Random processes and random fields

Let $(\Omega, \mathcal{B}, P)$ be a probability space. A real valued stochastic process in the time interval $[0, T]$ is a mapping

$$X : \Omega \times [0, T] \to \mathbb{R}. \tag{1}$$

The process can be continuous in time[1] (e.g., the Wiener process), discontinuous in time (e.g., telegrapher's random process), or time-discrete, e.g., represented by a sequence of random variables $X(t_j; \omega)$ $j = 1, \ldots, n$. In this sense, a random process can be thought of as an infinite set (countable or uncountable) of random variables indexed by the time t.

The properties of $X(t; \omega)$ very much depend on the way we characterize the process, i.e., the set of rules and specifications that allow us to fully characterize the process. Clearly, $X(t; \omega)$ is a random variable for each fixed t. This means that $X(t; \omega)$ admits a distribution function

$$F(x, t) = P\left(\{\omega : X(t; \omega) \leq x\}\right), \tag{4}$$

and eventually a probability density function

$$p(x, t) = \frac{dF(x, t)}{dx}. \tag{5}$$

With $F(x, t)$ or $p(x, t)$ available we can compute the statistical moments at time t, e.g.,

$$\mathbb{E}\left\{X(t; \omega)^k\right\} = \int_{-\infty}^{\infty} x^k p(x, t) dx, \quad k \in \mathbb{N}. \tag{6}$$

However, the PDF $p(x, t)$ provides very limited statistical information about the process $X(t; \omega)$. In fact, it does not allow us to compute any joint statistics at different times, for example the auto-correlation function

$$\mathbb{E}\{X(t; \omega) X(s; \omega)\} = \int_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2, t, s) dx_1 dx_2, \tag{7}$$

[1] The notion of continuity for real-valued functions can be generalized big time to stochastic processes. For example,

Continuity in probability:	$\lim_{s \to t} P(\{\omega : \lvert X(t; \omega) - X(s; \omega)\rvert > \epsilon\}) = 0$	for all $\epsilon > 0$.
Continuity with probability one:	$P\left(\{\omega : \lim_{s \to t} \lvert X(s; \omega) - X(t; \omega)\rvert = 0\}\right) = 1$	for all $t \geq 0$.
Mean-square continuity:	$\lim_{s \to t} \mathbb{E}\{\lvert X(t; \omega) - X(s; \omega)\rvert^2\} = 0.$	
Continuity in distribution:	$\lim_{s \to t} F(x, s) = F(x, t)$	($F(x, t)$ distribution function of $X(t; \omega)$).

In general, it can be shown that:

Continuity in mean-square $\Rightarrow$ continuity in probability $\Rightarrow$ continuity in distribution.

Continuity in probability follows from mean-square continuity thanks to the Markov's inequality

$$P(\{\omega : |X(t; \omega) - X(s; \omega)| > \epsilon\}) \leq \frac{1}{\epsilon^2} \mathbb{E}\{|X(t; \omega) - X(s; \omega)|^2\} \qquad \forall t, s \in [0, T]. \tag{2}$$

Note: there are pathwise continuous stochastic processes that are not mean-square continuous. On the other hand Mean-square continuity implies that the mean process $\mathbb{E}\{X(t; \omega)\}$ is continuous in t (but not necessarily pathwise continuous). In fact, using the inequality $|\mathbb{E}\{X\}|^2 \leq \mathbb{E}\{X^2\}$ we obtain

$$|\mathbb{E}\{X(t; \omega) - X(s; \omega)\}| = |\mathbb{E}\{X(t; \omega)\} - \mathbb{E}\{X(s; \omega)\}| \leq \sqrt{\mathbb{E}\{|X(t; \omega) - X(s; \omega)|^2\}}. \tag{3}$$

Similarly, if the process is mean-square continuous then the auto-correlation function $\mathbb{E}\{X(t; \omega) X(s; \omega)\}$ continuous in both s and t.

where $p(x_1, x_2, t, s)$ is the joint probability density function of the random variables $X(t;\omega)$ and $X(s;\omega)$, where t and s can vary in $[0, T]$.

A straightforward generalization of what we just said leads us to construct the joint PDF of a vector $\{X(t_1;\omega), \ldots, X(t_n;\omega)\}$ representing the values of the process $X(t;\omega)$ at an arbitrary number of time instants $t_i \in [0, T]$. Similarly, we can construct the joint characteristic function of the random process $X(t;\omega)$ at $\{t_1, \ldots, t_n\}$ as

$$\phi(a_1, \ldots, a_n; t_1, \ldots, t_n) = \mathbb{E}\left\{e^{ia_1X(t_1;\omega)+\cdots+ia_nX(t_n;\omega)}\right\}. \tag{8}$$

This expression can be obtained (at least formally) from the so-called Hopf characteristic functional [9, 2] associated with the stochastic process $X(t;\omega)$, i.e.,

$$\Phi([\theta(t)]) = \mathbb{E}\left\{\exp\left(i\int_0^T X(\tau;\omega)\theta(\tau)d\tau\right)\right\}, \tag{9}$$

where $\theta(t)$ is a deterministic test function which we are free to choose. For example, if we pick

$$\theta(t) = \sum_{i=1}^{n} a_i\delta(t - t_i), \tag{10}$$

and substitute it into (9) then we obtain (8). The Hopf functional[2] (9) provides full statistical information about the stochastic process $X(t;\omega)$, including all joint statistical moments, all multi-time PDFs, etc. For instance, the functional derivatives of Φ evaluated at $\theta = 0$ coincide with the statistical moments (see, e.g. [7])

$$\left.\frac{\delta^q\Phi([\theta])}{\delta\theta(t)^q}\right|_{\theta=0} = \frac{1}{i^q}\mathbb{E}\{X(t;\omega)^q\}, \qquad \left.\frac{\delta^{q+p}\Phi([\theta])}{\delta\theta(t)^q\delta\theta(s)^p}\right|_{\theta=0} = \frac{1}{i^{q+p}}\mathbb{E}\{X(t;\omega)^qX(s;\omega)^p\}. \tag{11}$$

In [2] the Hopf functional is determined for various types of stochastic processes.

Remark: To fully characterize a stochastic process it is not necessary to identify or provide the Hopf functional. A stochastic process can be defined in many different ways, some of which are not even explicit. However, if the Hopf characteristic functional is available, then the process is fully specified, perhaps in the most compact possible way (see [5] for applications of Hopf functional methods to turbulence).

Gaussian processes. The Hopf characteristic functional of a Gaussian process is (see, e.g., [2])

$$\Phi([\theta(t)]) = \exp\left(i\int_0^T \mu(\tau)\theta(\tau) - \frac{1}{2}\int_0^T\int_0^T C(\tau, s)\theta(\tau)\theta(s)d\tau ds\right), \tag{12}$$

where

$$\mu(t) = \mathbb{E}\{X(t;\omega)\} \qquad \text{(mean)}, \tag{13}$$

$$C(t, s) = \mathbb{E}\{X(t;\omega)X(s;\omega)\} - \mu(t)\mu(s) \qquad \text{(covariance function)}. \tag{14}$$

Higher order moments can be computed using functional differentiation (e.g., (11)), or by noticing that the joint characteristic function the random process $X(t;\omega)$ at an arbitrary number of distinct time instants is

$$\phi(a_1, \ldots, a_n; t_1, \ldots, t_n) = \exp\left(i\sum_{k=1}^{n} a_k\mu(t_k) - \frac{1}{2}\sum_{k,j=1}^{n} C(t_k, t_j)a_ka_j\right). \tag{15}$$

[2]Recall that a functional is a mapping from a certain space of functions (or distributions) into the real line or the complex plane. The Hopf functional is a complex-valued nonlinear functional into $\mathbb{C}$.

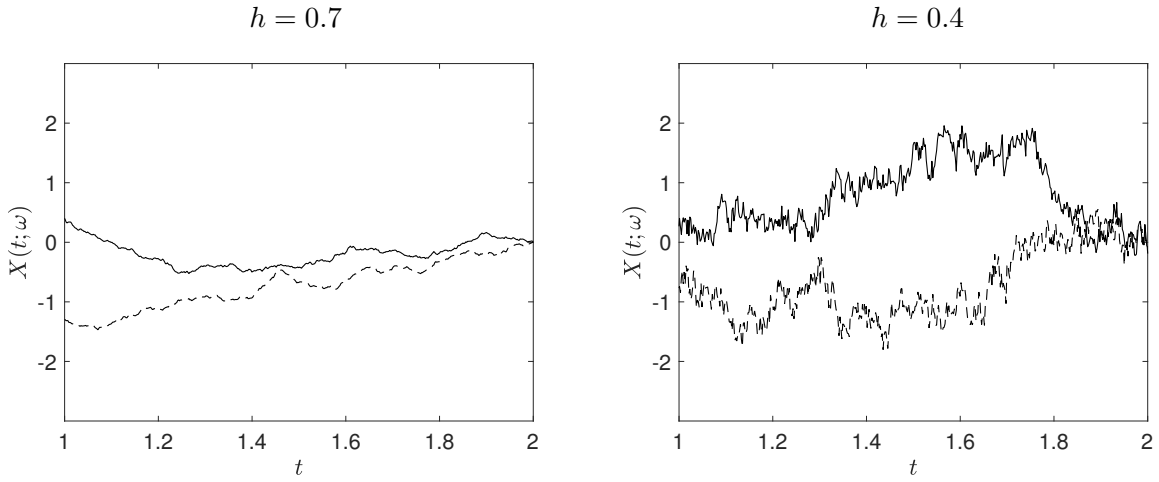

Figure 1: Samples of zero-mean Gaussian process with covariance function (17) and $\sigma = 1$. We show samples corresponding to different values of the Hurst parameter h.

