalising constant estimates.

6.3 Variance reduction and ordering

With terminal measure equivalence established in Proposition 6.7, we can describe the variance reduction from terminal knots for functions of the form $1 \otimes \varphi$. Recall that the model must be ϕ-extended before we can apply terminal knots. We start with Theorem 6.12, stating a general result for the difference in asymptotic variance after applying a terminal knot, before carefully specifying the models and test functions for which we can state an asymptotic variance ordering for.

Theorem 6.12 (Variance difference with a terminal knot)**.** *Consider models $\mathcal{M} = (M_{0:n}, G_{0:n})$ and $\mathcal{M}^* = \mathcal{K} * \mathcal{M}$ for terminal knot $\mathcal{K} = (n, R, K)$. If $\bar{\varphi} = 1 \otimes \varphi \in \hat{\mathcal{F}}(\mathcal{M})$ then*

$$\hat{\sigma}^2(\bar{\varphi}) - \hat{\sigma}^2_*(\bar{\varphi}) = \frac{\hat{\eta}_{n-1} R\{\mathrm{Cov}_K(H, H \cdot P_2[\{\phi^{-1} \cdot \varphi\}^2])\}}{\eta_n(G_n)^2},$$

where $M_n = P_1 \otimes P_2$ and $G_n = H \otimes \phi^{-1}$.

The equivalence of updated terminal measures under $\mathcal{M}$ and $\mathcal{M}^*$ in Theorem 6.12 for a test function $\bar{\varphi}$ is stated in Proposition 6.7. Clearly, models satisfying $P_2[\{\phi^{-1} \cdot \varphi\}^2] = 1$ almost surely will have a guaranteed variance reduction and there may be certain model classes where more general conclusions can be made. We state some sufficient conditions in Corollary 6.13 to ensure a variance reduction.

Corollary 6.13 (Variance reduction with a terminal knot)**.** *Consider model $\mathcal{M}^\circ \in \mathscr{M}_n$ and terminal knot $\mathcal{K} = (n, R, K)$ and let $\mathcal{M}^* = \mathcal{K} * \mathcal{M}^\circ$. For a reference model $\mathcal{M} = (M_{0:n}, G_{0:n})$ and target function ϕ, if*

1. *for some $m \in \mathbb{N}_1$ there exists a sequence of knots $\mathcal{K}_1, \ldots, \mathcal{K}_m$ such that $\mathcal{M}^\circ = \mathcal{K}_m * \cdots * \mathcal{K}_1 * \mathcal{M}^\phi$, and*
2. *$\varphi \in \hat{\mathcal{F}}(\mathcal{M})$ and $\phi = |\varphi|$, then*

$$\hat{\sigma}_*^2(1 \otimes \varphi) \leq \hat{\sigma}_\circ^2(1 \otimes \varphi) \leq \hat{\sigma}^2(\varphi).$$

Further, if $\mathcal{M}^\circ = \mathcal{M}^\phi$ *then* $\hat{\sigma}_\circ^2(1 \otimes \varphi) = \hat{\sigma}^2(\varphi)$,

$$\hat{\sigma}^2(\varphi) - \hat{\sigma}_*^2(1 \otimes \varphi) = \frac{\hat{\eta}_{n-1} R\{\mathrm{Var}_K(G_n \cdot |\varphi|)\}}{\eta_n(G_n)^2},$$

and the inequality $\hat{\sigma}_*^2(1 \otimes \varphi) \leq \hat{\sigma}^2(\varphi)$ *is strict if* $\hat{\eta}_{n-1} R\{\mathrm{Var}_K(G_n \cdot |\varphi|)\} > 0$.

Corollary 6.13 presents the incremental variance reduction from a terminal knot applied to a model that can be expressed as a ϕ-extended model with or without knots. It is written to emphasise the case of updated measures, since terminal knots are defined for an updated measure by convention, but includes predictive measures as a special case when $G_n = 1$. Multiple applications of terminal and standard knots are treated by the partial ordering described in Theorem 6.15. Importantly, we can only consider φ that are almost everywhere non-zero due the the conditions imposed by the ϕ-extension in Definition 6.1.

To reduce the variance for a particle approximation to a probability measure, i.e. $\hat{\eta}_n(\varphi)$, Corollary 6.13 implies that one should set $\phi = |\varphi - \hat{\eta}_n(\varphi)|$. However, this would require knowledge of $\hat{\eta}_n(\varphi)$ in advance. Iterative schemes could be used to approximate such a terminal knot, but the result would be approximate. We leave investigation of such iterative schemes for future work. As present, terminal knots are most amenable to normalising constant estimation which we consider specifically in Section 6.5.

Terminal knots can be used in conjunction with standard knots, but such a combination will only guarantee a reduction in the asymptotic variance for the target function $\phi = |\varphi|$. Equipped with terminal knots, we can define a partial ordering on Feynman–Kac models specifically for a test function φ.

Definition 6.14 (A partial ordering of Feynman–Kac with terminal knots)**.** *Consider two Feynman–Kac models,* $\mathcal{M}^\circ, \mathcal{M}^* \in \mathscr{M}_n$ *and target function* ϕ*. We say that* $\mathcal{M}^* \preccurlyeq_\phi \mathcal{M}^\circ$ *with respect to a reference model* $\mathcal{M} \in \mathscr{M}_n$ *if for some* $m, m' \in \mathbb{N}_1$

1. *there exists a sequences of knots* $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_m$ *such that* $\mathcal{M}^* = \mathcal{K}_m * \cdots * \mathcal{K}_1 * \mathcal{M}^\circ$,
2. *there exists a sequences of knots* $\mathcal{K}'_1, \mathcal{K}'_2, \ldots, \mathcal{K}'_{m'}$ *such that* $\mathcal{M}^\circ = \mathcal{K}'_m * \cdots * \mathcal{K}'_1 * \mathcal{M}^\phi$.

Each knot in the sequences can be a terminal knots or a standard knot.

Compared to Definition 4.3, this partial ordering now includes terminal knots but at the expense of generality: We are now tied to a single test function φ that satisfies $\phi = |\varphi|$ as the variance ordering states in Theorem 6.15.

Theorem 6.15 (Variance ordering with terminal knots)**.** *Consider a reference model* $\mathcal{M} \in \mathscr{M}_n$ *and let* $\varphi \in \hat{\mathcal{F}}(\mathcal{M})$*. If* $\mathcal{M}^* \preccurlyeq_\phi \mathcal{M}^\circ$ *with respect to* $\mathcal{M}$ *and* $\phi = |\varphi|$ *then*

$$\hat{\gamma}_n^*(1 \otimes \varphi) = \hat{\gamma}_n^\circ(1 \otimes \varphi) = \hat{\gamma}_n(\varphi) \quad \textit{and} \quad \hat{\sigma}_*^2(1 \otimes \varphi) \leq \hat{\sigma}_\circ^2(1 \otimes \varphi) \leq \hat{\sigma}^2(\varphi).$$

6.4 Optimality of adapted terminal knots

Analogously to their standard counterparts, applying an adapted terminal knot results in the largest variance reduction of any single terminal knot for the test function φ. We state this result in Theorem 6.16.

Theorem 6.16 (Adapted terminal knot optimality)**.** *Consider a model $\mathcal{M}^\circ$ satisfying Part 2 of Definition 6.14 with reference model $\mathcal{M}$. Let $\mathcal{K}$ and $\mathcal{K}^\diamond$ be terminal knots compatible with $\mathcal{M}^\circ$. If $\mathcal{K}^\diamond$ is the adapted terminal knot for $\mathcal{M}^\circ$ then $\mathcal{K}^\diamond * \mathcal{M}^\circ \preccurlyeq_\phi \mathcal{K} * \mathcal{M}^\circ$ with respect to $\mathcal{M}$.*

Beyond this optimality for a single terminal knot, we can also prove that adapted terminal knots allow for the asymptotic variance to be reduced to zero in some cases, in conjunction with standard knots.

Corollary 6.17 (Minimal variance from knotsets with terminal knot)**.** *For every model $\mathcal{M} \in \mathscr{M}_n$ and a.s. non-zero $\varphi \in \hat{\mathcal{F}}(\mathcal{M})$ there exists a sequence of knots $\mathcal{K}_{n+1}, \mathcal{K}_n, \ldots, \mathcal{K}_1$ such that*

$$\mathcal{M}^\star = \mathcal{K}_{n+1} * \mathcal{K}_n * \cdots * \mathcal{K}_1 * \mathcal{M}^{|\varphi|}$$

has asymptotic variance terms satisfying

$$\hat{v}^\star_{n,n}(\bar{\varphi}) = \hat{\eta}_n(|\varphi|)^2 - \hat{\eta}_n(\varphi)^2, \quad \hat{v}^\star_{p,n}(\varphi) = 0, \text{ for } p \in [0 : n-1]$$

where $v^\star_{p,n}(\varphi)$ are the asymptotic variance terms for $\mathcal{M}^\star$ and $\hat{\eta}_n$ is the terminal updated probability measure for $\mathcal{M}$.

With Corollary 6.17, we can state that when the target function φ is almost surely non-negative or non-positive, then the asymptotic variance of the particle approximation for φ is zero. This property is analogous to that of optimal importance functions in importance sampling. We can extended the result to non-negative or non-positive φ by replacing the terminal potential by $G^0_n = G_n \cdot 1_{\overline{S_0(\varphi)}}$. Noting that $\gamma_n(G_n \cdot \varphi) = \gamma_n(G^0_n \cdot \varphi)$ shows the equivalence, though the terminal probability measures will differ.

To construct particle estimates with zero asymptotic variance under $\hat{\gamma}_n$ for more general functions, one could adapt the strategy of "positivisation" from importance sampling (see for example, Owen and Zhou, 2000). We can write the terminal predictive measure of a fixed function $\varphi \in \mathcal{L}(\hat{\gamma}_n)$ as $\gamma_n(G_n \cdot \varphi) = \gamma_n(G^+_n \cdot |\varphi|) - \gamma_n(G^-_n \cdot |\varphi|)$, where $G^+_n = G_n \cdot 1_{\varphi > 0}$ and $G^-_n = G_n \cdot 1_{\varphi < 0}$. From this expression it is natural to consider two SMC algorithms; one with terminal potential G^+_n and the other with G^-_n. The underlying Feynman–Kac models and test functions of both algorithms now satisfy the conditions to achieve zero variance.

We now state an example model that achieves the variance in Corollary 6.17 which can be constructed by extending the sequence of models given in Example 5.4.

Example 6.18 (A sequence of adapted knot-models, continued)**.** *Consider the sequence of models in Example 5.4 with additional requirement that the initial model $\mathcal{M}_0 = \mathcal{M}^\phi$ is a ϕ-extension such that $\phi = |\varphi|$. Denote the predictive probability measures of $\mathcal{M} = (M_{0:n}, G_{0:n})$ as $\eta_{0:n}$. Let the next model in the sequence be $\mathcal{M}_{n+1} = \mathcal{K}_n^\diamond * \mathcal{M}_n$. If $\mathcal{K}_n^\diamond$ is the adapted terminal knot for $\mathcal{M}_n$ then the model $\mathcal{M}_{n+1} = (M_{0:n}^\star, G_{0:n}^\star)$ satisfies*

$$M_0^\star = \delta_0, \quad M_n^\star(0, \cdot) = \delta_0 \otimes \eta_n^{G_n \cdot \phi}, \quad M_p^\star = \mathrm{Id} \text{ for } p \in [1 : n-1],$$

with potential functions $G_p^\star = \eta_p(G_p)$ for $p \in [0 : n-1]$ and $G_n^\star = \eta_n(G_n \cdot \phi) \otimes \phi^{-1}$.

Having proven and demonstrated that a target model can be transformed until it has minimal asymptotic variance using knots, we can conclude that the partial ordering induced by knots includes the optimal model.

6.5 Estimating normalising constants

One of the most useful cases for terminal knots is when the normalising constant of the Feynman–Kac model, $\hat{\gamma}_n(1)$, of the model is of primary interest. If $\varphi = 1$ is the only test function of interest then it is possible to specify a simplified Feynman–Kac model, which we detail in Example 6.19, which results from applying a terminal knotset (see Definition 6.11).

Example 6.19 (Terminal knotset for normalising constant estimation)**.** *Consider a ϕ-extended model $\mathcal{M}^\phi \in \mathscr{M}_n$ and terminal knotset $\mathcal{K} = (R_{0:n}, K_{0:n})$ where $\phi = 1$. The knot-model $\mathcal{M}^* = \mathcal{K} * \mathcal{M}^\phi = (M_{0:n}^*, G_{0:n}^*)$ satisfies $M_0^* = R_0$, $M_p^* = K_{p-1}^{G_{p-1}} R_p$ for $p \in [1 : n-1]$, $M_n^* = K_{n-1}^{G_{n-1}} R_n \otimes K_n^{G_n}$, $G_p^* = K_p(G_p)$ for $p \in [0 : n-1]$, and $G_n^* = K_n(G_n) \otimes 1$.*

The model $\mathcal{M}^\dagger = (M_{0:n}^\dagger, G_{0:n}^\dagger)$ with

$$\begin{aligned} M_0^\dagger &= R_0, & G_0^\dagger &= K_0(G_0), \\ M_p^\dagger &= K_{p-1}^{G_{p-1}} R_p, & G_p^\dagger &= K_p(G_p), \quad p \in [1 : n], \end{aligned}$$

will satisfy $\hat{\gamma}_n^\dagger(1) = \hat{\gamma}_n^*(1) = \hat{\gamma}_n(1)$ and $\hat{\sigma}_\dagger^2(1) = \hat{\sigma}_*^2(1) \leq \hat{\sigma}^2(1)$.

