C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006,
ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

Appendix B

Gaussian Markov Processes

Particularly when the index set for a stochastic process is one-dimensional such as the real line or its discretization onto the integer lattice, it is very interesting to investigate the properties of Gaussian *Markov* processes (GMPs). In this Appendix we use $X(t)$ to define a stochastic process with continuous time parameter t. In the discrete time case the process is denoted $\ldots, X_{-1}, X_0, X_1, \ldots$ etc. We assume that the process has zero mean and is, unless otherwise stated, stationary.

A discrete-time autoregressive (AR) process of order p can be written as

AR process

$$X_t \;=\; \sum_{k=1}^{p} a_k X_{t-k} + b_0 Z_t, \tag{B.1}$$

where $Z_t \sim \mathcal{N}(0,1)$ and all Z_t's are i.i.d. . Notice the order-p Markov property that given the history $X_{t-1}, X_{t-2}, \ldots$, X_t depends only on the previous p X's. This relationship can be conveniently expressed as a graphical model; part of an AR(2) process is illustrated in Figure B.1. The name autoregressive stems from the fact that X_t is predicted from the p previous X's through a regression equation. If one stores the current X and the $p-1$ previous values as a state vector, then the AR(p) scalar process can be written equivalently as a vector AR(1) process.

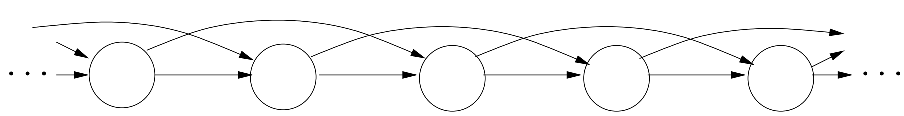

Figure B.1: Graphical model illustrating an AR(2) process.

Moving from the discrete time to the continuous time setting, the question arises as to how generalize the Markov notion used in the discrete-time AR process to define a continuoous-time AR process. It turns out that the correct generalization uses the idea of having not only the function value but also p of its derivatives at time t giving rise to the stochastic differential equation (SDE)[1]

SDE: stochastic differential equation

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006,
ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

$$a_p X^{(p)}(t) + a_{p-1} X^{(p-1)}(t) + \ldots + a_0 X(t) \;=\; b_0 Z(t), \tag{B.2}$$

where $X^{(i)}(t)$ denotes the ith derivative of $X(t)$ and $Z(t)$ is a white Gaussian noise process with covariance $\delta(t - t')$. This white noise process can be considered the derivative of the Wiener process. To avoid redundancy in the coefficients we assume that $a_p = 1$. A considerable amount of mathematical machinery is required to make rigorous the meaning of such equations, see e.g. Øksendal [1985]. As for the discrete-time case, one can write eq. (B.2) as a first-order vector SDE by defining the state to be $X(t)$ and its first $p-1$ derivatives.

We begin this chapter with a summary of some Fourier analysis results in section B.1. Fourier analysis is important to linear time invariant systems such as equations (B.1) and (B.2) because $e^{2\pi i s t}$ is an eigenfunction of the corresponding difference (resp differential) operator. We then move on in section B.2 to discuss continuous-time Gaussian Markov processes on the real line and their relationship to the same SDE on the circle. In section B.3 we describe discrete-time Gaussian Markov processes on the integer lattice and their relationship to the same difference equation on the circle. In section B.4 we explain the relationship between discrete-time GMPs and the discrete sampling of continuous-time GMPs. Finally in section B.5 we discuss generalizations of the Markov concept in higher dimensions. Much of this material is quite standard, although the relevant results are often scattered through different sources, and our aim is to provide a unified treatment. The relationship between the second-order properties of the SDEs on the real line and the circle, and difference equations on the integer lattice and the regular polygon is, to our knowledge, novel.

B.1 Fourier Analysis

We follow the treatment given by Kammler [2000]. We consider Fourier analysis of functions on the real line $\mathbb{R}$, of periodic functions of period l on the circle $\mathbb{T}_l$, of functions defined on the integer lattice $\mathbb{Z}$, and of functions on $\mathbb{P}_N$, the regular N-polygon, which is a discretization of $\mathbb{T}_l$.

For sufficiently well-behaved functions on $\mathbb{R}$ we have

$$f(x) \;=\; \int_{-\infty}^{\infty} \tilde{f}(s) e^{2\pi i s x}\, ds, \qquad \tilde{f}(s) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i s x}\, dx. \tag{B.3}$$

We refer to the equation on the left as the *synthesis* equation, and the equation on the right as the *analysis* equation.