Sampling Gaussian processes. To sample a Gaussian process with mean $\mu(t)$ and covariance function $C(s,t)$ it is sufficient to construct a temporal grid in $[0,T]$ and then sample a Gaussian random vector with mean $\mu_i = \mu(t_i)$, and covariance matrix with entries $C(t_k,t_j)$. To this end, recall that if $\boldsymbol{X}(\omega)$ is a zero-mean Gaussian random vector (column vector) with independent entries of variance one, and $\boldsymbol{C} = \boldsymbol{R}\boldsymbol{R}^T$ is the Cholesky decomposition of the covariance matrix[3] $\boldsymbol{C}$, then $\boldsymbol{Y} = \boldsymbol{R}\boldsymbol{X}$ is a zero-mean Gaussian random vector with covariance $\boldsymbol{C}$. In fact,

$$\mathbb{E}\{\boldsymbol{Y}(\omega)\boldsymbol{Y}^T(\omega)\} = \mathbb{E}\{\boldsymbol{R}\boldsymbol{X}(\omega)\boldsymbol{X}^T(\omega)\boldsymbol{R}^T\} = \boldsymbol{R}\underbrace{\mathbb{E}\{\boldsymbol{X}(\omega)\boldsymbol{X}^T(\omega)\}}_{\text{(identity matrix)}}\boldsymbol{R}^T = \boldsymbol{C}. \tag{16}$$

In Figure 1 we plot a few samples of a Gaussian random process with zero mean and covariance function

$$C(s,t) = \frac{\sigma}{2}\left(|s|^{2h} + |t|^{2h} - |s-t|^{2h}\right), \tag{17}$$

where $0 < h < 1$ is the so-called Hurst parameter. A Gaussian process with covariance function (17) is called fractional Brownian motion.

Gaussian random fields. The procedure we used to sample of Gaussian stochastic process with covariance $C(s,t)$ (e.g., (17)) can be extended to *Gaussian random fields* [6], i.e., random functions defined of a domain $V \subseteq \mathbb{R}^d$. For example, we can sample a zero-mean Gaussian random field $X(\boldsymbol{x};\omega)$ defined on the square domain $V = [0,1]\times[0,1]$ with covariance function

$$C(\boldsymbol{x},\boldsymbol{y}) = \frac{\sigma}{2}\left(\|\boldsymbol{x}\|_2^{2h} + \|\boldsymbol{y}\|_2^{2h} - \|\boldsymbol{x}-\boldsymbol{y}\|_2^{2h}\right). \tag{18}$$

To this end we, first construct the covariance matrix $C(\boldsymbol{x}_i,\boldsymbol{x}_j)$ and then use the procedure we just described before, i.e.

1. sample a zero-mean i.i.d. Gaussian random variable with variance one at each spatial location $\boldsymbol{x}_i$,
2. multiply the sample of the random vector constructed in this way by the matrix $\boldsymbol{R}$ obtained by the Cholesky decomposition of the covariance function (18).

In Figure 2 we provide a few samples of a zero mean Gaussian random field with covariance (18).

[3]The entries of the covariance matrix $\boldsymbol{C}$ are $C(t_i,t_j)$, where $C(t,s)$ is the covariance function of the random process.

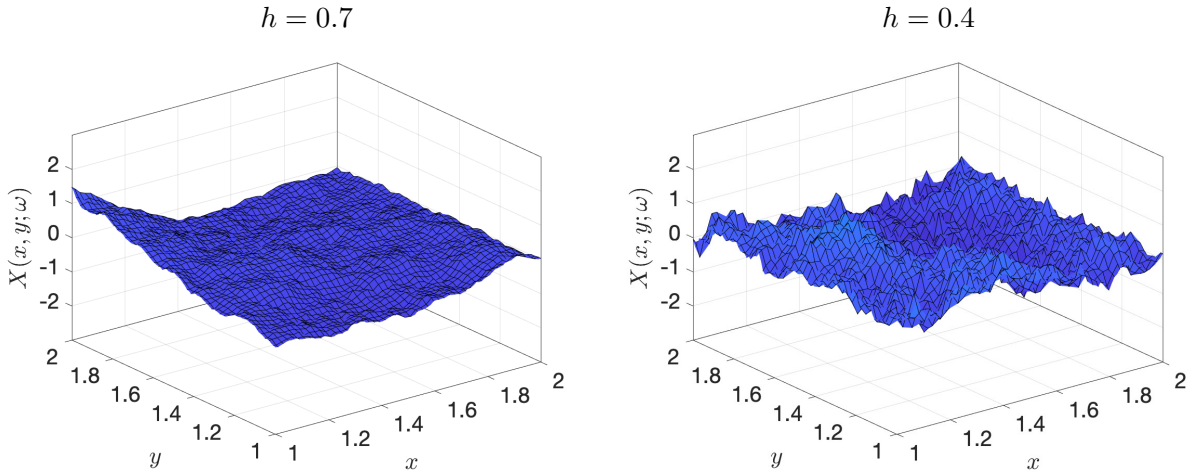

Figure 2: Samples of zero-mean Gaussian random field with covariance function (18) and $\sigma = 1$. We show samples corresponding to different values of the Hurst parameter h.

Discrete Markov processes

Consider a vector-valued discrete random process, i.e., a sequence of random vectors

$$\boldsymbol{X}_0(\omega) \mapsto \boldsymbol{X}_1(\omega) \mapsto \boldsymbol{X}_2(\omega) \mapsto \cdots \tag{19}$$

defined by some law that allows us to generate new elements of the sequence (hence the name "process"). To define the random process (19) we need to specify how a particular state $\boldsymbol{X}_n$ can be computed given past states. This is modeled effectively by the transition density of $\boldsymbol{X}_n$ given previous states. Alternatively, we can write down a deterministic evolution equation for $\boldsymbol{X}_j$, i.e., an equation that relates $\boldsymbol{X}_n$ to previous states.

Nonlinear autoregressive processes. The random process (19) is often defined via recurrence relations[4], e.g., of the form[5]

$$\boldsymbol{X}_{k+1}(\omega) = \boldsymbol{F}(\boldsymbol{X}_k(\omega)) + \boldsymbol{\xi}_k(\omega), \tag{21}$$

where $\boldsymbol{X}_0$ is a random vector with known PDF, and $\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots\}$ is a sequence of random vectors with known joint PDFs. The index "k" usually refers to time , but is can also refer to other quantities, e.g., the layer identity in a feed-forward neural network. Indeed, the recurrence relation (21) can model a recurrent neural network where the output of each layer is perturbed by some noise $\boldsymbol{\xi}_k$, and the components of $\boldsymbol{F}$ represent the neural activation functions (see [10]).

Markov property. Disregarding how we generate the sequence of random vectors $\{\boldsymbol{X}_j\}$ in (19), we can characterize the statistics of the process X_i in terms of the joint PDF of the random vectors $\{\boldsymbol{X}_0, \ldots, \boldsymbol{X}_n\}$, $p(\boldsymbol{x}_n, \ldots, \boldsymbol{x}_0)$. By using the definition of conditional probability density we have

$$p(\boldsymbol{x}_n, \ldots, \boldsymbol{x}_1, \boldsymbol{x}_0) = p(\boldsymbol{x}_n | \boldsymbol{x}_{n-1}, \ldots, \boldsymbol{x}_1, \boldsymbol{x}_0) p(\boldsymbol{x}_{n-1}, \ldots, \boldsymbol{x}_0). \tag{22}$$

[4]The discrete process defined by (21) is also called *nonlinear autoregressive process* [1].

[5]A discrete nonlinear autoregressive process that generalizes (21) is

$$\boldsymbol{X}_{k+1} = \boldsymbol{F}_k\left(\boldsymbol{X}_k, \boldsymbol{\xi}_k\right), \tag{20}$$

If the system is *memoryless* (or Markovian), we have that the conditional PDF of $\boldsymbol{X}_n$ given the entire history of $\boldsymbol{X}_i$ from $\boldsymbol{X}_0$ to $\boldsymbol{X}_{n-1}$ equals $p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})$, i.e.,

$$p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1},\ldots,\boldsymbol{x}_1,\boldsymbol{x}_0) = p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1}) \qquad \text{(Markov property)}. \tag{23}$$

In other words, the PDF of the random vector $\boldsymbol{X}_n(\omega)$ conditional to any set of "previous" vectors $\{\boldsymbol{X}_j(\omega)\}$ with $j < n$ equals to $p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})$, i.e., it depends only on $\boldsymbol{X}_{n-1}(\omega)$. By applying (23) recursively we obtain

$$p(\boldsymbol{x}_n,\boldsymbol{x}_{n-1},\ldots,\boldsymbol{x}_1,\boldsymbol{x}_0) = p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})p(\boldsymbol{x}_{n-1}|\boldsymbol{x}_{n-2})\cdots p(\boldsymbol{x}_1|\boldsymbol{x}_0)p(\boldsymbol{x}_0). \tag{24}$$

Hence, to fully specify a Markov process we just need to define

- The transition densities $p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)$,
- The PDF $p(\boldsymbol{x}_0)$ of the initial state $\boldsymbol{X}_0$.

Note also that by marginalizing (24) with respect to $(\boldsymbol{x}_0,\ldots,\boldsymbol{x}_{n-1})$ we obtain

$$p(\boldsymbol{x}_n) = \int p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})p(\boldsymbol{x}_{n-1})d\boldsymbol{x}_{n-1}. \tag{25}$$

The transition densities $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ depend on way the process is defined.