Note that $K_n^{G_n}$ does not need to be simulated in the model $\mathcal{M}^\dagger$. The asymptotic variance for model $\mathcal{M}^\dagger$ can also be further reduced by applying more knots. A special case of Example 6.19 is using the adapted terminal knotset. In this case, $M_0^\dagger = \delta_0$ and so long as $G_0^\dagger = M_0(G_0)$ is accounted for elsewhere in the algorithm, the first iteration of the SMC algorithm does not need to be run.

7 Applications and examples

7.1 Particle filters with 'full' adaptation

A particle filter with 'full' adaptation adapts each Markov kernel in the Feynman–Kac model to the current potential information through twisting. Originally proposed as a

type of auxiliary particle filter by Pitt and Shephard (1999), its modern interpretation does away with auxiliary variables, though it is still often referred to as a fully-adapted auxiliary particle filter. It is popular due to its empirical performance and its derivation which is motivated by identifying locally (i.e. conditional) optimal proposal distributions for the particle weights at each time step. We refer to this algorithm as a particle filter with 'full' adaptation. The Feynman–Kac model for such an algorithm is described in Example 7.1.

Example 7.1 (Particle filter with 'full' adaptation)**.** *Let $\mathcal{M} = (M_{0:n}, G_{0:n})$ be a model for a particle filter. The particle filter with 'full' adaptation with respect to $\mathcal{M}$ has model $\mathcal{M}^{\mathrm{F}} = (M^{\mathrm{F}}_{0:n}, G^{\mathrm{F}}_{0:n})$ satisfying*

$$\begin{aligned} M_0^{\mathrm{F}} &= M_0^{G_0}, \quad G_0^{\mathrm{F}} = M_0(G_0) \cdot M_1(G_1), \\ M_p^{\mathrm{F}} &= M_p^{G_p}, \quad G_p^{\mathrm{F}} = M_{p+1}(G_{p+1}), \qquad p \in [1:n-1] \\ M_n^{\mathrm{F}} &= M_n^{G_n}, \quad G_n^{\mathrm{F}} = 1. \end{aligned}$$

The adapted knot-model in Example 3.11 and the particle filter with 'full' adaptation in Example 7.1 share the same constituent twisted Markov kernels $M_p^{G_p}$ and expected potential functions $M_p(G_p)$, but differ in where these elements are located in time. In particular, $M_p^* = M_{p-1}^{G_{p-1}} = M_{p-1}^{\mathrm{F}}$ and $G_p^* = M_p(G_p) = G_{p-1}^{\mathrm{F}}$ for $p \in [2:n-1]$. A further crucial difference is that the adapted knot-model does not use the terminal twisted kernel $M_n^{G_n}$. Our theory on knots has shown that adapted knot-models order the asymptotic variance for all relevant test functions, whilst Johansen and Doucet (2008) contained a counterexample for such a result with the particle filter with 'full' adaptation. Note that they referred to this model as having 'perfect' adaptation.

We now restate the counterexample in Example 7.2, and demonstrate how adapted knot-models guarantee an asymptotic variance reduction whereas the fully-adapted particle filter does not.

Example 7.2 (Binary model of Johansen and Doucet, 2008)**.** *Let $\mathcal{B} = (M_{0:1}, G_{0:1})$ be a Feynman–Kac model with*

$$M_0(x_0) = \begin{cases} \frac{1}{2} & \text{if } x_0 = 0, \\ \frac{1}{2} & \text{if } x_0 = 1, \end{cases} \qquad M_1(x_0, x_1) = \begin{cases} 1-\delta & \text{if } x_1 = x_0, \\ \delta & \text{if } x_1 = 1 - x_0, \end{cases}$$

respectively, and potential functions

$$G_t(x_t) = \begin{cases} 1-\varepsilon & \text{if } x_t = y_t, \\ \varepsilon & \text{if } x_t = 1 - y_t, \end{cases} \qquad t \in \{0, 1\}$$

with fixed observations y_t such that $y_0 = 0$ and $y_1 = 1$.

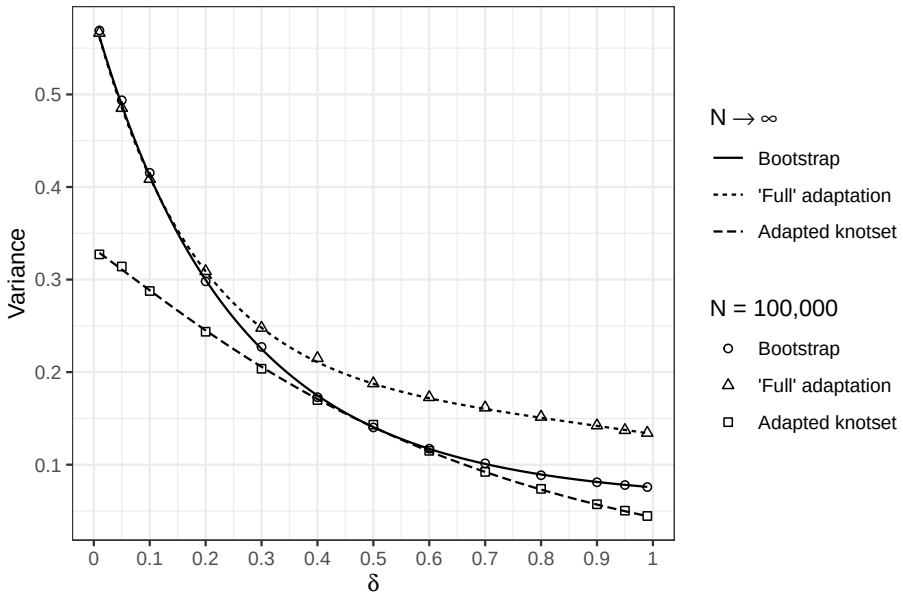

Figure 2: Analytical ($N \to \infty$) and empirical ($N = 100{,}000$) asymptotic variance of bootstrap particle filter, filter with 'full' adaptation, and adapted knotset particle filter for $\hat{\eta}_1^N(\varphi)$ in Example 7.2 with $\varepsilon = 0.25$ and $\varphi(x) = x$. The empirical variance is estimated from 50,000 independent replications.

Figure 2 compares the asymptotic variances of various particle filters (PF) approximating $\hat{\eta}_1^N(\varphi)$ with $\varphi(x) = x$. We consider the bootstrap PF, PF with 'full' adaptation, and adapted knotset PF in Example 7.2 with $\varepsilon = 0.25$ and $\delta \in (0, 1)$, both analytically and empirically. Figure 2 replicates the second figure in Johansen and Doucet (2008) with the addition of the adapted knotset PF. The PF with 'full' adaptation is specified by Example 7.1 with $n = 1$, whilst the adapted knotset PF has $M_0^* = \delta_0$, $M_1^*(0, \cdot) = M_0^{G_0} M_1$, $G_0^* = M_0(G_0)$, and $G_1^* = G_1$. When $\varepsilon = 0.25$, the adapted knotset PF outperforms the other PFs in this regime, whilst the bootstrap PF mostly outperforms the PF with 'full' adaptation. The existence of regimes where the bootstrap PF outperforms the PF with 'full' adaptation constitutes the counterexample of Johansen and Doucet (2008).

Figure 3 displays the excess analytical asymptotic variance of the PF with 'full' adaptation and adapted knotset PF, relative to the bootstrap PF, for $\varepsilon \in \{0.05, 0.1, 0.2, 0.4, 0.5\}$ and $\delta \in (0, 1)$. The excess variance of the adapted knotset PF is always less than or equal to zero, demonstrating the dominance over the bootstrap PF, whilst the PF with 'full' adaptation can be better or worse than the bootstrap depending on the regime. We also note that the PF with 'full' adaptation can outperform the adapted knotset PF for some parameter values and test functions. From our theory, knot-models or equivalent specifications guarantee variance ordering for all relevant test functions when we do not include a knot at the terminal time. Adapting the terminal time, as the PF with 'full' adaption does,

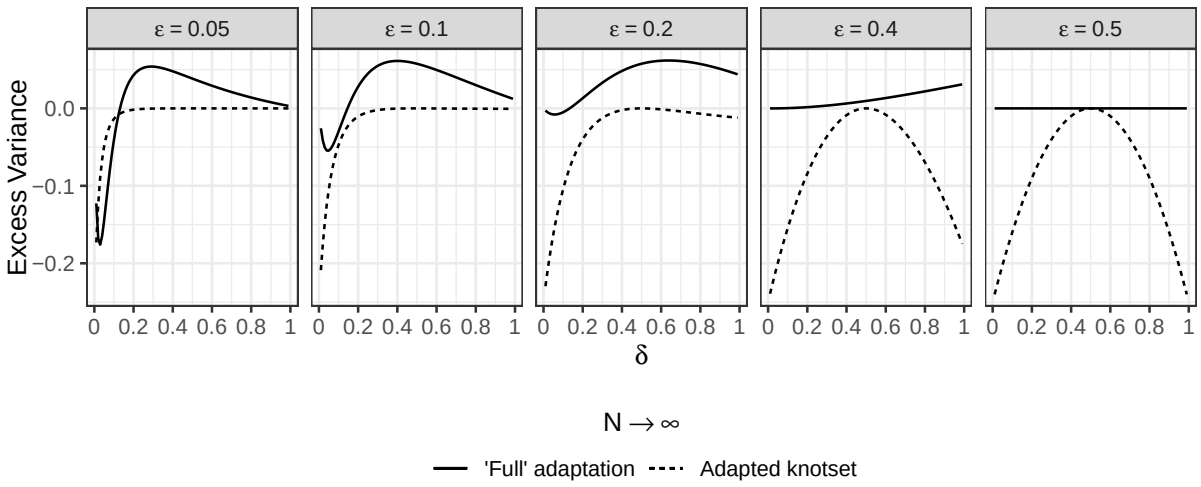

Figure 3: Analytical ($N \to \infty$) asymptotic excess variance of filter with 'full' adaptation and adapted knotset particle filter relative to the bootstrap particle filter for $\hat{\eta}_1^N(\varphi)$ in Example 7.2 with $\varepsilon = 0.25$ and $\varphi(x) = x$. The excess variance is calculated by subtracting the asymptotic variance of bootstrap particle filter. Negative values indicate an improvement over the bootstrap particle filter.

requires special consideration as discussed in Section 6. Using Example 7.2, we illustrate terminal knots and their relation to the PF with 'full' adaptation next.

We now consider Example 7.2 for normalising constant estimation, comparing the bootstrap PF, PF with 'full' adaption, and PF with an adapted terminal knotset. For the latter particle filter we use the simplified version stated in Example 6.19 with adapted knots, yielding components $M_0^* = \delta_0$, $M_1^*(0, \cdot) = M_0^{G_0}$, $G_0^* = M_0(G_0)$, and $G_1^* = M_1(G_1)$. Figure 4 compares the asymptotic variance of $\hat{\gamma}_1^N(1)/\hat{\gamma}_1(1)$ for each particle filter with $\varepsilon = 0.25$ and $\delta \in (0, 1)$ both analytically and empirically. The figure shows that the variance of the PF with 'full' adaptation and adapted terminal knotset PF coincide, as expected as the non-asymptotic estimates are equivalent. As such, the PF with 'full' adaptation guarantees variance reduction for normalising constant estimation.

Overall, our approach to variance reduction with knots explains why the particle filter with 'full' adaptation has good empirical performance in many different contexts. Firstly, it is only slightly different to a model (Example 3.11) that we can guarantee an asymptotic variance ordering for all relevant functions. Secondly, it does guarantee a variance reduction for normalising constant estimation because of its equivalence to an adapted terminal knot-model. Thus, we provide a cohesive explanation for the counterexample in Johansen and Doucet (2008) by clarifying that it is the adaptation to the terminal time that restricts the variance reduction guarantee, not the use of adaptation that is only locally (i.e. conditionally) optimal. Further, we have demonstrated how a minor modification to the particle

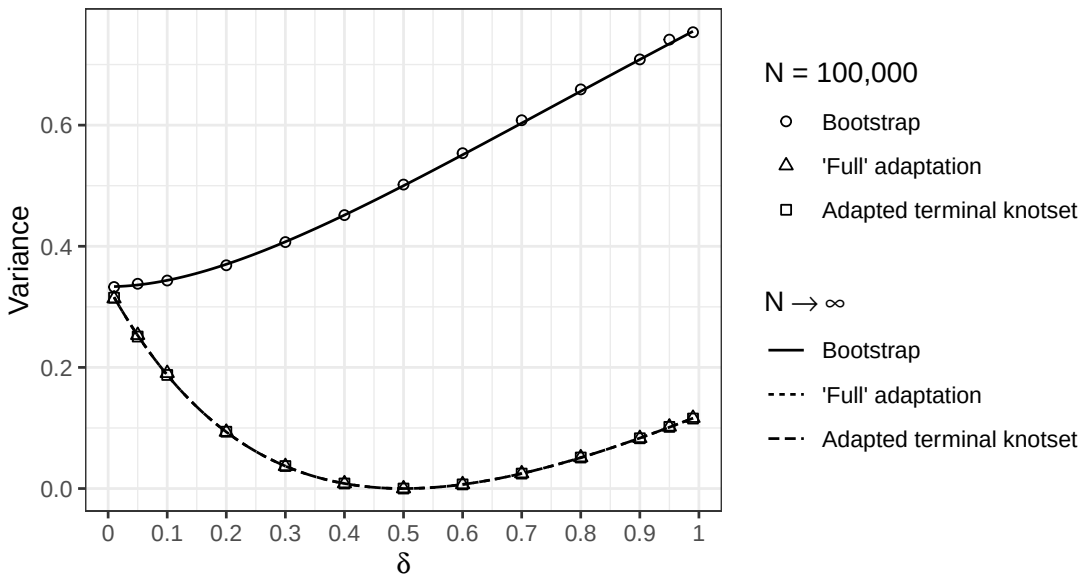

Figure 4: Analytical ($N \to \infty$) and empirical ($N = 100{,}000$) asymptotic variance of bootstrap PF, filter with 'full' adaptation, and adapted knotset PF for $\hat{\gamma}_1^N(1)/\hat{\gamma}_1(1)$ in Example 7.2 with $\varepsilon = 0.25$. The empirical variance is estimated from 50,000 independent replications. Note that the 'full' adaptation and adapted terminal knotset values are theoretically equal.

filter with 'full' adaptation, Example 3.11, guarantees variance ordering for all relevant test functions which has practical significance.