For functions on $\mathbb{T}_l$ we obtain the Fourier series representations

$$f(x) \;=\; \sum_{k=-\infty}^{\infty} \tilde{f}[k] e^{2\pi i k x/l}, \qquad \tilde{f}[k] \;=\; \frac{1}{l}\int_0^l f(x) e^{-2\pi i k x/l}\, dx, \tag{B.4}$$

[1] The a_k coefficients in equations (B.1) and (B.2) are not intended to have a close relationship. An approximate relationship might be established through the use of finite-difference approximations to derivatives.

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

where $\tilde{f}[k]$ denotes the coefficient of $e^{2\pi ikx/l}$ in the expansion. We use square brackets [ ] to denote that the argument is discrete, so that X_t and $X[t]$ are equivalent notations.

Similarly for $\mathbb{Z}$ we obtain

$$f[n] \;=\; \int_0^l \tilde{f}(s)e^{2\pi isn/l}\,ds, \qquad \tilde{f}(s) \;=\; \frac{1}{l}\sum_{n=-\infty}^{\infty} f[n]e^{-2\pi isn/l}. \tag{B.5}$$

Note that $\tilde{f}(s)$ is periodic with period l and so is defined only for $0 \le s < l$ to avoid aliasing. Often this transform is defined for the special case $l = 1$ but the general case emphasizes the duality between equations (B.4) and (B.5).

Finally, for functions on $\mathbb{P}_N$ we have the discrete Fourier transform

$$f[n] \;=\; \sum_{k=0}^{N-1} \tilde{f}[k]e^{2\pi ikn/N}, \qquad \tilde{f}[k] \;=\; \frac{1}{N}\sum_{n=0}^{N-1} f[n]e^{-2\pi ikn/N}. \tag{B.6}$$

Note that there are other conventions for Fourier transforms, particularly those involving $\omega = 2\pi s$. However, this tends to destroy symmetry between the analysis and synthesis equations so we use the definitions given above.

In the case of stochastic processes, the most important Fourier relationship is between the covariance function and the power spectrum; this is known as the Wiener-Khintchine theorem, see e.g. Chatfield [1989].

B.1.1 Sampling and Periodization

We can obtain relationships between functions and their transforms on $\mathbb{R}$, $\mathbb{T}_l$, $\mathbb{Z}$, $\mathbb{P}_N$ through the notions of *sampling* and *periodization*.

h-sampling

Definition B.1 h-sampling*: Given a function f on $\mathbb{R}$ and a spacing parameter $h > 0$, we construct a corresponding discrete function ϕ on $\mathbb{Z}$ using*

$$\phi[n] \;=\; f(nh), \qquad n \in \mathbb{Z}. \tag{B.7}$$

□

Similarly we can discretize a function defined on $\mathbb{T}_l$ onto $\mathbb{P}_N$, but in this case we must take $h = l/N$ so that N steps of size h will equal the period l.

periodization by summation

Definition B.2 Periodization by summation*: Let $f(x)$ be a function on $\mathbb{R}$ that rapidly approaches 0 as $x \to \pm\infty$. We can sum translates of the function to produce the l-periodic function*

$$g(x) \;=\; \sum_{m=-\infty}^{\infty} f(x - ml), \tag{B.8}$$

for $l > 0$. Analogously, when ϕ is defined on $\mathbb{Z}$ and $\phi[n]$ rapidly approaches 0 as $n \to \pm\infty$ we can construct a function γ on $\mathbb{P}_N$ by N-summation by setting

$$\gamma[n] \;=\; \sum_{m=-\infty}^{\infty} \phi[n - mN]. \tag{B.9}$$

□

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

Let $\phi[n]$ be obtained by h-sampling from $f(x)$, with corresponding Fourier transforms $\tilde{\phi}(s)$ and $\tilde{f}(s)$. Then we have

$$\phi[n] \;=\; f(nh) \;=\; \int_{-\infty}^{\infty} \tilde{f}(s) e^{2\pi i s n h}\, ds, \tag{B.10}$$

$$\phi[n] \;=\; \int_0^l \tilde{\phi}(s) e^{2\pi i s n/l}\, ds. \tag{B.11}$$

By breaking up the domain of integration in eq. (B.10) we obtain

$$\phi[n] \;=\; \sum_{m=-\infty}^{\infty} \int_{ml}^{(m+1)l} \tilde{f}(s) e^{2\pi i s n h}\, ds \tag{B.12}$$

$$\;=\; \sum_{m=-\infty}^{\infty} \int_0^l \tilde{f}(s' + ml) e^{2\pi i n h (s'+ml)}\, ds', \tag{B.13}$$

using the change of variable $s' = s - ml$. Now set $hl = 1$ and use $e^{2\pi i n m} = 1$ for $n,\ m$ integers to obtain