Nonlinear autoregressive Markov processes. Consider a slightly generalized version of the random process defined in (21), where $\boldsymbol{F}$ can be dependent of the particular iteration

$$\boldsymbol{X}_{k+1}(\omega) = \boldsymbol{F}_k(\boldsymbol{X}_k(\omega)) + \boldsymbol{\xi}_k(\omega), \tag{26}$$

In Appendix A we show that if the perturbation vectors $\{\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_n\}$ are mutually independent, and if each $\boldsymbol{\xi}_j$ is independent of past and current states, i.e., $\{\boldsymbol{X}_0,\ldots,\boldsymbol{X}_n\}$, then (26) defines a Markov process. The transition density of such process is given by

$$\begin{aligned} p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) &= \int p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k,\boldsymbol{\xi}_k)p_{\boldsymbol{\xi}_k}(\boldsymbol{\xi}_k)d\boldsymbol{\xi}_k \\ &= \int \delta(\boldsymbol{x}_{k+1}-\boldsymbol{F}_k(\boldsymbol{x}_k)-\boldsymbol{\xi}_k)p_{\boldsymbol{\xi}_k}(\boldsymbol{\xi}_k)d\boldsymbol{\xi}_k \\ &= p_{\boldsymbol{\xi}_k}(\boldsymbol{x}_{k+1}-\boldsymbol{F}_k(\boldsymbol{x}_k)), \end{aligned} \tag{27}$$

where $p_{\boldsymbol{\xi}_k}$ is the PDF of the noise vector $\boldsymbol{\xi}_k$. Note that if the noise vectors $\boldsymbol{\xi}_k$ are zero, i.e., if the autoregressive model is of the form

$$\boldsymbol{X}_{k+1} = \boldsymbol{F}_k(\boldsymbol{X}_k) \qquad k = 0,1,\ldots \tag{28}$$

with $\boldsymbol{X}_0$ random, then the transition density reduces to

$$p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) = \delta(\boldsymbol{x}_{k+1}-\boldsymbol{F}_k(\boldsymbol{x}_k)). \tag{29}$$

A substitution of (29) into (25) yields

$$p(\boldsymbol{x}_{k+1}) = \int \delta(\boldsymbol{x}_{k+1}-\boldsymbol{F}_k(\boldsymbol{x}_k))p(\boldsymbol{x}_k)d\boldsymbol{x}_k, \tag{30}$$

which is the expression we have seen when studying random vectors to transform the PDF of $\boldsymbol{X}_k$ into the of $\boldsymbol{X}_{k+1}$ via the nonlinear mapping (28).

Remark: Any discrete autoregressive process (state+noise) can be written as a Markovian process in a higher-dimensional state space. For example, consider

$$\boldsymbol{X}_{k+1} = \boldsymbol{F}(\boldsymbol{X}_k, \boldsymbol{X}_{k-1}) + \boldsymbol{\xi}_k, \tag{31}$$

$$\boldsymbol{\xi}_{k+1} = \sin(\boldsymbol{\xi}_k) + \boldsymbol{\eta}_k, \tag{32}$$

where $\boldsymbol{\eta}_k$ are i.i.d. and independent of $\{\boldsymbol{\xi}_k\}$ and $\{\boldsymbol{X}_k\}$. Clearly the noise $\boldsymbol{\xi}_k$ here is a Markov process (not an i.i.d. process), and it is driving the autoregressive process $\{\boldsymbol{X}_k\}$. Upon definition of $\boldsymbol{Z}_k = [\boldsymbol{X}_k,\, \boldsymbol{X}_{k-1},\, \boldsymbol{x}_k]^T$, it is possible to write (31)-(32) as a Markov process driven by the i.i.d. noise $\{\boldsymbol{\eta}_k\}$.

Statistically steady states. Consider the Markovian auto-regressive process (26), where $\boldsymbol{F}_k$ is independent of k, i.e.,

$$\boldsymbol{X}_{k+1} = \boldsymbol{F}(\boldsymbol{X}_k) + \boldsymbol{\xi}_k, \tag{33}$$

and the noise vectors $\{\boldsymbol{\xi}_k\}$ are independent and identically distributed. In these hypotheses the transition density $p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)$ has a functional form that is independent of k. Denote such functional form as $p(\boldsymbol{x}|\boldsymbol{y})$. A *statistically stationary state* $p^*(\boldsymbol{x})$ of the auto regressive Markov process (33) is a solution to the Volterra integral equation

$$p^*(\boldsymbol{x}) = \int p(\boldsymbol{x}|\boldsymbol{y}) p^*(\boldsymbol{y}) d\boldsymbol{y}. \tag{34}$$

Pseudo-random number generators. Random variables and other random objects are simulated by deterministic computer algorithms. The purpose of these algorithms is to produce sequences of numbers whose behavior is very hard to distinguish from that of their "truly random" counterparts, at least for the application of interest. Most algorithms commonly used to generate random numbers in a computer utilize autoregressive relations of the form (28). For example the function `rand()` to generate uniformly distributed random numbers in Matlab and other programming languages implements the so called "MRG32k3a sequence" originally proposed by L'Ecuyer in [3]. The sequence can be written as

$$X_k = (1403580 X_{k-2} - 810728 X_{k-3}) \mod m_1 \tag{35}$$

$$Y_k = (527612 Y_{k-1} - 1370589 Y_{k-3}) \mod m_2 \tag{36}$$

where $m_1 = 2^{32} - 209$, and $m_2 = 2^{32} - 22853$. The uniform "random number" U_k is then obtained as

$$U_k = \frac{(\widehat{X}_k - \widehat{Y}_k) \mod m_1}{m_1 + 1}. \tag{37}$$

where

$$\widehat{X}_k = \begin{cases} X_k & \text{if} \quad X_k \geq 0 \\ X_k - m_1 X_k & \text{if} \quad X_k < 0 \end{cases} \qquad \widehat{Y}_k = \begin{cases} Y_k & \text{if} \quad Y_k \geq 0 \\ Y_k - m_2 Y_k & \text{if} \quad Y_k < 0 \end{cases} \tag{38}$$

The period of this sequence (i.e., after how many iterations it repeats itself) is quite large, i.e., approximately 3.1×10^{57}. The sequence is usually initialized via a "seed" (hereafter set as 12345 for all components)

$$X_{-3} = X_{-2} = X_{-1} = Y_{-3} = Y_{-2} = Y_{-1} = 12345. \tag{39}$$

The seed must be positive and less than m_1 and m_2.

Particle filtering and recursive Bayesian state estimation

Consider a model described by a nonlinear autoregressive Markov process of the form[6]

$$\boldsymbol{X}_{k+1} = \boldsymbol{F}(\boldsymbol{X}_k) + \boldsymbol{\xi}_k, \qquad \text{(model equations)}, \tag{40}$$

[6]The model can be easily generalized the "non-autonomous" case by letting $\boldsymbol{F}$ depend on the particular iteration k, i.e. $\boldsymbol{X}_{k+1} = \boldsymbol{F}_k(\boldsymbol{X}_k)$, or even by including the noise vector $\boldsymbol{\xi}_k$ as an argument of $\boldsymbol{F}_k$, i.e., $\boldsymbol{X}_{k+1} = \boldsymbol{F}_k(\boldsymbol{X}_k, \boldsymbol{\xi}_k)$.

where $\{\boldsymbol{\xi}_j\}$ are i.i.d. random vectors independent of past and present states $\{\boldsymbol{X}_0,\ldots,\boldsymbol{X}_k\}$. Suppose that we are also given (noisy) measurements of a nonlinear function of the states $\{\boldsymbol{X}_k\}$

$$\boldsymbol{Y}_{k+1} = \boldsymbol{H}(\boldsymbol{X}_{k+1}) + \boldsymbol{\eta}_{k+1}, \qquad \text{(measurements)}, \tag{41}$$

where i.i.d. random vectors. The goal of particle filtering and recursive Bayesian state estimation is to construct the PDF of the current state $\boldsymbol{X}_k$, given all the available information, i.e., the discrete system (40) and the measured data $\{\boldsymbol{Y}_0,\ldots,\boldsymbol{Y}_k\}$ defined in (41). In principle, this PDF may be obtained recursively by computing a one-step PDF prediction and then update such PDF with the measurement

$$\text{Prediction} \quad \Rightarrow \quad \text{Update} \quad \Rightarrow \quad \text{Prediction} \quad \Rightarrow \quad \ldots$$

To show how, suppose we are given the PDF of the initial state $\boldsymbol{X}_0$, i.e. $p(\boldsymbol{x}_0)$. By using the Markovian property of (40) we can write the PDF of $\boldsymbol{X}_1$ as

$$p(\boldsymbol{x}_1) = \int p(\boldsymbol{x}_1|\boldsymbol{x}_0)p(\boldsymbol{x}_0)d\boldsymbol{x}_0 \qquad \text{(prediction 1)}. \tag{42}$$

The transition density $p(\boldsymbol{x}_1|\boldsymbol{x}_0)$ can be easily derived from (40) (see, e.g., Eq. (27)) as

$$p(\boldsymbol{x}_1|\boldsymbol{x}_0) = p_{\boldsymbol{\xi}_0}\left(\boldsymbol{x}_1 - \boldsymbol{F}(\boldsymbol{x}_0)\right). \tag{43}$$

At this point suppose that a measurement of $\boldsymbol{Y}_1$ comes available. By using the Bayes's rule we have

$$p(\boldsymbol{x}_1|\boldsymbol{y}_1) = \frac{p(\boldsymbol{y}_1|\boldsymbol{x}_1)p(\boldsymbol{x}_1)}{p(\boldsymbol{y}_1)} = \frac{p(\boldsymbol{y}_1|\boldsymbol{x}_1)p(\boldsymbol{x}_1)}{\displaystyle\int p(\boldsymbol{y}_1|\boldsymbol{x}_1)p(\boldsymbol{x}_1)d\boldsymbol{x}_1} \qquad \text{(update 1)}. \tag{44}$$

The transition density $p(\boldsymbol{y}_1|\boldsymbol{x}_1)$ is easily obtained from (41) as

$$p(\boldsymbol{y}_1|\boldsymbol{x}_1) = p_{\boldsymbol{\eta}_1}\left(\boldsymbol{y}_1 - \boldsymbol{H}(\boldsymbol{x}_1)\right). \tag{45}$$