7.2 Marginalisation as a knot

Model marginalisation in SMC is a well-known variance reduction technique that can be viewed as a special case of knots. Often referred to as Rao–Blackwellisation, the procedure involves analytically marginalising part of the state-space of the Feynman–Kac model, thus reducing the dimensionality of the estimation problem. Historically, Rao–Blackwellisation has been applied to models where it is applicable at all time points, and justified in the case of sequential importance samplers by appealing to a reduction in the variance of the weights (Doucet, 1998; Doucet et al., 2000). However, this justification does not relate Rao–Blackwellisation to the variance of particle approximations from a modern SMC algorithm which, in comparison, uses resampling. Framing model marginalisation as a knot operation, we prove this procedure will order the asymptotic variance of SMC algorithms for all relevant test functions when a knot is applied to a non-terminal time point.

We describe the marginalisation knot in Example 7.3 and the application of such a knot in Example 7.4 to demonstrate how certain model assumptions recover existing Rao–Blackwellisation in the literature.

Example 7.3 (Marginalisation knot)**.** *Consider a Markov kernel* $M_t = U \otimes V$ *such that* $U : (\mathsf{X}_{t-1}, \mathcal{Z}_1) \to [0,1]$ *and* $V : (\mathsf{Z}_1 \times \mathsf{X}_{t-1}, \mathcal{Z}_2) \to [0,1]$ *for measurable spaces* $(\mathsf{Z}_1, \mathcal{Z}_1)$ *and* $(\mathsf{Z}_2, \mathcal{Z}_2)$*. The knot* $\mathcal{K} = (t, R, K)$ *where*

$$R(x, \mathrm{d}y) = U(x, \mathrm{d}y_1)\delta_x(\mathrm{d}y_2), \quad K(y, \mathrm{d}z) = \delta_{y_1}(\mathrm{d}z_1)V(y, \mathrm{d}z_2)$$

is a marginalisation knot and $M_t = RK$*.*

The kernels U and V partition the state-space M_t is defined on. In particular, U is the marginal distribution of the first component of the partition and V is the conditional distribution of the second component of the partition. The result of applying a marginalisation knot to a model is considered next.

Example 7.4 (Model with marginalistion knot)**.** *Consider* $\mathcal{M} = (M_{0:n}, G_{0:n})$ *where* $M_t = U \otimes V$ *with* U, V*, and* $\mathcal{K} = (t, R, K)$ *defined as in Example 7.3. If* $\mathcal{K} * \mathcal{M} = (M^*_{0:n}, G^*_{0:n})$ *then*

$$M^*_t = R, \quad G^*_t(y) = V[F(y_1)](y), \quad M^*_{t+1} = K^{G_t} M_{t+1},$$

where F *is a functional such that* $F(z_1) = G_t([z_1, \cdot])$ *and* $K^{G_t}(y, \mathrm{d}z) = \delta_{y_1}(\mathrm{d}z_1) \otimes V^{F(y_1)}(y, \mathrm{d}z_2)$*.*

Example 7.4 generalises several existing particle filters using marginalised models. Doucet et al. (2000) consider the case where $M_{t+1}(z, \cdot) = P(z_1, \cdot)$, for some kernel P, and hence M_{t+1} does not depend on z_2. As such, the twisted kernel $V^{F(y_1)}$ is not necessary for particle filter implementation in practice. Andrieu and Doucet (2002) present the case where $G_t(z) = H(z_1)$, for some potential function H not depending on z_2 so that $G^*_t = G_t$ and $M^*_{t+1} = KM_{t+1}$. The authors assume M_t is Gaussian and apply the Kalman filter to calculate the form of the appropriate marginal and conditional distributions, U and V respectively. Extending this further, Schön et al. (2005) use a Kalman filter to marginalise the linear-Gaussian component of more general state-space models. Special cases include mixture Kalman filters (Chen and Liu, 2000) and model-marginalised particle filters for jump Markov linear systems (Doucet et al., 2002). For these examples, and any general (e.g. non-Gaussian) state-space model, this paper contributes a complete analysis of asymptotic variance reduction for terminal particle approximations arising from SMC when analytical marginalisation can be performed.

It is also instructive to note that a knot itself can be seen as model marginalisation. If $M_t = RK$ we could extend the Feynman–Kac model from $\mathcal{M} = (M_{0:n}, G_{0:n})$ to $\mathcal{M}' = (M'_{0:n+1}, G'_{0:n+1})$ where $M'_t = R$, $G'_t = 1$, $M'_{t+1} = K$, $G'_{t+1} = G_t$, and $M'_{p+1} = M_p$ with $G'_{p+1} = G_p$ for $p \in [t+1 : n]$. Marginalising the state $X'_{t+1} \sim K(x'_t, \cdot)$ in $\mathcal{M}'$, and collecting the remaining terms, results in the model $\mathcal{K} * \mathcal{M}$ where $\mathcal{K} = (t, R, K)$. As such, a knot can also be viewed as marginalisation, in particular model extension followed by marginalisation. Our procedures and theory present the most general case of this, as well as nuance around the use of knots at the terminal time.

7.3 Non-linear Student distribution state-space models

In this section we provide a numerical example to illustrate the use of knots in practice, and elucidate the connection between adapted knots and marginalisation knots. We consider a non-linear state-space model with latent variable driven by additive Student noise. The model uses non-linear functions $f_p : \mathbb{R}^d \to \mathbb{R}^d$ such that the latent space evolution can be described as

$$X_0 \sim \mathcal{T}(\mu, \Sigma, \nu), \quad (X_p \mid X_{p-1}) \sim \mathcal{T}(f_p(X_{p-1}), \Sigma, \nu) \text{ for } p \in [1:n]$$

where $\mathcal{T}(\mu, \Sigma, \nu)$ denotes the multivariate Student's t-distribution with mean $\mu \in \mathbb{R}^d$, positive definite scale matrix $\Sigma \in \mathrm{PD}(\mathbb{R}^{d\times d})$, and degrees of freedom ν. We assume the data are observed with Gaussian noise, that is $(Y_p \mid X_p) \sim \mathcal{N}(X_p, \Sigma')$ with $\Sigma' \in \mathrm{PD}(\mathbb{R}^{d\times d})$ for $p \in [0:n]$.

We will use the fact that a multivariate Student distribution can be represented as scale mixture of multivariate Gaussian distributions with transformed χ^2_ν distribution, that is a Chi-squared distribution with ν degrees of freedom. Noting this construction, the Feynman–Kac form of this state-space model has Markov kernels satisfying

$$\begin{aligned} M_0 &= (\delta_\mu \otimes S)K \\ M_p(x_{p-1}, \cdot) &= (\delta_{f_p(x_{p-1})} \otimes S)K \text{ for } p \in [1:n] \end{aligned} \tag{6}$$

where S is a χ^2_ν distribution and K is conditionally multivariate normal with $K([z,s],\cdot) = \mathcal{N}(z, \frac{\nu}{s}\Sigma)$. Hence, the knot $\mathcal{K}_p = (p, R_p, K)$ can be applied to the model where $R_0 = \delta_\mu \otimes S$ and $R_p(x_{p-1}, \cdot) = \delta_{f(x_{p-1})} \otimes S$ for $p \in [0:n-1]$. If we ϕ-extend the model, we can also apply the analogous terminal knot $\mathcal{K}_n$.

To test the variance reductions possible for this model, we compare the bootstrap particle filter and the terminal knotset particle filter in Example 6.19 with knots $\mathcal{K}_0, \ldots, \mathcal{K}_n$. We consider the problem of normalising constant estimation and assess performance with $R = 200$ repetitions of each particle filter. The particle filter implementations used $N = 2^{10}$ particles and adaptive resampling with threshold $\kappa = 0.5$. We test the particle filters for five independent datasets simulated by the data generating process that vary by dimension $d \in [1:5]$. The time horizon is fixed at $n = 10$, with initial mean $\mu = 0_d$, degrees of freedom $\nu = 4$, and identity variance matrices $\Sigma = \Sigma' = I_d$. We use a univariate non-linear function $g_p(x) = \frac{x}{2} + \frac{25x}{1+x^2} + 8\cos(1.2p)$ (see Kitagawa, 1996, and references therein) to construct a multivariate non-linear function $f_p(x) = A[g_p(x_1) \cdots g_p(x_d)]^\top$ with matrix $A \in \mathbb{R}^{d\times d}$ having unit diagonal, off-diagonals elements set to half (when $d \geq 2$), and all other elements set to zero (when $d \geq 3$).

Figure 5 displays the estimated normalising constants from each particle filter on the log-scale. For ease of comparison, the log-estimates were shifted by a constant (for each d) so that the terminal knotset particle filter estimates have a unit mean (zero on log-scale).

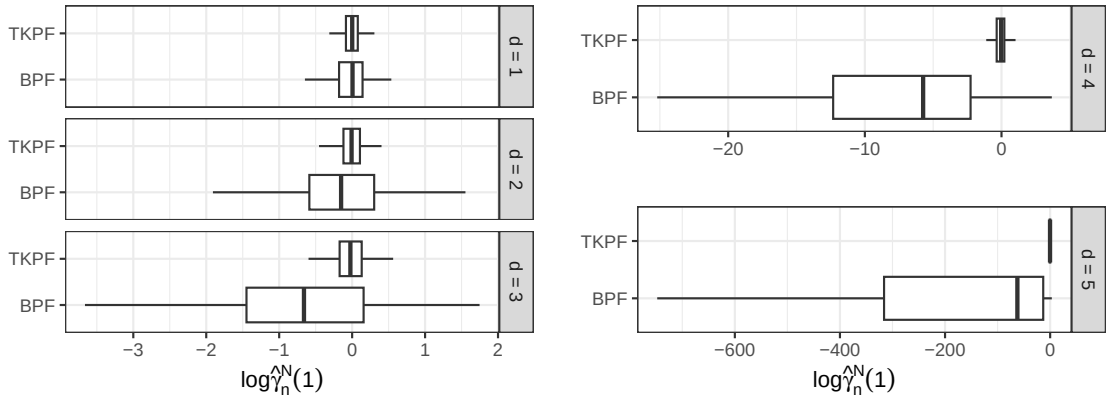

Figure 5: Box plots of (shifted) log normalising constant estimates of bootstrap PF and Terminal Knotset PF. The estimates were shifted by a constant (for each dimension d) for ease of comparison.

The figure demonstrates the variance reduction for normalising constant estimation for each dimension when using knots. We observe that the terminal knotset particle filter remains stable whilst the bootstrap particle filter becomes unstable as d increases.

This example demonstrates a variance reduction for a class of models that appears to be unconsidered in the literature. From the view of adaptation, Pitt and Shephard (1999) showed that 'full' adaptation could be implemented with non-linear functions and additive noise. Whilst in this example, we condition on the χ^2_ν distributed state and adapt only the Gaussian component in the extended state-space. Whereas from the marginalisation view, the general framework in Schön et al. (2005) does not include the possibility of marginalising a non-linear component, as we do here.

Many further generalisations of this example are possible. For example, a model in the form of (6) with a $\mathcal{K}_p$ knot leads to a tractable particle filter any conjugate K and G_p and arbitrary S. Such an S could represent an alternative distribution for the scale, or other components of the state-space, and need not be independent of past states.

8 Discussion

We have shown that knots unify and generalise 'full' adaptation and model-marginalisation as variance reduction techniques in sequential Monte Carlo algorithms. Our theory provides a comprehensive assessment of the asymptotic variance ordering implied by knots, the optimality of adapted knots, and demonstrates that repeated applications of adapted knots lead to algorithms with optimal variance.

In terms of particle filter design, we have re-emphasised the importance of 'full' adaptation (or adapted knots) by identifying the pitfall in the counterexample of Johansen and

Doucet (2008), and explaining how it can be avoided by not adapting at the terminal time point, or by adapting in a way that takes the test function φ into account. Further, given the guaranteed asymptotic variance ordering from knots, we can recommend that every particle filter be assessed to determine if there are one or more tractable knots that can be applied. In such an assessment, the cost of computing $K(G_p)$ and simulating from K^{G_p} should also be considered.