With this expression, we can compute the PDF $p(\boldsymbol{x}_1|\boldsymbol{y}_1)$ in (44). Next, we push forward the updated PDF $p(\boldsymbol{x}_1|\boldsymbol{y}_1)$ using the transition density $p(\boldsymbol{x}_2|\boldsymbol{x}_1)$ defined by our model equations (40), i.e.,

$$p(\boldsymbol{x}_2|\boldsymbol{y}_1) = \int p(\boldsymbol{x}_2|\boldsymbol{x}_1)p(\boldsymbol{x}_1|\boldsymbol{y}_1)d\boldsymbol{x}_1 \qquad \text{(prediction 2)}. \tag{46}$$

When a measurement of $\boldsymbol{Y}_2$ comes in we can updated the PDF as $p(\boldsymbol{x}_2|\boldsymbol{y}_1)$ as

$$p(\boldsymbol{x}_2|\boldsymbol{y}_1,\boldsymbol{y}_2) = \frac{p(\boldsymbol{y}_2|\boldsymbol{x}_2)p(\boldsymbol{x}_2|\boldsymbol{y}_1)}{\displaystyle\int p(\boldsymbol{y}_2|\boldsymbol{x}_2)p(\boldsymbol{x}_2|\boldsymbol{y}_1)d\boldsymbol{x}_2} \qquad \text{(update 2)}. \tag{47}$$

In summary, we need to iterate the following two equations

$$p(\boldsymbol{x}_{n+1}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n) = \int p(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n)p(\boldsymbol{x}_n|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n)d\boldsymbol{x}_n \qquad \text{(prediction)}, \tag{48}$$

$$p(\boldsymbol{x}_{n+1}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n,\boldsymbol{y}_{n+1}) = \frac{p(\boldsymbol{y}_{n+1}|\boldsymbol{x}_{n+1})p(\boldsymbol{x}_{n+1}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n)}{\displaystyle\int p(\boldsymbol{y}_{n+1}|\boldsymbol{x}_{n+1})p(\boldsymbol{x}_{n+1}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_n)d\boldsymbol{x}_{n+1}} \qquad \text{(update/filtering)}. \tag{49}$$

The recurrence relations (48)-(49) of constitute the formal solution to the Bayesian recursive estimation problem. Analytic solutions to this problem are only available for a relatively small and restrictive choice of system and measurement models, the most important being the Kalman filter (see hereafter), which assumes that both $\boldsymbol{F}$ and $\boldsymbol{H}$ in (40) and (41) are linear, and that $\{\boldsymbol{\eta}_k\}$ and $\{\boldsymbol{\eta}_k\}$ are zero-mean additive

Gaussian noises with known covariance matrices. This yields a sequence of Gaussian transition densities and Gaussian states. To represent such states, we only need a evolution equations for the mean and the covariance of the predicted and updated Gaussian PDFs. Indeed mean and covariance fully characterize the Gaussian.

Kalman filter. Suppose that both $\boldsymbol{F}$ and $\boldsymbol{H}$ in (40) and (41) are linear, and that $\{\boldsymbol{\eta}_k\}$ and $\{\boldsymbol{\eta}_k\}$ are zero-mean additive Gaussian noises with known covariance matrices. In these assumptions we write the system (40) and (41) as

$$\boldsymbol{X}_{k+1} = \boldsymbol{A}_k \boldsymbol{X}_k + \boldsymbol{\xi}_k \qquad \text{model,} \tag{50}$$

$$\boldsymbol{Y}_{k+1} = \boldsymbol{B}_{k+1} \boldsymbol{X}_{k+1} + \boldsymbol{\eta}_{k+1} \qquad \text{measurements.} \tag{51}$$

Since all PDFs in (48)-(49) are Gaussian, we just need to update the (conditional) mean and the (conditional) covariance of the PDF defined by the prediction and update/filtering steps. In the hypotheses of zero-mean Gaussian noise with known covariances[7] a straightforward calculation shows that

$$\text{prediction step:} \quad \begin{cases} \boldsymbol{\mu}_{k+1} = \boldsymbol{A}_k \boldsymbol{\mu}_k & \text{mean} \\ \boldsymbol{C}_{k+1} = \boldsymbol{A}_k \boldsymbol{C}_k \boldsymbol{A}_k^T + \boldsymbol{\Sigma}_k & \text{covariance} \end{cases} \tag{52}$$

where $\boldsymbol{\mu}_k = \mathbb{E}\{\boldsymbol{X}_k | \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_k\}$, $\boldsymbol{\mu}_{k+1} = \mathbb{E}\{\boldsymbol{X}_{k+1} | \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_k\}$, $\boldsymbol{C}_k$ is the covariance of $\boldsymbol{X}_k$ conditional to $\{\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_k\}$, $\boldsymbol{C}_{k+1}$ is the covariance of $\boldsymbol{X}_{k+1}$ conditional to $\{\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_k\}$, and $\boldsymbol{\Sigma}_k$ is the covariance of $\boldsymbol{\xi}_k$. Upon definition of

$$\boldsymbol{K}_{k+1} = \boldsymbol{C}_{k+1} \boldsymbol{B}_k^T \left[\boldsymbol{B}_k \boldsymbol{C}_{k+1} \boldsymbol{B}_k^T + \boldsymbol{Q}_{k+1} \right]^{-1} \qquad \text{(Kalman gain)}, \tag{53}$$

where $\boldsymbol{Q}_{k+1}$ is the covariance of $\boldsymbol{\eta}_{k+1}$, we can write the update/filtering step (49) as

$$\text{update step:} \quad \begin{cases} \widehat{\boldsymbol{\mu}}_{k+1} = \boldsymbol{\mu}_{k+1} + \boldsymbol{K}_{k+1} \left(\boldsymbol{y}_{k+1} - \boldsymbol{B}_{k+1} \boldsymbol{\mu}_{k+1} \right) & \text{mean} \\ \widehat{\boldsymbol{C}}_{k+1} = \left(\boldsymbol{I} - \boldsymbol{K}_{k+1} \boldsymbol{B}_{k+1} \right) \boldsymbol{C}_{k+1} & \text{covariance} \end{cases}. \tag{54}$$

Here, $\widehat{\boldsymbol{\mu}}_k = \mathbb{E}\{\boldsymbol{X}_{k+1} | \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_{k+1}\}$ is the posterior mean of $\boldsymbol{X}_{k+1}$ after the observation of $\boldsymbol{Y}_{k+1}$ (denoted as $\boldsymbol{y}_{k+1}$). Similarly, $\widehat{\boldsymbol{C}}_{k+1}$ is the covariance of $\boldsymbol{X}_{k+1}$ conditional to $\{\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_{k+1}\}$, i.e., after the observation of $\boldsymbol{Y}_{k+1}$ comes in.

In other words, $\boldsymbol{\mu}_{k+1}$ and $\boldsymbol{C}_{k+1}$ are the mean and the covariance of the PDF $p(\boldsymbol{x}_{k+1} | \boldsymbol{y}_1, \dots, \boldsymbol{y}_k)$ defined in (48), while $\widehat{\boldsymbol{\mu}}_{k+1}$ and $\widehat{\boldsymbol{C}}_{k+1}$ are the mean and the covariance of the PDF $p(\boldsymbol{x}_{k+1} | \boldsymbol{y}_1, \dots, \boldsymbol{y}_{k+1})$ defined in (49). With the updated mean $\widehat{\boldsymbol{\mu}}_{k+1}$ and covariance $\widehat{\boldsymbol{C}}_{k+1}$, we go back to (52) and use such updated mean/covariance to do another prediction step forward, and so on so forth.

Example (polynomial prediction of a random signal): Consider the problem of predicting the next element of a random sequence of numbers defined on an evenly-spaced temporal grid using a simple second-order polynomial interpolation/extrapolation scheme. With reference to Figure 3, we have that the second order polynomial that interpolates a function $Z(t)$ at $\{t_{k-2}, t_{k-1}, t_k\}$ can be written in a Lagrangian form as

$$X(t) = Z(t_{k-2}) l_{k-2}(t) + Z(t_{k-1}) l_{k-1}(t) + Z(t_k) l_k(t) \tag{55}$$

where $\{l_j(t)\}$ are Lagrange characteristic polynomials. We have $Z(t_k) = X_k$ (interpolation condition). Moreover, by evaluating $X(t)$ at $t = t_{k+1}$ we obtain the linear multi-step relation

$$X_{k+1} = 3X_k - 3X_{k-1} + X_{k-2}. \tag{56}$$

[7] In general the covariance matrices of the noise sequences $\{\xi_k\}$ and $\{\eta_k\}$ is not known and has to be estimated from data.

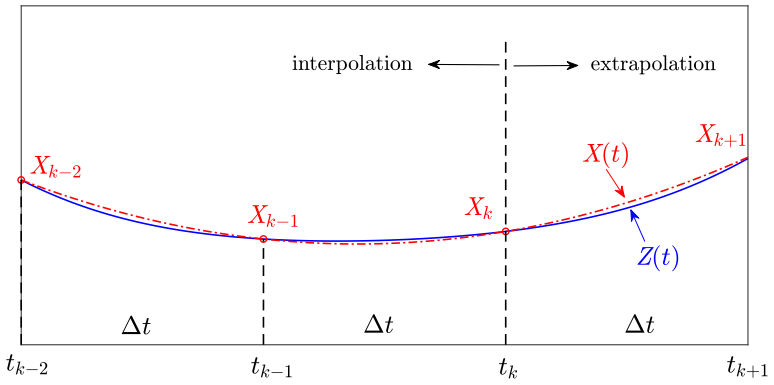

Figure 3: Sketch of the second-order polynomial interpolation/extrapolation process to construct the linear forecasting model (57). The function $Z(t)$ is interpolated at $\{t_{k-2}, t_{k-1}, t_k\}$ by the second-order polynomial $X(t)$, which is then extrapolated at t_{k+1}.

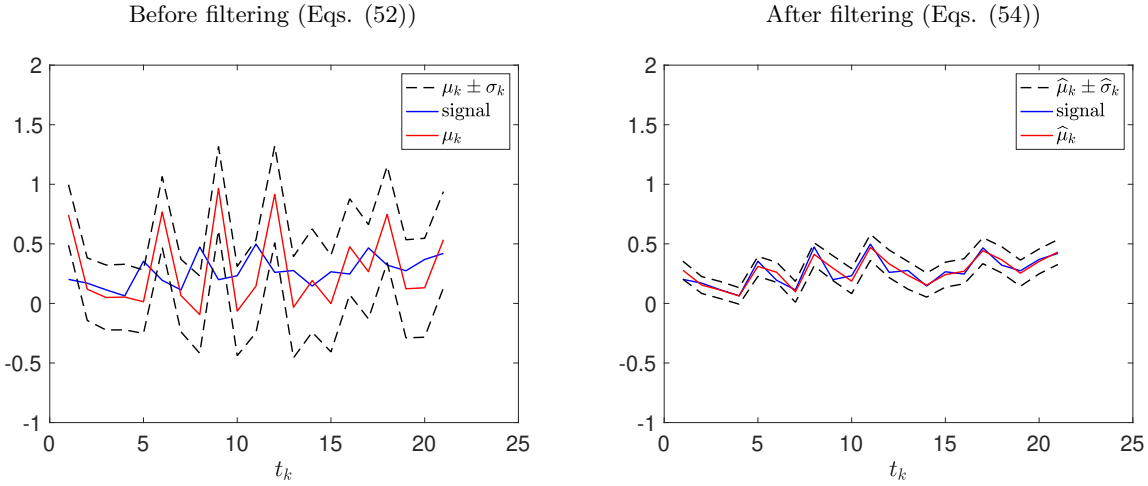

Figure 4: Kalman filtering of a second-order polynomial model to predict (via simple extrapolation) a random signal at the next time step. Shown are the mean and a confidence interval of the predictor at a particular time. We show results of predictor mean and variance before and after filtering.

This allows us to write the following linear prediction model for X_{k+1} given $\{X_k, X_{k-1}, X_{k-2}\}$

$$\underbrace{\begin{bmatrix} X_{k-1} \\ X_k \\ X_{k+1} \end{bmatrix}}_{\boldsymbol{X}_{k+1}} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} X_{k-2} \\ X_{k-1} \\ X_k \end{bmatrix}}_{\boldsymbol{X}_k} + \underbrace{\begin{bmatrix} \xi_{k-2} \\ \xi_{k-1} \\ \xi_k \end{bmatrix}}_{\boldsymbol{\xi}_k}. \tag{57}$$

Since $\boldsymbol{X}_k$ is a random vector, this model should be interpreted as an infinite number of interpolations and extrapolations, one for each sample of $\boldsymbol{X}_k$. This yields a whole ensemble of X_{k+1}, which is then perturbed by noise. Suppose we observe the full state vector $\boldsymbol{X}_k$ at each time, with some additive noise, i.e.,

$$\boldsymbol{Y}_{k+1} = \boldsymbol{X}_{k+1} + \boldsymbol{\eta}_k. \tag{58}$$

In Figure 4 we plot the results obtained by applying the Kalman filter to the prediction model (57).

Karhunen-Loève expansion of random processes

Let $X(t;\omega)$ be a zero-mean square-integrable stochastic process defined on the probability space $(\Omega,\mathcal{B},P)$. "Square-integrable" means that $X(t;\omega)$ has finite second order moment, i.e.,

$$\mathbb{E}\left\{\int_0^T X(t;\omega)^2 dt\right\} < \infty. \tag{59}$$

By using the properties of $L^2(\Omega,\mathcal{B},P)$ spaces (probability spaces of square integrable random variables), it can be shown that $X(t;\omega)$ admits a series expansion

$$X(t;\omega) = \sum_{k=1}^{\infty} \sqrt{\lambda_k}\xi_k(\omega)\psi_k(t), \tag{60}$$

where $\{\xi_1(\omega),\xi_2(\omega),\ldots\}$ is a set of uncorrelated, i.e., orthonormal, random variables satisfying

$$\mathbb{E}\left\{\xi_i(\omega)\xi_j(\omega)\right\} = \delta_{ij}, \tag{61}$$

and $\{\psi_1(t),\psi_2(t),\ldots\}$ are orthonormal (in $L^2([0,T])$) temporal modes

$$\int_0^T \psi_i(t)\psi_j(t)dt = \delta_{ij}. \tag{62}$$

By using the orthogonality properties (61)-(62), we obtain the so-called dispersion relations[8]

$$\xi_k(\omega) = \frac{1}{\sqrt{\lambda_k}}\int_0^T X(t;\omega)\psi_k(t)dt, \tag{64}$$

$$\psi_k(t) = \frac{1}{\sqrt{\lambda_k}}\mathbb{E}\{\xi_k(\omega)X(t;\omega)\}. \tag{65}$$

A substitution of (64) into (65) yields the eigenvalue problem

$$\int_0^T C(t,s)\psi_k(s)ds = \lambda_k^2\psi_k(t). \tag{66}$$

where

$$C(t,s) = \mathbb{E}\left\{X(t;\omega)X(s;\omega)\right\} \tag{67}$$

is the autocorrelation function of the process. In other words, the KL temporal modes are eigenfunctions of the the auto-correlation function of the process. Since $C(t,s)$ is a Mercer's kernel (continuous symmetric non-negative definite kernel) we have that $\{\psi_k(t)\}$ is a complete orthonormal basis of $L^2([0,T])$.

Example: Let us compute the KL expansion of a stochastic process with *exponential* auto-correlation function

$$C(t,s) = \frac{\sigma^2}{2\tau}e^{-|t-s|/\tau}, \tag{68}$$

where τ denotes the correlation time. Note that (68) is an element of a Dirac delta sequence. This implies that

$$\lim_{\tau\to 0}\frac{\sigma^2}{2\tau}e^{-|t-s|/\tau} = \sigma^2\delta(t-s). \tag{69}$$

[8]It is straightforward to show that (65) follows form the variational principle

$$\min_{\psi_k} E([\psi_1,\psi_2,\ldots]) = \min_{\psi_k}\int_0^T \mathbb{E}\left\{\left|X(t;\omega) - \sum_{k=1}^{\infty}\sqrt{\lambda_k}\xi_k(\omega)\psi_k(t)\right|^2\right\}dt. \tag{63}$$

The eigenvalue problem (68) with $C(t,s)$ defined in(68) admits the analytical solution[9] (see [4])

$$\psi_k(t) = \frac{\tau z_k \cos(z_k t) + \sin(z_k t)}{\sqrt{\frac{1}{2}\left(\tau^2 z_k^2 + 1\right) T + \left(\tau^2 z_k^2 - 1\right) \frac{\sin(2 z_k T)}{4 z_k} + \frac{\tau}{2}\left(1 - \cos(2 z_k T)\right)}}, \tag{73}$$

where z_k are solution of the transcendental equation

$$\left(z_k^2 - \frac{1}{\tau}\right) \tan(z_k T) - \frac{2 z_k}{\tau} = 0, \tag{74}$$

and

$$\lambda_k = \frac{\sigma^2}{\left(z_k^2 \tau^2 + 1\right)} \tag{75}$$

are the KL eigenvalues. The KL eigenvalues become smaller and smaller as z_k increases. The eigenvalue decay is more pronounced for larger correlation lengths τ, while for very small correlation lengths the eigenvalue decay rate is very small, eventually zero for zero correlation length.

For practical purposes, the KL series expansion (60) is usually truncated to a finite number of terms. As we just discussed, the number of terms is inversely proportional to τ: the smaller τ the larger the number of terms. The number of terms M in the KL series expansion (60) is usually chosen by thresholding the relative "energy" of the process as

$$\frac{\sum_{k=1}^{M} \lambda_k}{\sum_{k=1}^{\infty} \lambda_k} \simeq 0.95. \tag{76}$$

This implies that the modes we retain in the series capture about 95% of the process "energy". In Figure 5 we plot samples of the exponentially correlated Gaussian random process

$$X(t;\omega) = \sin(t) + \frac{\sigma}{2\tau} \sum_{k=1}^{M} \sqrt{\lambda}_k \xi_k(\omega) \psi_k(t) \qquad t \in [0, 20], \tag{77}$$

for $\tau = 1$ and $\tau = 0.1$.

Remark: In the case where (66) cannot be solved analytically, we can resort to numerical method for Fredholm eigenvalue problems, e.g., Finite-difference methods, spectral methods, or Galerkin methods (see

[9] To compute the analytical solution of the KL eigenvalue problem (66) with exponential covariance (68) let us first rewrite it as

$$\int_0^T e^{-c|t-s|} \psi_k(s) ds = \hat{\lambda}_k \psi_k(t), \qquad c = \frac{1}{\tau}, \qquad \hat{\lambda}_k = \frac{2\tau}{\sigma^2} \lambda_k. \tag{70}$$

Differentiating with respect to t the equivalent expression

$$\int_0^t e^{-c(t-s)} \psi_k(s) ds + \int_t^T e^{c(t-s)} \psi_k(s) ds = \hat{\lambda}_k \psi_k(t) \tag{71}$$

yields the second-order boundary value problem

$$\begin{cases} \dfrac{d^2 \psi_k}{dt^2} = \dfrac{c^2 \hat{\lambda}_k - 2c}{\hat{\lambda}_k} \psi_k(t) \\ \dfrac{d\psi_k(t)}{dt} = c\psi(0) \\ \dfrac{d\psi_k(T)}{dt} = c\psi(T) \end{cases} \tag{72}$$

The solution of the BVP (72) is (73)-(75).