There are several future research directions to explore. Adapted knots have a connection with twisted Feynman–Kac models (Guarniero et al., 2017; Heng et al., 2020). On an extended state-space, a knot can be thought of as decomposing a kernel of a Feynman–Kac model into R_t and K_t, adapting K_t to the potential G_t, and simplifying the resulting model. Whilst a "twist" at time t decomposes the potential function into ψ_t and $\frac{G_t}{\psi}$ and adapts M_t to ψ_t. When $K_t = M_t$ and $\psi_t = G_t$ the knot-model and twist-model are equivalent up to a time-shift for $t < n$. Further developing this connection may suggest new methods for twisted Feynman–Kac models in particle filters. Similarly, look-ahead particle filters should also be considered (Lin et al., 2013). For now we note that, except for normalising constant estimation, it may be beneficial for a twisted model to use $\psi_n = 1$ as we know that terminal knots do not guarantee an asymptotic variance reduction for all relevant test functions.

Though we described how terminal knots reduce the asymptotic variance of terminal updated probability measures, we noted a complication arising from the requirement that the target function be $\phi = |\varphi - \hat{\eta}_n(\varphi)|$, where $\hat{\eta}_n(\varphi)$ is not known ahead of time. It would be possible to estimate $\hat{\eta}_n(\varphi)$ after the $n-1$ step of the particle filter, and use this to construct the requisite terminal Markov kernel and potential, from the application of a terminal knot to the appropriate ϕ-extended model, adaptively. Further, it is possible that this procedure could be done for multiple test functions, resulting in an SMC algorithm run until time $n-1$ which then branches for each function of interest. Methodological development of such an algorithm is left for future work.

The application of knots and related variance reductions for SMC samplers is also an area for future research. Consider an SMC sampler with target distribution $\hat{\eta}_{t-1}$ at time t with $M_t = RK$. Further, let R be k iterations of some $\hat{\eta}_{t-1}$-invariant MCMC kernel V, that is $R = V \otimes \cdots \otimes V$, and let K be a random uniform selection over the path generated by R. Applying the knot (t, R, K) to the resulting Feynman–Kac model recovers the proposed kernel and potential function used in the recent waste-free SMC algorithm (Dau and Chopin, 2022) at time t. Related connections to recent work on Hamiltonian Monte Carlo integrator snippets in an SMC-like algorithm (Andrieu et al., 2024) are also of interest.

References

Andrieu, C. and A. Doucet (2002). Particle filtering for partially observed Gaussian state space models. *Journal of the Royal Statistical Society Series B: Statistical Methodology 64*(4), 827–836.

Andrieu, C., M. C. Escudero, and C. Zhang (2024). Monte Carlo sampling with integrator snippets. *arXiv preprint arXiv:2404.13302*.

Bengtsson, H. (2025). *matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors)*. R package version 1.5.0.

Besançon, M., T. Papamarkou, D. Anthoff, A. Arslan, S. Byrne, D. Lin, and J. Pearson (2021). Distributions.jl: Definition and modeling of probability distributions in the juliastats ecosystem. *Journal of Statistical Software 98*(16), 1–30.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah (2017). Julia: A fresh approach to numerical computing. *SIAM Review 59*(1), 65–98.

Billingsley, P. (1995). *Probability and Measure* (3rd ed.). New York, NY: John Wiley & Sons, Inc.

Bouchet-Valat, M. and B. Kamiński (2023). Dataframes.jl: Flexible and fast tabular data in julia. *Journal of Statistical Software 107*(4), 1–32.

Cardoso, G., Y. J. el idrissi, S. L. Corff, and E. Moulines (2024). Monte Carlo guided denoising diffusion models for Bayesian linear inverse problems. In *The Twelfth International Conference on Learning Representations*.

Cérou, F., P. Del Moral, T. Furon, and A. Guyader (2012). Sequential Monte Carlo for rare event estimation. *Statistics and computing 22*(3), 795–808.

Chen, R. and J. S. Liu (2000). Mixture Kalman filters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62*(3), 493–508.

Creal, D. (2012). A survey of sequential Monte Carlo methods for economics and finance. *Econometric reviews 31*(3), 245–296.

Dai, C., J. Heng, P. E. Jacob, and N. Whiteley (2022). An invitation to sequential Monte Carlo samplers. *Journal of the American Statistical Association 117*(539), 1587–1600.

Dau, H.-D. and N. Chopin (2022). Waste-free sequential Monte Carlo. *Journal of the Royal Statistical Society Series B: Statistical Methodology 84*(1), 114–148.

Del Moral, P. (2004). *Feynman-Kac formulae*. New York: Springer-Verlag.

Del Moral, P., A. Doucet, and A. Jasra (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology 68*(3), 411–436.

Del Moral, P., A. Doucet, and A. Jasra (2012). On adaptive resampling strategies for sequential Monte Carlo methods. *Bernoulli 18*(1), 252–278.

Doucet, A. (1998). On sequential simulation-based methods for Bayesian filtering. Technical Report CUED-F-ENG-TR310, University of Cambridge, Department of Engineering.

Doucet, A., S. Godsill, and C. Andrieu (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and computing 10*, 197–208.

Doucet, A., N. J. Gordon, and V. Krishnamurthy (2002). Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on signal processing 49*(3), 613–624.

Doucet, A. and X. Wang (2005). Monte Carlo methods for signal processing: a review in the statistical signal processing context. *IEEE Signal Processing Magazine 22*(6), 152–170.

Endo, A., E. Van Leeuwen, and M. Baguelin (2019). Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers. *Epidemics 29*, 100363.

Fearnhead, P. and H. R. Künsch (2018). Particle filters and data assimilation. *Annual Review of Statistics and Its Application 5*(1), 421–449.

Gordon, N. J., D. J. Salmond, and A. F. Smith (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing) 140*(2), 107–113.

Guarniero, P., A. M. Johansen, and A. Lee (2017). The iterated auxiliary particle filter. *Journal of the American Statistical Association 112*(520), 1636–1647.

Gustafsson, F., F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund (2002). Particle filters for positioning, navigation, and tracking. *IEEE Transactions on signal processing 50*(2), 425–437.

Heng, J., A. N. Bishop, G. Deligiannidis, and A. Doucet (2020). Controlled sequential Monte Carlo. *The Annals of Statistics 48*(5), 2904–2929.

Johansen, A. M. and A. Doucet (2008). A note on auxiliary particle filters. *Statistics & Probability Letters 78*(12), 1498–1504.

Jouin, M., R. Gouriveau, D. Hissel, M.-C. Péra, and N. Zerhouni (2016). Particle filter-based prognostics: Review, discussion and perspectives. *Mechanical Systems and Signal Processing 72*, 2–31.

Kantas, N., A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin (2015). On particle methods for parameter estimation in state-space models. *Statistical Science 30*(3), 328–351.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of computational and graphical statistics 5*(1), 1–25.

Kong, A., J. S. Liu, and W. H. Wong (1994). Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association 89*(425), 278–288.

Lazaric, A., M. Restelli, and A. Bonarini (2007). Reinforcement learning in continuous action spaces through sequential Monte Carlo methods. *Advances in neural information processing systems 20*, 1–8.

Lee, A. and N. Whiteley (2018). Variance estimation in the particle filter. *Biometrika 105*(3), 609–625.

Lin, D., J. M. White, S. Byrne, D. Bates, A. Noack, J. Pearson, A. Arslan, K. Squire, D. Anthoff, T. Papamarkou, M. Besançon, J. Drugowitsch, M. Schauer, and contributors (2019, July). JuliaStats/Distributions.jl: a Julia package for probability distributions and associated functions.

Lin, M., R. Chen, and J. S. Liu (2013). Lookahead strategies for sequential Monte Carlo. *Statistical science 28*(1), 69–94.

Lioutas, V., J. W. Lavington, J. Sefas, M. Niedoba, Y. Liu, B. Zwartsenberg, S. Dabiri, F. Wood, and A. Scibior (2023). Critic sequential Monte Carlo. In *The Eleventh International Conference on Learning Representations*.

Liu, J. S. and R. Chen (1995). Blind deconvolution via sequential imputations. *Journal of the American Statistical Association 90*(430), 567–576.

Lopes, H. F. and R. S. Tsay (2011). Particle filters and Bayesian inference in financial econometrics. *Journal of Forecasting 30*(1), 168–209.

Macfarlane, M., E. Toledo, D. Byrne, P. Duckworth, and A. Laterre (2024). SPO: Sequential Monte Carlo policy optimisation. *Advances in Neural Information Processing Systems 37*, 1019–1057.

Mihaylova, L., A. Y. Carmi, F. Septier, A. Gning, S. K. Pang, and S. Godsill (2014). Overview of Bayesian sequential Monte Carlo methods for group and extended object tracking. *Digital Signal Processing 25*, 1–16.

Owen, A. and Y. Zhou (2000). Safe and effective importance sampling. *Journal of the American Statistical Association 95*(449), 135–143.

Pedersen, T. L. (2025). *patchwork: The Composer of Plots*. R package version 1.3.2.

Phillips, A., H.-D. Dau, M. J. Hutchinson, V. De Bortoli, G. Deligiannidis, and A. Doucet (2024, 21–27 Jul). Particle denoising diffusion sampler. In R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett, and F. Berkenkamp (Eds.), *Proceedings of the 41st International Conference on Machine Learning*, Volume 235 of *Proceedings of Machine Learning Research*, pp. 40688–40724. PMLR.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association 94*(446), 590–599.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schön, T., F. Gustafsson, and P.-J. Nordlund (2005). Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on signal processing 53*(7), 2279–2289.

Stachniss, C. and W. Burgard (2014). Particle filters for robot navigation. *Foundations and Trends® in Robotics 3*(4), 211–282.

Temfack, D. and J. Wyse (2024). A review of sequential Monte Carlo methods for real-time disease modeling. *arXiv preprint arXiv:2408.15739*.

Thrun, S. (2002). Particle filters in robotics. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 511–518.

Van Leeuwen, P. J., H. R. Künsch, L. Nerger, R. Potthast, and S. Reich (2019). Particle filters for high-dimensional geoscience applications: A review. *Quarterly Journal of the Royal Meteorological Society 145*(723), 2335–2365.

Wang, X., T. Li, S. Sun, and J. M. Corchado (2017). A survey of recent advances in particle filters and remaining challenges for multitarget tracking. *Sensors 17*(12), 2707.

Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani (2019). Welcome to the tidyverse. *Journal of Open Source Software 4*(43), 1686.

A Proof of main results

Proposition 3.12

Proof. Part 1. From Proposition 3.9 $\gamma_0^* = R_0$. Further, using the recursion (1), for $p \in [1:n-1]$, we have

$$\gamma_p^*(\varphi) = \hat{\gamma}_{p-1}^* M_p^*(\varphi) = \gamma_{p-1}^* \{K_{p-1}(G_{p-1}) \cdot K_{p-1}^{G_{p-1}} R_p(\varphi)\}.$$

Then by Proposition B.1, $\gamma_{p-1}^* \{K_{p-1}(G_{p-1}) \cdot K_{p-1}^{G_{p-1}} R_p(\varphi)\} = \gamma_{p-1}^* K_{p-1}\{G_{p-1} \cdot R_p(\varphi)\}$, and hence for $p \in [1:n-1]$,

$$\gamma_p^*(\varphi) = \gamma_{p-1}^* K_{p-1}\{G_{p-1} \cdot R_p(\varphi)\}.$$

From this we can see that $\gamma_1^*(\varphi) = \hat{\gamma}_0 R_1(\varphi)$, then $\gamma_2^*(\varphi) = \hat{\gamma}_1 R_2(\varphi)$, and the proof for Part 1 can conclude by induction.

For Part 2, we use Part 1 to see that $\gamma_0^* K_0(\varphi) = R_0 K_0(\varphi) = M_0(\varphi) = \gamma_0(\varphi)$ and further that $\gamma_p^* K_p(\varphi) = \hat{\gamma}_{p-1} R_p K_p(\varphi) = \hat{\gamma}_{p-1} M_p(\varphi) = \gamma_p(\varphi)$ for $p \in [1:n-1]$. □

Proposition 3.13

Proof. For Part 1, we expand (1) at time n to get

$$\begin{aligned}
\gamma_n^*(\varphi) &= \gamma_{n-1}^* \{G_{n-1}^* \cdot M_n^*(\varphi)\} \\
&= \hat{\gamma}_{p-2} R_{n-1}\{K_{n-1}(G_{n-1}) \cdot K_{n-1}^{G_{n-1}} M_n(\varphi)\} \\
&= \hat{\gamma}_{p-2} R_{n-1} K_{n-1}[G_{n-1} \cdot M_n(\varphi)] \\
&= \hat{\gamma}_{p-2} M_{n-1}[G_{n-1} \cdot M_n(\varphi)] \\
&= \gamma_n(\varphi)
\end{aligned}$$

from Proposition 3.9, Proposition 3.12, and Proposition B.1. Since $G_n^* = G_n$ the updated terminal measures are also equivalent.

Part 2 follows directly from Part 1 by normalising.

For Part 3, for the predictive measures we have $\gamma_n^*(1) = \gamma_n(1)$ from Part 1. Further, for $p \in [0:n-1]$, Proposition 3.12 (Part 2) yields $\gamma_p^* K_p(1) = \gamma_p(1)$ then we note that $\gamma_p^* K_p(1) = \gamma_p^*(1)$ to gain $\gamma_p^*(1) = \gamma_p(1)$.

For the updated measures, for $p \in [1:n-1]$, Proposition 3.12 (Part 1) yields $\gamma_p^*(1) = \hat{\gamma}_{p-1} R_p(1) = \hat{\gamma}_{p-1}(1)$. Then note that $\gamma_p^*(1) = \hat{\gamma}_{p-1}^*(1)$ to gain $\hat{\gamma}_p^*(1) = \hat{\gamma}_p(1)$ for $p \in [0:n-2]$. For $p = n-1$ we have $\hat{\gamma}_{n-1}^*(1) = \gamma_n^*(1) = \gamma_n(1) = \hat{\gamma}_{n-1}(1)$ by Part 1, and for $p = n$, $\hat{\gamma}_n^*(1) = \hat{\gamma}_n(1)$ again by Part 1.

□

Theorem 4.1

Proof. From Proposition B.2 we have $Q^*_{p,n} = K_p Q_{p,n}$ almost everywhere for $p \in [0:n-1]$, and from Proposition 3.12 $\gamma^*_0 = R_0$ and $\gamma^*_p = \hat{\gamma}_{p-1} R_p$ for $p \in [1:n-1]$. Therefore, using Jensen's inequality, for $p \in [1:n-1]$

$$\gamma^*_p\{Q^*_{p,n}(\varphi)^2\} = \hat{\gamma}_{p-1} R_p\{K_p Q_{p,n}(\varphi)^2\} \leq \hat{\gamma}_{p-1} R_p K_p\{Q_{p,n}(\varphi)^2\} = \gamma_p\{Q_{p,n}(\varphi)^2\}, \text{ and}$$
$$\gamma^*_0\{Q^*_{0,n}(\varphi)^2\} = R_0\{K_0 Q_{0,n}(\varphi)^2\} \leq R_0 K_0\{Q_{0,n}(\varphi)^2\} = \gamma_0\{Q_{0,n}(\varphi)^2\}.$$

Whilst for $p = n$, and $Q^*_{n,n} = Q_{n,n} = \mathrm{Id}$ by definition and $\gamma^*_n = \gamma_n$ from Proposition 3.13 so $\gamma^*_n\{Q_{n,n}(\varphi)^2\} = \gamma_n\{Q_{n,n}(\varphi)^2\}$. From the above inequalities, and using $\gamma^*_p(1) = \gamma_p(1)$ and $\eta^*_n = \eta_n$ from Proposition 3.13 we can state that, for $p \in [0:n-1]$,

$$v^*_{p,n}(\varphi) = \frac{\gamma^*_p(1)\gamma^*_p(Q^*_{p,n}(\varphi)^2)}{\gamma^*_n(1)^2} - \eta^*_n(\varphi)^2 \leq \frac{\gamma_p(1)\gamma_p(Q_{p,n}(\varphi)^2)}{\gamma_n(1)^2} - \eta_n(\varphi)^2 = v_{p,n}(\varphi),$$

and therefore $\sigma^2_*(\varphi) \leq \sigma^2(\varphi)$ by also noting that $v^*_{n,n}(\varphi) = v_{n,n}(\varphi)$.

To quantify the reduction in variance, we can see that

$$\begin{aligned}
\gamma_n\{Q_{n,n}(\varphi)^2\} - \gamma^*_n\{Q_{n,n}(\varphi)^2\} &= 0, \\
\gamma_p\{Q_{p,n}(\varphi)^2\} - \gamma^*_p\{Q^*_{p,n}(\varphi)^2\} &= \hat{\gamma}_{p-1} R_p K_p\{Q_{p,n}(\varphi)^2\} - \hat{\gamma}_{p-1} R_p\{K_p Q_{p,n}(\varphi)^2\} \\
&= \hat{\gamma}_{p-1} R_p \left[K_p\{Q_{p,n}(\varphi)^2\} - K_p Q_{p,n}(\varphi)^2\right] \\
&= \hat{\gamma}_{p-1} R_p \left[\mathrm{Var}_{K_p}\{Q_{p,n}(\varphi)\}\right], \text{ for } p \in [1:n-1], \\
\gamma_0\{Q_{0,n}(\varphi)^2\} - \gamma^*_0\{Q^*_{0,n}(\varphi)^2\} &= R_0 K_0\{Q_{0,n}(\varphi)^2\} - R_0\{K_0 Q_{0,n}(\varphi)^2\} \\
&= R_0 \left[K_0\{Q_{0,n}(\varphi)^2\} - K_0 Q_{0,n}(\varphi)^2\right] \\
&= R_0 \left[\mathrm{Var}_{K_0}\{Q_{0,n}(\varphi)\}\right]
\end{aligned}$$

which combined with the measure equivalences with Proposition 3.13 yields the desired result. From this quantification and the original inequality we can conclude that the inequality is indeed strict if the ν_p-averaged variance terms in Theorem 4.1 are strictly positive. □

Corollary 4.2

Proof. For Part 1, if $\varphi \in \mathcal{F}(\mathcal{M})$ then $\varphi - \eta_n(\varphi) \in \mathcal{F}(\mathcal{M})$ then from Theorem 4.1 we have $\sigma^2_*(\varphi - \eta^*_n(\varphi)) \leq \sigma^2(\varphi - \eta_n(\varphi))$, noting that $\eta^*_n(\varphi) = \eta_n(\varphi) < \infty$.

For Part 2, using (4) the updated asymptotic variance of $\mathcal{M}^*$ can be written as $\hat{\sigma}^2_*(\varphi) = \sigma^2_*(G^*_n \cdot \varphi)/\eta^*_n(G^*_n)^2$. Then we can state

$$\hat{\sigma}^2_*(\varphi) = \frac{\sigma^2_*(G^*_n \cdot \varphi)}{\eta^*_n(G^*_n)^2} = \frac{\sigma^2_*(G_n \cdot \varphi)}{\eta_n(G_n)^2}$$

since we have $G_n^* = G_n$ by Proposition 3.9 and $\eta_n^* = \eta_n$ by Proposition 3.13. Lastly, if $\varphi \in \hat{\mathcal{F}}(\mathcal{M})$ then $G_n \cdot \varphi \in \mathcal{F}(\mathcal{M})$ and so from Theorem 4.1 $\sigma_*^2(G_n \cdot \varphi) \leq \sigma^2(G_n \cdot \varphi)$. Therefore,

$$\hat{\sigma}_*^2(\varphi) \leq \frac{\sigma^2(G_n \cdot \varphi)}{\eta_n(G_n)^2} = \hat{\sigma}^2(\varphi),$$

and the inequality is strict under the same conditions as Theorem 4.1. Part 3 follows in the same manner as Part 1, but for updated models. □

Theorem 4.4

Proof. If $\mathcal{M}^* \preccurlyeq \mathcal{M}$ then there exists a sequence of knotsets $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_m$ such that $\mathcal{M}^* = \mathcal{K}_m * \cdots * \mathcal{K}_1 * \mathcal{M}$ for some $m \in \mathbb{N}_1$. Let $\mathcal{M}_s = \mathcal{K}_s * \mathcal{M}_{s-1}$ for $s \in [1:m]$ with $\mathcal{M}_0 = \mathcal{M}$ and let the predictive and asymptotic variance maps of $\mathcal{M}_s$ be σ_s^2 and $\hat{\sigma}_s^2$ respectively. We note that $\mathcal{M}_m = \mathcal{M}^*$. From Theorem 4.1 and Corollary 4.2 we can state that $\sigma_s^2(\varphi) \leq \sigma_{s-1}^2(\varphi)$ for $\varphi \in \mathcal{F}(\mathcal{M})$ and $\hat{\sigma}_s^2(\varphi) \leq \hat{\sigma}_{s-1}^2(\varphi)$ for $\varphi \in \hat{\mathcal{F}}(\mathcal{M})$ over $s \in [1:m]$. Each inequality will be strict under the same conditions as Theorem 4.1. Therefore we can state that $\sigma_*^2(\varphi) = \sigma_m^2(\varphi) \leq \sigma_0^2(\varphi) = \sigma^2(\varphi)$ if $\varphi \in \mathcal{F}(\mathcal{M})$ and $\hat{\sigma}_*^2(\varphi) = \hat{\sigma}_m^2(\varphi) \leq \hat{\sigma}_0^2(\varphi) = \hat{\sigma}^2(\varphi)$ if $\varphi \in \hat{\mathcal{F}}(\mathcal{M})$. The analogous results for the probability measures follow since $\varphi - \eta_n(\varphi) \in \mathcal{F}(\mathcal{M})$ and $\varphi - \hat{\eta}_n(\varphi) \in \hat{\mathcal{F}}(\mathcal{M})$. □

Theorem 5.1

Proof. First consider the case of knots. For $t > 0$ and $\mathcal{K} = (t, R, K)$, let $\mathcal{R} = (t, \mathrm{Id}, R)$ then $\mathcal{R} * \mathcal{K} * \mathcal{M} = \mathcal{K}^\diamond * \mathcal{M}$ by Proposition B.4 and hence $\mathcal{K}^\diamond * \mathcal{M} \preccurlyeq \mathcal{K} * \mathcal{M}$. For a knot at time $t = 0$, $\mathcal{K} = (0, R, K)$ where R is a probability measure. As such, let $\mathcal{R} = (0, \delta_0, K_0)$ where $K_0(0, \cdot) = R$ to reach the same conclusion.

For knotsets, Proposition B.5 ensures the existence of a knot $\mathcal{R}$ such that $\mathcal{R} * \mathcal{K} * \mathcal{M} = \mathcal{K}^\diamond * \mathcal{M}$ for any model $\mathcal{M}$ and knotset $\mathcal{K}$. Therefore, $\mathcal{K}^\diamond * \mathcal{M} \preccurlyeq \mathcal{K} * \mathcal{M}$. □

Theorem 5.2

Proof. We have that $\mathcal{M}_m = \mathcal{K}_{m-1} * \mathcal{M}_{m-1}$ and hence, by Proposition B.5, there exists a knotset $\mathcal{R}_m$ that completes $\mathcal{K}_{m-1}$, that is $\mathcal{R}_m * \mathcal{M}_m = \mathcal{K}_{m-1}^\diamond * \mathcal{M}_{m-1}$ where $\mathcal{K}_{m-1}^\diamond$ is the adapted knotset for $\mathcal{M}_{m-1}$. We use Proposition B.7 to find

$$\begin{aligned}
\mathcal{R}_m * \mathcal{M}_m &= \mathcal{K}_{m-1}^\diamond * \mathcal{K}_{m-2} * \cdots * \mathcal{K}_0 * \mathcal{M}_0 \\
&= \mathcal{J}_{m-1} * \cdots * \mathcal{J}_1 * \mathcal{K}_0^\star * \mathcal{M}_0 \\
&= \mathcal{J}_{m-1} * \cdots * \mathcal{J}_1 * \mathcal{M}_1^\star,
\end{aligned}$$

for some knotsets $\mathcal{J}_t$ for $t \in [1:m-1]$, where the first equality follows by definition of $\mathcal{M}_{m-1}$. The process of completing the first knotset (now $\mathcal{J}_{m-1}$) with a $\mathcal{R}_{m-1}$ and changing

the adapted knotset in the sequence can be repeated until we have

$$\mathcal{R}_1 * \cdots * \mathcal{R}_{m-1} * \mathcal{R}_m * \mathcal{M}_m = \mathcal{M}_m^*,$$

and hence $\mathcal{M}_m^\star \preccurlyeq \mathcal{M}_m$. □

Theorem 5.3

Proof. The model $\mathcal{M}_n$ in Example 5.4 satisfies the requirements on the asymptotic variance terms for a sequence of knots. This can be seen by noting all potential functions are constant in this model before the terminal time, hence $v_{p,n}^*(\varphi) = 0$ for $p \in [0 : n-1]$. As for the final term we have

$$\begin{aligned} v_{n,n}^*(\varphi) &= \gamma_n^*(G_n^2) - \eta_n^*(\varphi)^2 \\ &= \prod_{p=0}^{n-1} \eta_p(G_p)\eta_n(G_n^2) - \eta_n(\varphi)^2 \\ &= \gamma_n(1)\eta_n(G_n^2) - \eta_n(\varphi)^2 \\ &= \gamma_n(G_n^2) - \eta_n(\varphi)^2 \\ &= v_{n,n}(\varphi). \end{aligned}$$

As knots are a special case of knotsets, a sequence of knotsets satisfying the variance conditions also exists. □

Proposition 6.2

Proof. For Part 1, since no elements of the Feynman–Kac model are changed prior to time n, we have that $\hat{\gamma}_{n-1}^\phi = \hat{\gamma}_{n-1}$ and then

$$\gamma_n^\phi(\varphi_1 \otimes \varphi_2) = \hat{\gamma}_{n-1} M_n^\phi(\varphi_1 \otimes \varphi_2) = \hat{\gamma}_{n-1} M_n(\varphi_1 \cdot \varphi_2) = \gamma_n(\varphi_1 \cdot \varphi_2). \tag{7}$$

From this result, we see that $\gamma_n^\phi(1 \otimes \varphi) = \gamma_n(\varphi)$ as required for the predictive measure. As for the updated measure in Part 2 we have

$$\begin{aligned} \hat{\gamma}_n^\phi(1 \otimes \varphi) &= \gamma_n^\phi(G_n^\phi \cdot [1 \otimes \varphi]) \\ &= \gamma_n^\phi([G_n \cdot \phi] \otimes [\varphi \cdot \phi^{-1}]) \\ &= \gamma_n(G_n \cdot \varphi \cdot \phi \cdot \phi^{-1}) \\ &= \hat{\gamma}_n(\varphi \cdot \phi \cdot \phi^{-1}) \\ &= \hat{\gamma}_n(\varphi), \end{aligned}$$

using (7) for the third equality and the fact that ϕ is $\hat{\gamma}_n$-a.e. positive for the final equality. For Part 3, starting with the $Q_{p,n}$ terms we first consider Q_n^ϕ for

$$\begin{aligned} Q_n^\phi(G_n^\phi \cdot [1 \otimes \varphi]) &= G_{n-1}^\phi \cdot M_n^\phi(G_n^\phi \cdot [1 \otimes \varphi]) \\ &= G_{n-1}^\phi \cdot (M_n \otimes \mathrm{Id})([G_n \cdot \phi] \otimes [\varphi \cdot \phi^{-1}]) \\ &= G_{n-1} \cdot M_n(G_n \cdot \varphi \cdot \phi \cdot \phi^{-1}) \\ &= Q_n(G_n \cdot \varphi), \end{aligned}$$

almost everywhere w.r.t. γ_{n-1}. Then since $M_p^\phi = M_p$ and $G_p^\phi = G_p$ for $p \in [0:n-1]$, we have $Q_p^\phi = Q_p$ for $p \in [1:n-1]$, and can state that $Q_{p,n}^\phi(G_n^\phi) = Q_{p,n}(G_n \cdot \varphi)$ almost everywhere under γ_p for $p \in [0:n-1]$. It is also true that $Q_{n,n}^\phi = \mathrm{Id}$ and $Q_{n,n} = \mathrm{Id}$ by definition, where each identity kernel is defined on their respective measure space. As such, combining with Part 2 we can state that $\hat{v}_{p,n}^\phi(1 \otimes \varphi) = \hat{v}_{p,n}(\varphi)$ for $p \in [0:n-1]$.