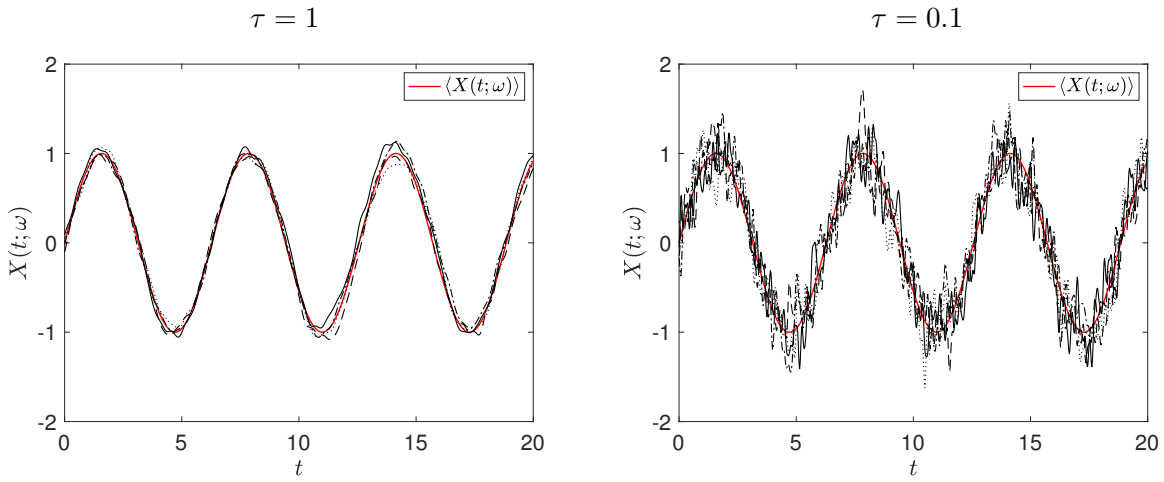

Figure 5: Samples of the exponentially correlated Gaussian random process (77) for different correlation times τ. The KL The mean of the process is shown in red. The truncation threshold for the number of terms M is set at 95% of the energy of the process (see Eq. (76)).

e.g., [8]). Of course it is also possible to define KL expansions of *random fields* by simply generalizing the bi-orthogonal series (60) as

$$X(\boldsymbol{x};\omega) = \sum_{k=1}^{\infty} \sqrt{\lambda}_k \xi_k(\omega)\psi_k(\boldsymbol{x}). \tag{78}$$

The computation of the KL expansion follows exactly the same steps as before, i.e., $\psi_k(\boldsymbol{x})$ are solutions to the eigenvalue problem

$$\int_V C(\boldsymbol{x},\boldsymbol{y})\psi_k(\boldsymbol{y})d\boldsymbol{y} = \lambda_k \psi_k(\boldsymbol{x}), \tag{79}$$

where V is some spatial domain.

Remark: To sample realizations of the random process (60) we need to sample the random variables $\{\xi_1,\dots,\xi_M\}$. Such random variables are (by construction) orthonormal (see (61)), i.e., they are uncorrelated and have variance equal to one. Clearly, if $\{\xi_1,\dots,\xi_M\}$ are jointly Gaussian then we know that the condition (61) is necessary and sufficient for independence. Hence, in the Gaussian case, sampling the joint PDF of $\{\xi_1,\dots,\xi_M\}$ reduces to sampling the PDF of an independent set of one-dimensional Gaussian random variables with zero mean and variance one. More generally, if we have available the joint PDF $p(\xi_1,\dots,\xi_M)$, e.g., by computing (64), then we can sample it using Markov Chain Monte Carlo (MCMC) methods, e.g., the Metropolis-Hastings algorithm or Gibbs sampling.

Wiener process

The Wiener process (also known as "Brownian motion") is a zero-mean continuous-time random process satisfying the following conditions:

- The increment $X(t+\tau;\omega) - X(t;\omega)$ is a Gaussian random variable with zero mean and variance τ. In other words, the conditional probability density of $X(t+\tau;\omega)$ given $X(t;\omega)$, i.e., $p(x(t+\tau)|x(t))$ is Gaussian with mean $X(t;\omega)$ and variance τ[10].

[10]Random processes with conditional mean at time t_{k+1} equal to the process observation at time t_k (given the whole past history) are called *martingales*. In a discrete setting a martingale is defined as

$$E\left\{X_{k+1}|X_k = x_k,\dots,X_0 = x_0\right\} = x_k. \tag{80}$$

- The random variables (increments)

$$X(t_1;\omega) - X(t_0;\omega) \qquad \text{and} \qquad X(t_3;\omega) - X(t_2;\omega) \tag{81}$$

  are statistically independent for $t_0 < t_1 \leq t_2 < t_3$. In other words, the Wiener process is an *independent increment* process.

- The process $X(t;\omega)$ is continuous with probability one, i.e.,

$$P\left(\{\omega : \lim_{s\to t} |X(s;\omega) - X(t;\omega)| = 0\}\right) = 1 \qquad \text{for all } t \geq 0. \tag{82}$$

  This means that almost all (except sets of measure zero) sample paths are continuous in the classical sense, but the process $X(t;\omega)$ is nowhere differentiable. Continuity with probability one implies continuity in probability, and therefore mean square continuity and continuity in distribution.

An very clear construction of the Wiener process is provided by Wiener himself in [11, Lecture 1]. The simplest algorithm to sample a Wiener process leverages the fact that the process has Gaussian distributed independent increments. Let $\{t_k\}_{k=1,\ldots,n}$ be n distinct time instants

$$0 = t_0 < t_1 < \ldots < t_n. \tag{83}$$

Then

$$X(t_k;\omega) = \sum_{j=1}^{k} \sqrt{\Delta t_j}\xi_j(\omega) \qquad \Delta t_j = t_j - t_{j-1}, \tag{84}$$

where $\{\xi_j(\omega)\}$ are independent random variables with mean zero and variance 1. A closer look at (84), reveals

$$X(t_1;\omega) = \sqrt{\Delta t_1}\xi_1(\omega), \tag{85}$$

$$X(t_2;\omega) = X(t_1;\omega) + \sqrt{\Delta t_2}\xi_2(\omega) = \sqrt{\Delta t_1}\xi_1(\omega) + \sqrt{\Delta t_2}\xi_2(\omega) \tag{86}$$

$$\ldots$$

Since $X(t_k;\omega)$ is a superimposition of essentially an infinite number of independent random vairable, it is rather straightforward to show that the one time PDF of $X(t;\omega)$ is

$$p(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}, \tag{87}$$

i.e., Gaussian. This equation also follows from the conditional PDF identity

$$p(x,t) = \int_{\infty}^{\infty} p(x,t|y,s)p(y,s)dy \qquad t > s, \tag{88}$$

where $p(x,t|y,s)$ is the transition density[11], and $p(y,s)$ is the PDF of $X(s;\omega)$. If we set $s = 0$ then $p(y,0) = \delta(y)$ and, of course, this yields (87). The auto-correlation function of the Wiener process is

$$C(t,s) = \min(t,s) \tag{90}$$

The Wiener process is therefore a martingale.

[11] From the recurrence relation

$$X(t_{k+1};\omega) = X(t_k;\omega) + \sqrt{\Delta t_{k+1}}\xi_{k+1}(\omega) \tag{89}$$

with $\xi_{k+1}(\omega)$ Gaussian with zero mean and variance one we see that the conditional PDF $p(x,t|y,s)$ is Gaussian with mean $X(t_k)$ and variance $t - s$.

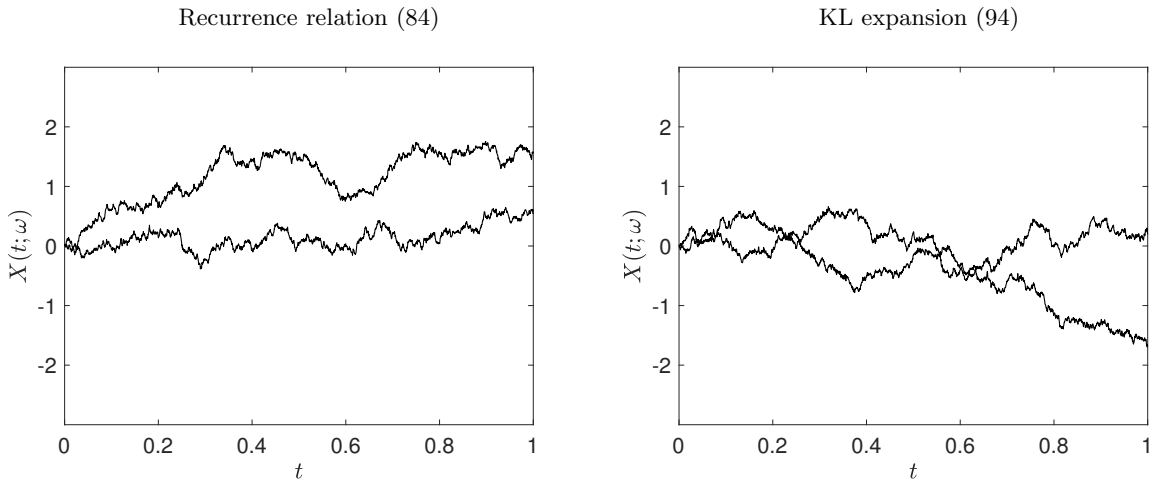

Figure 6: Wiener processes obtained by sampling the Karhunen-Loève expansion (94) with 10^5 terms on a temporal grid with 2000 points in $[0,1]$, and by iterating (84) on the same temporal grid.