Whereas for $p = n$, we first note $\{G_n^\phi \cdot [1 \otimes \varphi]\}^2 = [G_n \cdot \phi]^2 \otimes [\varphi \cdot \phi^{-1}]^2$, so

$$\begin{aligned} M_n^\phi(\{G_n^\phi \cdot [1 \otimes \varphi]\}^2) &= (M_n \otimes \mathrm{Id})([G_n \cdot \phi]^2 \otimes [\varphi \cdot \phi^{-1}]^2) \\ &= M_n([G_n \cdot \varphi]^2), \end{aligned}$$

almost everywhere under $\hat{\gamma}_{n-1}$. Then for the nth variance term

$$\begin{aligned} \hat{v}_{n,n}^\phi([1 \otimes \varphi]) &= \frac{\gamma_n^\phi(1)\gamma_n^\phi(\{G_n^\phi \cdot [1 \otimes \varphi]\}^2)}{\gamma_n^\phi(G_n^\phi)^2} - \eta_n^\phi(G_n^\phi \cdot [1 \otimes \varphi]) \\ &= \frac{\gamma_n(1)\hat{\gamma}_{n-1}M_n^\phi(\{G_n^\phi \cdot [1 \otimes \varphi]\}^2)}{\gamma_n(G_n)^2} - \eta_n(G_n \cdot [1 \otimes \varphi]) \\ &= \frac{\gamma_n(1)\hat{\gamma}_{n-1}M_n(\{G_n \cdot \varphi\}^2)}{\gamma_n(G_n)^2} - \eta_n(G_n \cdot [1 \otimes \varphi]) \\ &= \hat{v}_{n,n}(\varphi), \end{aligned}$$

using Part 2 for equality of marginal measures. Hence, $\hat{\sigma}_\phi^2([1 \otimes \varphi]) = \hat{\sigma}^2(\varphi)$ since all $p \in [0:n]$ terms are equal. □

Proposition 6.5

Proof. By compatibility condition (i), $\mathcal{M}^\circ = (M_{0:n}^\circ, G_{0:n}^\circ)$ satisfies $M_n^\circ = U_1 \otimes V_1^{G_n \cdot \phi}$ and $G_n^\circ = V_1(G_n \cdot \phi) \otimes \phi^{-1}$, for some Markov kernels U_1 and V_1, reference model $\mathcal{M} = (M_{0:n}, G_{0:n})$, and target function ϕ. Let $\mathcal{M}^* = \mathcal{K} * \mathcal{M}^\circ = (M_{0:n}^*, G_{0:n}^*)$.

First case. If $\mathcal{K}$ is a non-terminal knot then $M_n^* = K^{G_{n-1}}U_1 \otimes V_1^{G_n \cdot \phi}$ and $G_n^* = V_1(G_n \cdot \phi) \otimes \phi^{-1}$ for some Markov kernel K. Letting $U_2 = K^{G_{n-1}}U_1$ and $V_2 = V_1$ shows that $\mathcal{M}^*$ satisfies compatibility condition (i).

Second case. If $\mathcal{K}$ is a terminal knot then $M_n^* = R \otimes K^{V_1(G_n \cdot \phi)}V_1^{G_n \cdot \phi}$ and $G_n^* = KV_1(G_n \cdot \phi) \otimes \phi^{-1}$ for some Markov kernels R and K such that $RK = U_1$. By Proposition B.3, we can state $M_n^* = R \otimes (KV_1)^{G_n \cdot \phi}$ and hence letting $U_2 = R$ and $V_2 = KV_1$ shows that $\mathcal{M}^*$ satisfies compatibility condition (i). □

Proposition 6.7

Proof. For Part 1, since $M_p^* = M_p$ and $G_p^* = G_p$ for $p \in [0:n-1]$ all marginal measures are equal at these times.

By Definition 6.3 compatibility condition (i) there exists P_1 P_2, H, and ϕ such that $M_n = P_1 \otimes P_2$ and $G_n = H \otimes \phi^{-1}$. Therefore for Part 2 we can consider

$$\begin{aligned}
\hat{\gamma}_n^*(1 \otimes \varphi) &= \hat{\gamma}_{n-1}^* M_n^*(G_n^* \cdot [1 \otimes \varphi]) \\
&= \hat{\gamma}_{n-1}(R \otimes K^H P_2)(K(H) \otimes [\phi^{-1} \cdot \varphi]) \\
&= \hat{\gamma}_{n-1} R\{K(H) \cdot K^H P_2(\phi^{-1} \cdot \varphi)\} \\
&= \hat{\gamma}_{n-1} RK[H \cdot P_2(\phi^{-1} \cdot \varphi)]
\end{aligned}$$

by Proposition B.1. Then, by Definition 6.3 compatibility condition (ii), we have

$$\begin{aligned}
\hat{\gamma}_n^*(1 \otimes \varphi) &= \hat{\gamma}_{n-1} P_1[H \cdot P_2(\phi^{-1} \cdot \varphi)] \\
&= \hat{\gamma}_{n-1}(P_1 \otimes P_2)([H \otimes \phi^{-1}] \cdot [1 \otimes \varphi]) \\
&= \hat{\gamma}_{n-1} M_n(G_n \cdot [1 \otimes \varphi]) \\
&= \hat{\gamma}_n(1 \otimes \varphi).
\end{aligned}$$

□

Theorem 6.12

Proof. For the $Q_{p,n}$ terms we first consider Q_n^* for

$$Q_n^*(G_n^* \cdot \bar{\varphi}) = G_{n-1}^* \cdot M_n^*(G_n^* \cdot \bar{\varphi}) = G_{n-1} \cdot M_n(G_n \cdot \bar{\varphi}) = Q_n(G_n \cdot \bar{\varphi})$$

from Proposition B.8 and noting $G_{n-1}^* = G_{n-1}$. Then since $M_p^* = M_p$ and $G_p^* = G_p$ for $p \in [0:n-1]$, we have $Q_p^* = Q_p$ for $p \in [1:n-1]$, and can state that $Q_{p,n}^*(G_n^* \cdot \bar{\varphi}) = Q_{p,n}(G_n \cdot \bar{\varphi})$ for $p \in [0:n-1]$ and $Q_{n,n}^* = Q_{n,n} = \mathrm{Id}$ by definition. As such, combining with Proposition 6.7 (Part 1), we can state that $\hat{v}_{p,n}^*(\bar{\varphi}) = \hat{v}_{p,n}(\bar{\varphi})$ for $p \in [0:n-1]$. Therefore, we can state that $\hat{\sigma}^2(\bar{\varphi}) - \hat{\sigma}_*^2(\bar{\varphi}) = \hat{v}_{n,n}(\bar{\varphi}) - \hat{v}_{n,n}^*(\bar{\varphi})$.

From (4) we find

$$\begin{aligned}
\hat{v}_{n,n}(\bar{\varphi}) - \hat{v}_{n,n}^*(\bar{\varphi}) &= \frac{\eta_n(\{G_n \cdot \bar{\varphi}\}^2) - \eta_n(G_n \cdot \bar{\varphi})^2}{\eta_n(G_n)^2} - \frac{\eta_n^*(\{G_n^* \cdot \bar{\varphi}\}^2) - \eta_n^*(G_n^* \cdot \bar{\varphi})^2}{\eta_n^*(G_n^*)^2} \\
&= \frac{\hat{\eta}_{n-1} M_n(\{G_n \cdot \bar{\varphi}\}^2)}{\eta_n(G_n)^2} - \frac{\hat{\eta}_{n-1}^* M_n^*(\{G_n^* \cdot \bar{\varphi}\}^2)}{\eta_n(G_n)^2} \\
&= \frac{\hat{\eta}_{n-1}\left[M_n(\{G_n \cdot \bar{\varphi}\}^2) - M_n^*(\{G_n^* \cdot \bar{\varphi}\}^2)\right]}{\eta_n(G_n)^2},
\end{aligned}$$

using $\eta_n^*(G_n^* \cdot \bar{\varphi}) = \frac{\hat{\gamma}_n^*(\bar{\varphi})}{\gamma_n^*(1)} = \frac{\hat{\gamma}_n(\bar{\varphi})}{\gamma_n(1)} = \eta_n(G_n \cdot \bar{\varphi})$ and $\hat{\eta}_{n-1}^* = \hat{\eta}_{n-1}$ from Proposition 6.7.

Then we compute the individual terms of the difference, first noting that $M_n = P_1 \otimes P_2$ and $G_n = H \otimes \phi^{-1}$ for some P_1, P_2, H, and ϕ by compatibility. As such, we have

$$\begin{aligned} M_n^*(\{G_n^* \cdot \bar{\varphi}\}^2) &= (R \otimes K^H P_2)(K(H)^2 \otimes [\phi^{-1} \cdot \varphi]^2) \\ &= R\{K(H)^2 \cdot K^H P_2([\phi^{-1} \cdot \varphi]^2)\} \\ &= R\{K(H) \cdot K[H \cdot P_2([\phi^{-1} \cdot \varphi]^2)]\} \end{aligned}$$

by Proposition B.1, and by simplification

$$\begin{aligned} M_n(\{G_n \cdot \bar{\varphi}\}^2) &= (P_1 \otimes P_2)(H^2 \otimes [\phi^{-1} \cdot \varphi]^2) \\ &= P_1\{H^2 \cdot P_2([\phi^{-1} \cdot \varphi]^2)\} \\ &= RK\{H^2 \cdot P_2([\phi^{-1} \cdot \varphi]^2)\}. \end{aligned}$$

Let $H' = H \cdot P_2([\phi^{-1} \cdot \varphi]^2)$ then we can state

$$\begin{aligned} M_n(\{G_n \cdot \bar{\varphi}\}^2) - M_n^*(\{G_n^* \cdot \bar{\varphi}\}^2) &= R\left\{K(H \cdot H') - K(H) \cdot K(H')\right\} \\ &= R\{\mathrm{Cov}_K(H, H')\}, \end{aligned}$$

completing the proof. □

Corollary 6.13

Proof. From (repeated application of) Proposition 6.5 there exists Markov kernels U, V such that $M_n^\circ = U \otimes V^{G_n \cdot \phi}$ and $G_n^\circ = V(G_n \cdot \phi) \otimes \phi^{-1}$. With $P_1 = U$, $P_2 = V^{G_n \cdot \phi}$, and $H = V(G_n \cdot \phi)$, we note that

$$\begin{aligned} H \cdot P_2[\{\phi^{-1} \cdot \varphi\}^2] &= V(G_n \cdot \phi) \cdot V^{G_n \cdot \phi}[\{\phi^{-1} \cdot \varphi\}^2] \\ &= V(G_n \cdot \phi) = H, \end{aligned}$$

since $\phi = |\varphi|$. Then from Theorem 6.12 we can state

$$\hat{\sigma}_\circ^2(1 \otimes \varphi) - \hat{\sigma}_*^2(1 \otimes \varphi) = \frac{\hat{\eta}_{n-1}^\circ R\{\mathrm{Var}_K(V[G_n \cdot \phi])\}}{\eta_n^\circ(G_n^\circ)}. \tag{8}$$

Hence, we have

$$\hat{\sigma}_*^2(1 \otimes \varphi) \leq \hat{\sigma}_\circ^2(1 \otimes \varphi) \tag{9}$$

and the inequality will be strict if $\hat{\eta}_{n-1}^\circ R\{\mathrm{Var}_K(V[G_n \cdot \phi])\} > 0$.

Equation (9) establishes that terminal knots applied to ϕ-extended models with knots lead to an asymptotic variance ordering for test functions of the form $1 \otimes \varphi$ when $\phi = |\varphi|$. Since $\mathcal{M}^\circ$ has this form, every knot in the assumed knot-model sequence reduces the variance of a test function of the form $1 \otimes \varphi$. This follows from Proposition 4.1 (if a standard knot) or by (9) (if a terminal knot). Hence we can state that $\hat{\sigma}_\circ^2(1 \otimes \varphi) \leq \hat{\sigma}_\phi^2(1 \otimes \varphi)$

where $\hat{\sigma}^2_\phi$ is the asymptotic variance map for $\mathcal{M}^\phi$. Then from Proposition 6.2 we have $\hat{\sigma}^2_\phi(1 \otimes \varphi) = \hat{\sigma}^2(\varphi)$ to complete the first part of the proof.

If $\mathcal{K}$ is the first knot to be applied to $\mathcal{M}^\phi$ then we have $V = \mathrm{Id}$, $\hat{\eta}^\circ_{n-1} = \hat{\eta}^\phi_{n-1}$, and $\hat{\sigma}^2_\circ(1 \otimes \varphi) = \hat{\sigma}^2(\varphi)$ as $\mathcal{M}^\circ = \mathcal{M}^\phi$ and $\hat{\eta}^\phi_{n-1} = \hat{\eta}_{n-1}$ as ϕ-extension does not change the non-terminal measures. Using these result in conjunction with (8) leads to the final result. □

Theorem 6.15

Proof. By definition $\mathcal{M}^* = \mathcal{K}_m * \cdots * \mathcal{K}_1 * \mathcal{M}^\circ$ so the variance inequalities follow from iterated applications of Corollary 6.13 (if a terminal knot) or Theorem 4.1 (if a standard knot). Further, $\mathcal{M}^\circ = \mathcal{K}'_{m'} * \cdots * \mathcal{K}'_1 * \mathcal{M}^\phi$ by definition so the equalities between measures follows by iterated applications of Proposition 6.7 (if a terminal knot) and Proposition 3.13 (if a standard knot) to the entire sequence of models. Finally, Proposition 6.2 ensures the equivalence of the ϕ-extended model to the reference model $\mathcal{M}$. □

Theorem 6.16

Proof. First note that since $\mathcal{M}^\circ$ satisfies Part 2 of Definition 6.14 with respect to $\mathcal{M}$, by definition, $\mathcal{K} * \mathcal{M}^\circ$ also satisfies this condition. Then by (repeated application of) Proposition 6.5 there exists Markov kernels U, V such that $M^\circ_n = U \otimes V^{G_n \cdot \phi}$ and $G^\circ_n = V(G_n \cdot \phi) \otimes \phi^{-1}$. Let $\mathcal{K} = (n, R, K)$ and $\mathcal{R} = (n, \mathrm{Id}, R)$, consider the model $\mathcal{M}^* = \mathcal{R} * \mathcal{K} * \mathcal{M}^\circ$, noting that $U = RK$ by compatibility. Hence $M^*_n = \mathrm{Id} \otimes R^{K(H)} K^H V^{G_n \cdot \phi}$ and $G^*_n = RK(H) \otimes \phi^{-1}$ where $H = V(G_n \cdot \phi)$ from the sequential application of the terminal knots. We can simplify the Markov kernel $M^*_n = \mathrm{Id} \otimes (RKV)^{G_n \cdot \phi} = \mathrm{Id} \otimes (UV)^{G_n \cdot \phi}$ using Proposition B.3 twice. The potential function simplifies to $G^*_n = UV(G_n \cdot \phi) \otimes \phi^{-1}$.

Now consider the model $\mathcal{M}^\diamond = \mathcal{K}^\diamond * \mathcal{M}^\circ$ where $\mathcal{K}^\diamond$ is the adapted kernel for $\mathcal{M}^\circ$. The adapted knot is $\mathcal{K}^\diamond = (n, \mathrm{Id}, U)$ and hence $M^\diamond_n = \mathrm{Id} \otimes U^H V^{G_n \cdot \phi} = \mathrm{Id} \otimes (UV)^{G_n \cdot \phi}$ by Proposition B.3 and $G^*_n = U(H) \otimes \phi^{-1} = UV(G_n \cdot \phi) \otimes \phi^{-1}$. Hence, $M^\diamond_n = M^*_n$ and $G^\diamond_n = G^*_n$, so that we can conclude $\mathcal{M}^* = \mathcal{M}^\diamond$.

Finally, we can state $\mathcal{M}^* = \mathcal{R} * \mathcal{K} * \mathcal{M}^\circ = \mathcal{K}^\diamond * \mathcal{M}^\circ$ and hence $\mathcal{K}^\diamond * \mathcal{M}^\circ \preccurlyeq_\phi \mathcal{K} * \mathcal{M}^\circ$ with respect to $\mathcal{M}$. □

Corollary 6.17

Proof. Consider $\mathcal{M}_n$ defined by the sequence in Example 5.4 using initial model $\mathcal{M}_0 = \mathcal{M}^\phi$ with target function $\phi = |\varphi|$ and reference model $\mathcal{M}$. Let $\mathcal{M}^\star = \mathcal{K}^\diamond_n * \mathcal{M}_n$ where $\mathcal{K}^\diamond_n$ is the adapted terminal knot for $\mathcal{M}_n$.

First note that from Theorem 5.3 we have $\hat{v}^\star_{p,n}(\varphi) = 0$ for $p \in [0 : n-1]$ as the application of the terminal knot $\mathcal{K}^\diamond_n$ to $\mathcal{M}_n$ will not change the asymptotic variance terms at earlier times.

From Example 6.18 we can state $\eta_n^\star = \delta_0 \otimes \eta_n^{G_n \cdot \phi}$ and $G_n^\star = \eta_n(G_n \cdot \phi) \otimes \phi^{-1}$. Hence, $\eta_n^\star(G_n^\star \cdot \bar{\varphi}) = (\delta_0 \otimes \eta_n^{G_n \cdot \phi})\{\eta_n(G_n \cdot \phi) \otimes (\phi^{-1} \cdot \varphi)\} = \eta_n(G_n \cdot \varphi)$ and $\eta_n^\star(\{G_n^\star \cdot \bar{\varphi}\}^2) = (\delta_0 \otimes \eta_n^{G_n \cdot \phi})\{\eta_n(G_n \cdot \phi)^2 \otimes 1\} = \eta_n(G_n \cdot \phi)^2$.

Combining these results with (4) we find that

$$\begin{aligned}\hat{v}_{n,n}^*(\bar{\varphi}) &= \frac{\eta_n^*(\{G_n^* \cdot \bar{\varphi}\}^2) - \eta_n^*(G_n^* \cdot \bar{\varphi})^2}{\eta_n^*(G_n^*)^2} \\ &= \frac{\eta_n(G_n \cdot \phi)^2 - \eta_n(G_n \cdot \varphi)^2}{\eta_n(G_n)^2} \\ &= \hat{\eta}_n(\phi)^2 - \hat{\eta}_n(\varphi)^2.\end{aligned}$$

□

B Supporting results

Proposition B.1 (Kernel untwisting)**.** *If $K^H : (\mathsf{X}, \mathcal{Y}) \to [0,1]$ is a twisted Markov kernel and $\varphi : \mathsf{Y} \to [-\infty, \infty]$ is measurable then*

$$K(H) \cdot K^H(\varphi) = K(H \cdot \varphi),$$

λ-a.e. for any measure λ on $(\mathsf{X}, \mathcal{X})$.

Proof. Let $\mathsf{X}_+ = \{x \in \mathsf{X} : K(H)(x) > 0\}$ and $\mathsf{X}_0 = \{x \in \mathsf{X} : K(H)(x) = 0\}$ noting $\mathsf{X} = \mathsf{X}_+ \cup \mathsf{X}_0$ and $\mathsf{X}_+ \cap \mathsf{X}_0 = \emptyset$.

First case. If $x \in \mathsf{X}_+$ then $K(H)(x)K^H(\varphi)(x) = K(H \cdot \varphi)(x)$.

Second case. If $x \in \mathsf{X}_0$ we make three observations. Firstly, $K(H)(x)K^H(\varphi)(x) = K(H)(x)K(\varphi)(x)$ by definition. Secondly, $H = 0$ almost surely w.r.t. $K(x, \cdot)$ since H is non-negative. Hence, $K(H \cdot \varphi)(x) = 0$ using the standard measure-theoretic convention $0 \times \infty = 0$ (see for example, Billingsley, 1995, p. 199). Lastly, $K(H)(x)K(\varphi)(x) = 0$ for $K(\varphi)(x) \in [-\infty, \infty]$ by the same convention, since $K(H)(x) = 0$. Therefore, $K(H)(x)K^H(\varphi)(x) = K(H \cdot \varphi)(x) = 0$ for almost every $x \in \mathsf{X}_0$. □

Proposition B.2 (Form of $Q_{p,n}$ with knots)**.** *Suppose $\mathcal{M}^* = \mathcal{K} * \mathcal{M}$ with $Q_{p,n}^*$ terms for $p \in [0\!:\!n]$. For any measure λ_p on $(\mathsf{Y}_p, \mathcal{Y}_p)$, if $\mathcal{K} = (R_{0:n-1}, K_{0:n-1})$ and $\varphi : \mathsf{X}_n \to [-\infty, \infty]$ then $Q_{p,n}^*(\varphi) = K_p Q_{p,n}(\varphi)$ λ_p-a.e. for $p \in [0\!:\!n-1]$ and $Q_{n,n}^*(\varphi) = Q_{n,n}(\varphi)$.*

Proof. Let R_p a probability measure on $(\mathsf{Y}_p, \mathcal{Y}_p)$. Starting with Q_p^* terms under $\mathcal{M}^*$, for $p \in [1\!:\!n-1]$ and $\varphi_p : \mathsf{Y}_p \to [-\infty, \infty]$ and we have

$$\begin{aligned}Q_p^*(\varphi_p) &= K_{p-1}(G_{p-1}) \cdot K_{p-1}^{G_{p-1}} R_p(\varphi_p) \\ &= K_{p-1}\{G_{p-1} \cdot R_p(\varphi_p)\},\end{aligned} \tag{10}$$

λ_{p-1}-a.e. by Proposition B.1. As for Q_n^* we have

$$\begin{aligned} Q_n^*(\varphi) &= K_{n-1}(G_{n-1}) \cdot K_{n-1}^{G_{n-1}} M_n(\varphi) \\ &= K_{n-1}\{G_{n-1} \cdot M_n(\varphi)\} \\ &= K_{n-1} Q_n(\varphi), \end{aligned}$$

λ_{n-1}-a.e. by Proposition B.1. Similarly, it follows that

$$Q_{n-1,n}^*(\varphi) = Q_n^*(\varphi) = K_{n-1} Q_n(\varphi) = K_{n-1} Q_{n-1,n}(\varphi).$$

Now, assume $Q_{p+1,n}^*(\varphi) = K_{p+1} Q_{p+1,n}(\varphi)$ for a given $p \in [0 : n-2]$. Then by (10)

$$\begin{aligned} Q_{p,n}^*(\varphi) &= Q_{p+1}^* Q_{p+1,n}^*(\varphi) \\ &= \int K_p(\cdot, \mathrm{d}x_p) G_p(x_p) R_{p+1}(x_p, \mathrm{d}y_{p+1}) K_{p+1}(y_{p+1}, \mathrm{d}x_{p+1}) Q_{p+1,n}(\varphi)(x_{p+1}) \\ &= \int K_p(\cdot, \mathrm{d}x_p) G_p(x_p) M_{p+1}(x_p, \mathrm{d}x_{p+1}) Q_{p+1,n}(\varphi)(x_{p+1}) \\ &= K_p Q_{p+1} Q_{p+1,n}(\varphi) \\ &= K_p Q_{p,n}(\varphi), \end{aligned}$$

λ_p-a.e. Therefore by induction we have $Q_{p,n}^*(\varphi) = K_p Q_{p,n}(\varphi)$ λ_p-a.e. for $p \in [0 : n-1]$, and $Q_{n,n}^*(\varphi) = Q_{n,n}(\varphi)$ since $Q_{n,n}^* = Q_{n,n} = \mathrm{Id}$ by definition. □

Proposition B.3 (Twisting kernels equivalence)**.** *Let $R : (\mathsf{X}, \mathcal{Y}) \to [0,1]$ and $K : (\mathsf{Y}, \mathcal{Z}) \to [0,1]$ be two Markov kernels for measurable spaces $(\mathsf{Y}, \mathcal{Y})$ and $(\mathsf{Z}, \mathcal{Z})$, and let H be a non-negative real-valued function. If $R^{K(H)}$ and K^H are twisted Markov kernels then $R^{K(H)} K^H = (RK)^H$ λ-a.e. for any measure λ on $(\mathsf{X}, \mathcal{X})$.*

Proof. Note that since $K(H)$ is integrable w.r.t. $R(x, \cdot)$ by definition and $R\{K(H)\} = RK(H)$ then we have H integrable w.r.t. $RK(x, \cdot)$ for $x \in \mathsf{X}$.