With the autocorrelation function available we can compute a KL expansion of the Wiener process following the procedure outlined in the previous section. If we consider the time interval $[0,1]$ this yields the eigenvalue problem

$$\int_0^1 C(t,s)\psi_k(s) = \lambda_k \psi_k(t), \tag{91}$$

the solution of which is

$$\psi_k(t) = \sqrt{2}\sin\left(\left[k-\frac{1}{2}\right]\pi t\right) \quad k = 1,2,\dots \tag{92}$$

and

$$\lambda_k = \frac{4}{\pi^2\left(2k-1\right)^2}. \tag{93}$$

Substituting (92) and (93) into (60) yields

$$X(t;\omega) = \sum_{k=1}^{\infty} \frac{2\sqrt{2}}{\pi\left(2k-1\right)}\xi_k(\omega)\sin\left(\left[k-\frac{1}{2}\right]\pi t\right), \tag{94}$$

where $\xi_k(\omega)$ are independent Gaussian random variables with zero mean and variance one (they satisfy (61)). The series expansion in (94) can be eventually truncated to a finite number of terms, depending on the threshold set on the eigenvalues (93) (which decay as $1/k$). In Figure 6 we plot a few samples of the Wiener process we obtain by sampling (94) with 10^5 terms on a temporal grid with 2000 points in $[0,1]$, and the Wiener process we obtain by iterating (84) on the same temporal grid. Note that if $X(t;\omega)$ is a Wiener process in $t \in [0,1]$ then

$$\sqrt{T}X\left(\frac{t}{T};\omega\right) \qquad t \in [0,T] \tag{95}$$

is a Wiener process in $[0,T]$. This expression is obtained by simply changing the variables in the integral equation (91). The expression (95) shows that features of a Wiener process do not change while zooming in or out. In other words, the Wiener process is *self-similar*.

ODEs driven by Gaussian white noise

Consider the following stochastic differential equation[12] (SDE)

$$d\boldsymbol{X}(t) = \boldsymbol{m}(\boldsymbol{X}(t,t))dt + d\boldsymbol{W}(t) \qquad t \in [0,T], \tag{97}$$

where $\boldsymbol{m}(\boldsymbol{x},t)$ is a smooth vector field, and $\boldsymbol{W}(t)$ is a vector of independent Wiener processes (each component is a Wiener process). We discretize the SDE (97) with the Euler-Maruyama scheme on an evenly-spaced grid with N points in $[0,T]$ to obtain

$$\boldsymbol{X}_{k+1} = \boldsymbol{X}_k + \boldsymbol{m}(\boldsymbol{X}_k, t_k)\Delta t + \Delta \boldsymbol{W}_k \tag{98}$$

where $\Delta \boldsymbol{W}_k$ is a Gaussian random vector with zero mean and variance $\Delta t = T/N$ (same variance in each component). Since the Wiener process is an independent-increment process, we have that the increments $\{\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots\}$ are statistically independent and identically distributed. Specifically, they are all jointly Gaussian with zero mean and diagonal covariance with Δt along the diagonal. This implies that the nonlinear autoregressive process (98) is Markovian (see Appendix A) with transition density

$$p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) = \frac{1}{(2\pi\Delta t)^{n/2}} \exp\left(-\frac{1}{2\Delta t}\left\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k - \boldsymbol{m}(\boldsymbol{x}_k, t_k)\Delta t\right\|_2^2\right). \tag{99}$$

Fokker-Planck equation. For simplicity, let us first consider the one-dimensional case with time-independent "drift" $m(x,t)$. In this case the discrete dynamical system (98) is written as

$$X_{k+1} = X_k + m(X_k)\Delta t + \Delta W_k. \tag{100}$$

Define the random variable $X(t_k) = X_k$ and the corresponding probability density $p(x,t_k) = p(x_k)$. We are interested in deriving an evolution equation for $p(x,t)$ by taking the limit

$$\lim_{\Delta t \to 0} \frac{p(x, t_k + \Delta t) - p(x,t_k)}{\Delta t} \tag{101}$$

To this end, let us first push forward in time (by a small amount Δt) the PDF $p(x,t_k)$ using the transition density (99), i.e.,

$$\begin{aligned} p(x, t_k + \Delta t) &= \int_{-\infty}^{\infty} p(x|y)p(y,t_k)dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\Delta t}} \exp\left[-\frac{1}{2\Delta t}\left(x - y - m(y)\Delta t\right)^2\right] p(y,t_k)dy. \end{aligned} \tag{102}$$

Note that for small Δt and any given x we have that the kernel $p(x|y)$ (considered as a function of y) is very much concentrated around x. This allows us to expand the integrand in (102) in a Taylor series in y for y very close to x. To this end, it is convenient to change the integration variable y as

$$\epsilon = y - x \tag{103}$$

and (102) as

$$p(x, t_k + \Delta t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\Delta t}} \exp\left[-\frac{1}{2\Delta t}\left(\epsilon + m(x+\epsilon)\Delta t\right)^2\right] p(x+\epsilon, t_k)d\epsilon. \tag{104}$$

[12] The SDE (97) can also be written as

$$\frac{d\boldsymbol{X}(t)}{dt} = \boldsymbol{m}(\boldsymbol{X}(t),t) + \dot{\boldsymbol{W}}(t). \tag{96}$$

The quantity $\dot{\boldsymbol{W}}(t) = d\boldsymbol{W}(t)/dt$ is the "derivative" of the Wiener process, which is known as *Gaussian white noise*.

Expanding the exponential in a Taylor series at $\epsilon = 0$ yields

$$p(x,t_k+\Delta t) \simeq \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\Delta t}} \exp\left[-\frac{1}{2\Delta t}\left(\epsilon(1+m'(x)\Delta t) + m(x)\Delta t + \cdots\right)^2\right] \times \left[p(x,t_k) + \frac{\partial p(x,t_k)}{\partial x}\epsilon + \frac{1}{2}\frac{\partial^2 p(x,t_k)}{\partial x^2}\epsilon^2 + \frac{1}{6}\frac{\partial^3 p(x,t_k)}{\partial x^3}\epsilon^3 + \cdots\right] d\epsilon. \tag{105}$$

To compute the Gaussian integrals, we consider another change of variable[13]

$$\eta = \epsilon\gamma(\Delta t) \qquad \gamma(\Delta t) = 1 + m'(x)\Delta t \tag{107}$$

Substituting (107) into (105) yields,

$$\begin{aligned} p(x,t_k+\Delta t) \simeq & \frac{1}{\gamma(\Delta t)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\Delta t}} \exp\left[-\frac{1}{2\Delta t}\left(\eta + m(x)\Delta t\right)^2\right] \times \\ & \left[p(x,t_k) + \frac{1}{\gamma(\Delta t)}\frac{\partial p(x,t_k)}{\partial x}\eta + \frac{1}{2\gamma^2(\Delta t)}\frac{\partial^2 p(x,t_k)}{\partial x^2}\eta^2 + \frac{1}{6\gamma^3(\Delta t)}\frac{\partial^3 p(x,t_k)}{\partial x^3}\eta^3 + \cdots\right] d\eta \\ = & \frac{1}{\gamma(\Delta t)}\left[p(x,t_k) - \frac{m(x)\Delta t}{\gamma(\Delta t)}\frac{\partial p(x,t_k)}{\partial x} + \frac{(m(x)\Delta t)^2 + \Delta t}{2\gamma^2(\Delta t)}\frac{\partial^2 p(x,t_k)}{\partial x^2} + \right. \\ & \left. \frac{(m(x)\Delta t)^3 - 3m(x)\Delta t^2}{6\gamma^3(\Delta t)}\frac{\partial^3 p(x,t_k)}{\partial x^3} + \cdots\right]. \end{aligned} \tag{108}$$

In fact the second- and third-order moments of a Gaussian distribution mean $\mu = -m(x)\Delta t$ and variance $\sigma^2 = \Delta t$ are $\mu^2 + \sigma^2$ and $\mu^3 + 3\mu\sigma^2$, respectively. Lastly, note that for small Δt we have

$$\frac{p(x,t_k)}{1+m'(x)\Delta t} \simeq p(x,t_k) - \Delta t m'(x) p(x,t_k). \tag{109}$$

Putting all this together, we finally obtain

$$\begin{aligned} \frac{\partial p(x,t_k)}{\partial t} &= \lim_{\Delta t\to 0} \frac{p(x,t_k+\Delta t) - p(x,t_k)}{\Delta t} \\ &= \lim_{\Delta t\to 0}\left[-m'(x)p(x,t_k) - \frac{m(x)}{\gamma^2(\Delta t)}\frac{\partial p(x,t_k)}{\partial x} + \frac{m(x)^2\Delta t + 1}{2\gamma^3(\Delta t)}\frac{\partial^2 p(x,t_k)}{\partial x^2} + \right. \\ &\qquad \left. \frac{m(x)^3\Delta t^2 - 3m(x)\Delta t}{6\gamma^4(\Delta t)}\frac{\partial^3 p(x,t_k)}{\partial x^3} + \cdots\right] \\ &= -m'(x)p(x,t_k) - m(x)\frac{\partial p(x,t_k)}{\partial x} + \frac{1}{2}\frac{\partial^2 p(x,t_k)}{\partial x^2}. \end{aligned} \tag{110}$$

Hence we obtain

$$\frac{\partial p(x,t)}{\partial t} + \frac{\partial}{\partial x}\left[m(x)p(x,t)\right] = \frac{1}{2}\frac{\partial^2 p(x,t)}{\partial x^2}. \tag{111}$$

This equation is an advection-diffusion equation known as Fokker-Planck equation, and it governs the one-time PDF $p(x,t)$ of the random process defined by the SDE

$$dX(t) = m(X(t))dt + dW(t), \qquad X(0) = X_0(\omega). \tag{112}$$

[13]Note that $\gamma(\Delta t) = 1 + m'(x)\Delta t$ in (107) is a perturbation of 1 for Δt sufficiently small. Moreover,

$$\lim_{\Delta t\to 0}\gamma(\Delta t) = 1. \tag{106}$$

More generally, the Fokker Planck equation corresponding to SDE[14]

$$dX(t) = m(X(t),t)dt + s(X(t),t)dW(t), \qquad X(0) = X_0(\omega). \tag{114}$$

is

$$\frac{\partial p(x,t)}{\partial t} + \frac{\partial}{\partial x}\left[m(x,t)p(x,t)\right] = \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[D(x,t)p(x,t)\right], \qquad D(x,t) = s(x,t)^2. \tag{115}$$