By definition of the twisted kernel we have

$$R^{K(H)} K^H(x, \mathrm{d}z) = \begin{cases} \int_{\mathsf{Y}} \frac{R(x, \mathrm{d}y) K(H)(y)}{RK(H)(x)} K^H(y, \mathrm{d}z) & \text{if } RK(H)(x) > 0, \\ RK^H(x, \mathrm{d}z), & \text{if } RK(H)(x) = 0. \end{cases}$$

In the first case,

$$\begin{aligned} \int_{\mathsf{Y}} \frac{R(x, \mathrm{d}y) K(H)(y)}{RK(H)(x)} K^H(y, \mathrm{d}z) &= \int_{\mathsf{Y}} \frac{R(x, \mathrm{d}y) K(y, \mathrm{d}z) H(z)}{RK(H)(x)} \\ &= \frac{RK(x, \mathrm{d}z) H(z)}{RK(H)(x)} \end{aligned}$$

by Proposition B.1. As for the second case, $RK(H)(x) = 0$ implies $K(H) = 0$ almost surely w.r.t. R, hence $RK^H = RK$, by definition of K^H. Thus, $R^{K(H)} K^H = (RK)^H$ λ-a.e. □

Proposition B.4 (Simplification of two t-knots)**.** *Let $\mathcal{K}_1 = (t, R_1, K_1)$ and $\mathcal{K}_2 = (t, R_2, K_2)$ be knots for $t \in [0:n-1]$. If $\mathcal{K}_1$ is compatible with $\mathcal{M} \in \mathscr{M}_n$ and $\mathcal{K}_2$ is compatible with $\mathcal{K}_1 * \mathcal{M}$ then $\mathcal{K}_2 * \mathcal{K}_1 * \mathcal{M} = \mathcal{K}_3 * \mathcal{M}$ where $\mathcal{K}_3 = (t, R_2, K_2 K_1)$.*

Proof. Let $\mathcal{M} = (M_{0:n}, G_{0:n})$. By repeated use of Definition 3.3, the model $\mathcal{K}_2 * \mathcal{K}_1 * \mathcal{M} = (M^{(2)}_{0:n}, G^{(2)}_{0:n})$ has

$$M^{(2)}_t = R_2, \quad M^{(2)}_{t+1} = K_2^{K_1(G_t)} K_1^{G_t} M_{t+1}, \quad G^{(2)}_t = K_2 K_1(G_t),$$

and the remaining kernels and potentials are the same as those in $\mathcal{M}$. By Proposition B.3 we can state $M^{(2)}_{t+1} = (K_2 K_1)^{G_t} M_{t+1}$. Therefore the kernels and potential of $\mathcal{K}_2 * \mathcal{K}_1 * \mathcal{M}$ are equal to that of $\mathcal{K} * \mathcal{M}$ where $\mathcal{K} = (t, R_2, K_2 K_1)$. □

Proposition B.5 (Completion of a knotset)**.** *Suppose $\mathcal{K}$ is a knotset compatible with $\mathcal{M}$ and $\mathcal{K}^\diamond$ is the adapted knotset for $\mathcal{M}$. Then there exists a knotset $\mathcal{R}$ such that $\mathcal{R} * \mathcal{K} * \mathcal{M} = \mathcal{K}^\diamond * \mathcal{M}$.*

Proof. Consider $\mathcal{K} = (R_{0:n-1}, K_{0:n-1})$ and $\mathcal{M} = (M_{0:n}, G_{0:n})$ and recall $\mathcal{M}^* = \mathcal{K} * \mathcal{M}$ has Markov kernels $M^*_0 = R_0$, $M^*_p = K_{p-1}^{G_{p-1}} R_p$ for $p \in [1:n-1]$, and $M^*_n = K_{n-1}^{G_{n-1}} M_n$, whilst the potentials are $G^*_p = K_p(G_p)$ for $p \in [0:n-1]$ and $G^*_n = G_n$. Let $\mathcal{R} = (U_{0:n-1}, V_{0:n-1})$ where $U_p = K_{p-1}^{G_{p-1}}$ and $V_p = R_p$ for $p \in [1:n-1]$, whilst $U_0 = \delta_0$ and some V_0 satisfying $V_0(0, \cdot) = R_0$. The model $\mathcal{M}' = \mathcal{R} * \mathcal{M}^* = (M'_{0:n}, G'_{0:n})$ satisfies

$$\begin{aligned} &M'_0 = \delta_0, \quad M'_1(0, \cdot) = V_0^{K_0(G_0)} K_0^{G_0}, \quad M'_n = R_{n-1}^{K_{n-1}(G_{n-1})} K_{n-1}^{G_{n-1}} M_n, \\ &M'_p = R_{p-1}^{K_{p-1}(G_{p-1})} K_{p-1}^{G_{p-1}} \text{ for } p \in [2:n-1]. \end{aligned}$$

By Proposition B.3 we can state $V_0^{K_0(G_0)} K_0^{G_0} = (V_0 K_0)^{G_0}$ and $R_p^{K_p(G_p)} K_p^{G_p} = (R_p K_p)^{G_p}$ for $p \in [1:n-1]$. Hence, by compatibility

$$\begin{aligned} &M'_0 = \delta_0, \quad M'_1(0, \cdot) = M_0^{G_0}, \quad M'_n = M_{n-1}^{G_{n-1}} M_n, \\ &M'_p = M_{p-1}^{G_{p-1}} \text{ for } p \in [2:n-1]. \end{aligned}$$

Whilst the potential functions satisfy $G'_p = R_p K_p(G_p) = M_p(G_p)$ for $p \in [0:n-1]$ and $G'_n = G_n$. Hence, $\mathcal{M}' = \mathcal{K}^\diamond * \mathcal{M}$ which completes the proof. □

Proposition B.6 (Adapted knotset equivalence)**.** *Let $\mathcal{K}$ be a knotset and suppose $\mathcal{K}^\diamond$ is the adapted knotset for $\mathcal{M}$ and $\mathcal{K}^\diamond_*$ is the adapted knotset for $\mathcal{M}^* = \mathcal{K} * \mathcal{M}$. Then there exists a knotset $\mathcal{K}'$ such that $\mathcal{K}^\diamond_* * \mathcal{K} * \mathcal{M} = \mathcal{K}' * \mathcal{K}^\diamond * \mathcal{M}$.*

Proof. Consider $\mathcal{K} = (R_{0:n-1}, K_{0:n-1})$, $\mathcal{M} = (M_{0:n}, G_{0:n})$, and let $\mathcal{M}_1 = \mathcal{K}^\diamond_* * \mathcal{M}^* = \mathcal{K}^\diamond_* * \mathcal{K} * \mathcal{M}$. The model $\mathcal{M}^* = (M^*_{0:n}, G^*_{0:n})$ follows Proposition 3.9 and since $\mathcal{K}^\diamond_*$ is the adapted knotset for $\mathcal{M}^*$, we can state $\mathcal{M}_1 = (M_{1,0:n}, G_{1,0:n})$ where the kernels are $M_{1,0} = \delta_0$, $M_{1,p} = (M^*_{p-1})^{G^*_{p-1}}$ for $p \in [1:n-1]$, and $M_{1,n} = (M^*_{n-1})^{G^*_{n-1}} M_n$, whilst the potentials are

$G_{1,p} = M_p^*(G_p^*)$ for $p \in [0:n-1]$ and $G_{1,n} = G_n$. Noting that $R_pK_p = M_p$, $(M_0^*)^{G_0^*} = R_0^{K_0(G_0)}$, and $(M_p^*)^{G_p^*} = (K_{p-1}^{G_{p-1}}R_p)^{K_p(G_p)} = (K_{p-1}^{G_{p-1}})^{M_p(G_p)}R_p^{K_p(G_p)}$ for $p \in [1:n-1]$ by Proposition B.3 we can state

$$\begin{aligned}
M_{1,0} &= \delta_0, & G_{1,0} &= M_0(G_0),\\
M_{1,1}(0,\cdot) &= R_0^{K_0(G_0)} & G_{1,1} &= K_0^{G_0}M_1(G_1),\\
M_{1,p} &= (K_{p-2}^{G_{p-2}})^{M_{p-1}(G_{p-1})}R_{p-1}^{K_{p-1}(G_{p-1})} & G_{1,p} &= K_{p-1}^{G_{p-1}}M_p(G_p), \quad p \in [2:n-1]\\
M_{1,n} &= (K_{n-2}^{G_{n-2}})^{M_{n-1}(G_{n-1})}R_{n-1}^{K_{n-1}(G_{n-1})}M_n, & G_{1,n} &= G_n.
\end{aligned}$$

Next, consider $\mathcal{M}^\diamond = \mathcal{K}^\diamond * \mathcal{M} = (M_{0:n}^\diamond, G_{0:n}^\diamond)$, as in Example 3.11, and note that it can be rewritten as

$$\begin{aligned}
M_0^\diamond &= \delta_0, & G_0^\diamond &= M_0(G_0),\\
M_1^\diamond(0,\cdot) &= R_0^{K_0(G_0)}K_0^{G_0}, & G_p^\diamond &= M_p(G_p),\\
M_p^\diamond &= R_{p-1}^{K_{p-1}(G_{p-1})}K_{p-1}^{G_{p-1}}, & G_p^\diamond &= M_p(G_p), \quad p \in [2:n-1]\\
M_n^\diamond &= R_{n-1}^{K_{n-1}(G_{n-1})}K_{n-1}^{G_{n-1}}M_n, & G_n^\diamond &= G_n
\end{aligned}$$

as $M_p^{G_p} = R_p^{K_p(G_p)}K_P^{G_p}$ by Proposition B.3 and since $R_pK_p = M_p$. Finally, let $\mathcal{K}' = (R'_{0:n-1}, K'_{0:n-1})$ and $\mathcal{M}_2 = \mathcal{K}' * \mathcal{M}^\diamond$ where $R'_0 = \delta_0$ and $K'_0 = \mathrm{Id}$, $R'_p = R_{p-1}^{K_{p-1}(G_{p-1})}$ and $K'_p = K_{p-1}^{G_{p-1}}$ for $p \in [1:n-1]$. Therefore from Proposition 3.9, $\mathcal{M}_2 = (M_{2,0:n}, G_{2,0:n})$ satisfies $M_{2,p} = M_{1,p}$ and $G_{2,p} = G_{1,p}$ for $p \in [0:n]$ and hence $\mathcal{M}_2 = \mathcal{M}_1$ completing the proof. □

Proposition B.7 (Repeated adapted knotset equivalence)**.** *Let $t \in [2:\infty)$ and suppose $\mathcal{K}_s^\diamond$ is the adapted knotset for $\mathcal{M}_s$ for $s \in \{1,t\}$ and $\mathcal{K}_s$ for $s \in [1:t-1]$ are knotsets such that $\mathcal{M}_t = \mathcal{K}_{t-1} * \cdots * \mathcal{K}_1 * \mathcal{M}_1$. Then there exists knotsets $\mathcal{J}_s$ for $s \in [2:t]$ such that*

$$\mathcal{K}_t^\diamond * \mathcal{M}_t = \mathcal{K}_t^\diamond * \mathcal{K}_{t-1} * \cdots * \mathcal{K}_1 * \mathcal{M}_1 = \mathcal{J}_t * \cdots * \mathcal{J}_2 * \mathcal{K}_1^\diamond * \mathcal{M}_1.$$

Proof. The proof follows from repeated applications of Proposition B.6. □

Proposition B.8 (Terminal knot kernel equivalence)**.** *Consider model $\mathcal{M} = (M_{0:n}, G_{0:n})$ where $M_n = P_1 \otimes P_2$ and $G_n = H \otimes \phi^{-1}$, and terminal knot $\mathcal{K} = (n, R, K)$ and let $\mathcal{K} * \mathcal{M} = (M_{0:n}^*, G_{0:n}^*)$. The terminal kernels and potential functions satisfy*

$$M_n^*(G_n^* \cdot [1 \otimes \varphi]) = M_n(G_n \cdot [1 \otimes \varphi]).$$

Proof. We have,

$$\begin{aligned}
M_n^*(G_n^* \cdot [1 \otimes \varphi]) =& (R \otimes K^H P_2)([K(H) \otimes \phi^{-1}] \cdot [1 \otimes \varphi])\\
=& (R \otimes K^H P_2)([K(H) \otimes [\phi^{-1} \cdot \varphi])\\
=& R\{K(H) \cdot K^H P_2(\phi^{-1} \cdot \varphi)\}\\
=& R\{K[H \cdot P_2(\phi^{-1} \cdot \varphi)]\},
\end{aligned}$$

where the last equality follows from Proposition B.1. Finally, using Definition 6.3 compatibility condition (ii) we have

$$\begin{aligned} M_n^*(G_n^* \cdot [1 \otimes \varphi]) =& RK\{H \cdot P_2(\phi^{-1} \cdot \varphi)\} \\ &= (P_1 \otimes P_2)(H \otimes [\phi^{-1} \cdot \varphi]) \\ &= (P_1 \otimes P_2)(H \otimes [\phi^{-1} \cdot \varphi]) \\ &= M_n(G_n \cdot [1 \otimes \varphi]). \end{aligned}$$

□

SUPPLEMENTARY MATERIAL

Code to reproduce the experiments and visualisations in this paper is available online: `https://github.com/bonStats/KnotsNonLinearSSM.jl`. We gladly acknowledge the `tidyverse` (Wickham et al., 2019), `matrixStats` (Bengtsson, 2025), and `patchwork` (Pedersen, 2025) packages in programming language `R` (R Core Team, 2025). As well as the `Julia` language (Bezanson et al., 2017) and packages `DataFrames.jl` (Bouchet-Valat and Kamiński, 2023) and `Distributions.jl` (Besançon et al., 2021; Lin et al., 2019).