Multivariate Fokker-Planck equation. By following a similar derivation, it can be shown that the Fokker Plank equation corresponding to the general n-dimensional nonlinear SDE driven by m-dimensional Gaussian white noise

$$d\boldsymbol{X}(t) = \boldsymbol{m}(\boldsymbol{X}(t),t)dt + \boldsymbol{S}(\boldsymbol{X}(t),t)d\boldsymbol{W}(t), \tag{116}$$

where $\boldsymbol{m} \in \mathbb{R}^n$ and $\boldsymbol{S}$ is a $n \times m$ matrix, and in the limit $\Delta t \to 0$ is

$$\frac{\partial p(\boldsymbol{x},t)}{\partial t} + \sum_{j=1}^{n}\frac{\partial}{\partial x_j}\left[m_j(\boldsymbol{x})p(\boldsymbol{x},t)\right] = \frac{1}{2}\sum_{k,j=1}^{n}\frac{\partial^2}{\partial x_k \partial x_k}\left[D_{jk}(\boldsymbol{x},t)p(\boldsymbol{x},t)\right]. \tag{117}$$

Here

$$D_{jk}(\boldsymbol{x},t) = \sum_{p=1}^{m} S_{jp}(\boldsymbol{x},t)S_{kp}(\boldsymbol{x},t) \tag{118}$$

is the diffusion matrix induced by $\boldsymbol{S}(\boldsymbol{x},t)$.

Statistical steady states. The Fokker-Planck equation may or may not converge (asymptotically in t to an equilibrium PDF. Even in the case where the drift coefficient $\boldsymbol{m}(\boldsymbol{X},t)$ is time-independent, and the diffusion $\boldsymbol{S}(\boldsymbol{X},t)$ is constantly equal to one there may not be a any statistically steady states. Moreover, such steady states, whenever they exist, are usually not Gaussian.

Ito's Lemma. Let $X(t)$ be a stochastic process defined by the SDE (112). Consider a smooth function $g(x,t)$ (at least of class C^2 in x and t). Ito's lemma states that the stochastic process (phase space function of $X(t)$)

$$Y(t) = g(X(t),t) \tag{119}$$

satisfies the SDE

$$dY(t) = \left[\frac{\partial g(X(t),t)}{\partial t} + m(X(t))\frac{\partial g(X(t),t)}{\partial x} + \frac{1}{2}\frac{\partial^2 g(X(t),t)}{\partial x^2}\right]dt + \frac{\partial g(X(t),t)}{\partial x}dW(t). \tag{120}$$

A more general version of Ito's lemma is available for the multivariate SDE (116).

[14] A simple example of and SDE of the form (114) is

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t), \tag{113}$$

which defines the so-called *geometric Brownian motion* process. This is often used in finance to model stock prices.

Appendix A: Nonlinear autoregressive Markov processes

In this Appendix we prove that if the random vectors $\{\boldsymbol{\xi}_0,\ldots,\boldsymbol{\xi}_{n-1}\}$ in (21) are statistically independent, and if $\boldsymbol{\xi}_n$ is independent of past and current states, i.e., $\{\boldsymbol{X}_0,\ldots,\boldsymbol{X}_n\}$, then (21) defines a Markov process[15]. In other words, we show that discrete nonlinear autoregressive processes driven by white noise (not necessarily Gaussian) are Markov processes. To this end, we first notice that the full statistical information of the autoregressive process is defined by the joint probability density function of $\{\boldsymbol{X}_n,\ldots,\boldsymbol{X}_0,\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\}$. Let us denote by $p\left(\boldsymbol{x}_n,\ldots,\boldsymbol{x}_0,\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\right)$ such joint density function. By using well-known identities for conditional PDFs we can write

$$\begin{aligned} p\left(\boldsymbol{x}_n,\ldots,\boldsymbol{x}_0,\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\right) =& p\left(\boldsymbol{x}_n|\boldsymbol{x}_{n-1},\ldots,\boldsymbol{x}_0,\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\right)\times \\ & p\left(\boldsymbol{x}_{n-1}|\boldsymbol{x}_{n-2},\ldots,\boldsymbol{x}_0,\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\right)\times\cdots \\ & p\left(\boldsymbol{x}_0|\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\right)p\left(\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\right). \end{aligned} \tag{121}$$

Clearly, from equation (21) it follows that

$$p\left(\boldsymbol{x}_n|\boldsymbol{x}_{n-1},\ldots,\boldsymbol{x}_0,\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\right)=p\left(\boldsymbol{x}_n|\boldsymbol{x}_{n-1},\boldsymbol{\xi}_{n-1}\right). \tag{122}$$

This allows us to rewrite (121) as

$$p\left(\boldsymbol{x}_n,\ldots,\boldsymbol{x}_0,\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\right)=p\left(\boldsymbol{x}_n|\boldsymbol{x}_{n-1},\boldsymbol{\xi}_{n-1}\right)p\left(\boldsymbol{x}_{n-1}|\boldsymbol{x}_{n-2},\boldsymbol{\xi}_{n-2}\right)\cdots p\left(\boldsymbol{x}_1|\boldsymbol{x}_0,\boldsymbol{\xi}_0\right)p\left(\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\right). \tag{123}$$

If the vectors $\{\boldsymbol{\xi}_{n_1},\ldots,\boldsymbol{\xi}_0\}$ are statistically independent, and if $\boldsymbol{\xi}_k$ is independent of $\{\boldsymbol{X}_0,\ldots,\boldsymbol{X}_k\}$ then

$$\begin{aligned} p\left(\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\right) =& p\left(\boldsymbol{\xi}_{n-1}\right)\cdots p\left(\boldsymbol{\xi}_0\right) \\ =& p\left(\boldsymbol{\xi}_{n-1}|\boldsymbol{x}_{n-1}\right)\cdots p\left(\boldsymbol{\xi}_0|\boldsymbol{x}_0\right). \end{aligned} \tag{124}$$

Substituting (124) into (123) and integrating over $\{\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\}$ yields

$$\begin{aligned} p\left(\boldsymbol{x}_n,\ldots,\boldsymbol{x}_0\right) =& \underbrace{\left(\int p\left(\boldsymbol{x}_n|\boldsymbol{x}_{n-1},\boldsymbol{\xi}_{n-1}\right)p\left(\boldsymbol{\xi}_{n-1}|\boldsymbol{x}_{n-1}\right)d\boldsymbol{\xi}_{n-1}\right)}_{p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})}\times\cdots \\ & \underbrace{\left(\int p\left(\boldsymbol{x}_1|\boldsymbol{x}_0,\boldsymbol{\xi}_0\right)p\left(\boldsymbol{\xi}_0|\boldsymbol{x}_0\right)d\boldsymbol{\xi}_0\right)}_{p(\boldsymbol{x}_1|\boldsymbol{x}_0)}p(\boldsymbol{x}_0) \\ =& p\left(\boldsymbol{x}_n|\boldsymbol{x}_{n-1}\right)p\left(\boldsymbol{x}_{n-1}|\boldsymbol{x}_{n-2}\right)\cdots p\left(\boldsymbol{x}_1|\boldsymbol{x}_0\right)p(\boldsymbol{x}_0), \end{aligned} \tag{125}$$

which clearly represents the joint PDF of a Markov process. Note that the Markovian property of the process $\{\boldsymbol{X}_n\}$ relies heavily on the fact that the joint PDF of the random vectors $\{\boldsymbol{\xi}_{n-1},\ldots,\boldsymbol{\xi}_0\}$ can be factorized as a product of conditional densities as in (124), i.e., that the random vectors are statistically independent, and also that $\boldsymbol{\xi}_n$ is independent of past and current states, i.e., $\{\boldsymbol{X}_0,\ldots,\boldsymbol{X}_n\}$.

[15] Note that $\boldsymbol{X}_n$ depends on $\boldsymbol{\xi}_{n-1}$ via the recursion (21).

References

[1] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control*. Wiley, 2008.

[2] V. I. Klyatskin. *Dynamics of stochastic systems*. Elsevier Publishing Company, 2005.

[3] Pierre L'Ecuyer. *Random Number Generation*, pages 35–71. Springer Berlin Heidelberg, 2012.

[4] M. Jardak M, C. H. Su, and G. E. Karniadakis. Spectral polynomial chaos solutions of the stochastic advection equation. *Journal of Scientific Computing*, 17:319–338, 2002.

[5] A. S. Monin and A. M. Yaglom. *Statistical Fluid Mechanics, Volume II: Mechanics of Turbulence*. Dover, 2007.

[6] H. Rue and L. Held. *Gaussian Markov random fields*. Chapman & Hall/CRC, 2005.

[7] D. Venturi. The numerical approximation of nonlinear functionals and functional differential equations. *Physics Reports*, 732:1–102, 2018.

[8] D. Venturi, M. Choi, and G. E. Karniadakis. Supercritical quasi-conduction states in stochastic Rayleigh-Bénard convection. *Int. J. Heat and Mass Transfer*, 55(13-14):3732–3743, 2012.

[9] D. Venturi and A. Dektor. Spectral methods for nonlinear functionals and functional differential equations. *Res. Math. Sci.*, 8(27):1–39, 2021.

[10] D. Venturi and X. Li. The Mori-Zwanzig formulation of deep learning. *ArXiv*, (2209.05544):1–40, 2022.

[11] N. Wiener. *Nonlinear problems in random theory*. MIT Press, 1966.