$$\phi[n] \;=\; \int_0^l \Big(\sum_{m=-\infty}^{\infty} \tilde{f}(s + ml) \Big) e^{2\pi i s n/l}\, ds, \tag{B.14}$$

which implies that

$$\tilde{\phi}(s) \;=\; \sum_{m=-\infty}^{\infty} \tilde{f}(s + ml), \tag{B.15}$$

with $l = 1/h$. Alternatively setting $l = 1$ one obtains $\tilde{\phi}(s) = \frac{1}{h}\sum_{m=-\infty}^{\infty} \tilde{f}(\frac{s+m}{h})$. Similarly if f is defined on $\mathbb{T}_l$ and $\phi[n] = f(\frac{nl}{N})$ is obtained by sampling then

$$\tilde{\phi}[n] \;=\; \sum_{m=-\infty}^{\infty} \tilde{f}[n + mN]. \tag{B.16}$$

Thus we see that sampling in x-space causes periodization in Fourier space.

Now consider the periodization of a function $f(x)$ with $x \in \mathbb{R}$ to give the l-periodic function $g(x) \triangleq \sum_{m=-\infty}^{\infty} f(x - ml)$. Let $\tilde{g}[k]$ be the Fourier coefficients of $g(x)$. We obtain

$$\tilde{g}[k] \;=\; \frac{1}{l}\int_0^l g(x) e^{-2\pi i k x/l}\, dx \;=\; \frac{1}{l}\int_0^l \sum_{m=-\infty}^{\infty} f(x - ml) e^{-2\pi i k x/l}\, dx \tag{B.17}$$

$$\;=\; \frac{1}{l}\int_{-\infty}^{\infty} f(x) e^{-2\pi i k x/l}\, dx \;=\; \frac{1}{l}\tilde{f}\Big(\frac{k}{l}\Big), \tag{B.18}$$

assuming that $f(x)$ is sufficiently well-behaved that the summation and integration operations can be exchanged. A similar relationship can be obtained for the periodization of a function defined on $\mathbb{Z}$. Thus we see that periodization in x-space gives rise to sampling in Fourier space.

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

B.2 Continuous-time Gaussian Markov Processes

We first consider continuous-time Gaussian Markov processes on the real line, and then relate the covariance function obtained to that for the stationary solution of the SDE on the circle. Our treatment of continuous-time GMPs on $\mathbb{R}$ follows Papoulis [1991, ch. 10].

B.2.1 Continuous-time GMPs on $\mathbb{R}$

We wish to find the power spectrum and covariance function for the stationary process corresponding to the SDE given by eq. (B.2). Recall that the covariance function of a stationary process $k(t)$ and the power spectrum $S(s)$ form a Fourier transform pair.

The Fourier transform of the stochastic process $X(t)$ is a stochastic process $\tilde{X}(s)$ given by

$$\tilde{X}(s) = \int_{-\infty}^{\infty} X(t)e^{-2\pi ist}\,dt, \qquad X(t) = \int_{-\infty}^{\infty} \tilde{X}(s)e^{2\pi ist}\,ds, \tag{B.19}$$

where the integrals are interpreted as a mean-square limit. Let * denote complex conjugation and $\langle \ldots \rangle$ denote expectation with respect to the stochastic process. Then for a stationary Gaussian process we have

$$\begin{aligned} \langle \tilde{X}(s_1)\tilde{X}^*(s_2)\rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \langle X(t)X^*(t')\rangle e^{-2\pi is_1t}e^{2\pi is_2t'}\,dt\,dt' & \text{(B.20)}\\ &= \int_{-\infty}^{\infty} dt' e^{-2\pi i(s_1-s_2)t'}\int_{-\infty}^{\infty} d\tau k(\tau)e^{-2\pi is_1\tau} & \text{(B.21)}\\ &= S(s_1)\delta(s_1-s_2), & \text{(B.22)} \end{aligned}$$

using the change of variables $\tau = t - t'$ and the integral representation of the delta function $\int e^{-2\pi ist}dt = \delta(s)$. This shows that $\tilde{X}(s_1)$ and $\tilde{X}(s_2)$ are uncorrelated for $s_1 \neq s_2$, i.e. that the Fourier basis are eigenfunctions of the differential operator. Also from eq. (B.19) we obtain

$$X^{(k)}(t) = \int_{-\infty}^{\infty} (2\pi is)^k \tilde{X}(s)e^{2\pi ist}\,ds. \tag{B.23}$$

Now if we Fourier transform eq. (B.2) we obtain

$$\sum_{k=0}^{p} a_k(2\pi is)^k\tilde{X}(s) = b_0\tilde{Z}(s), \tag{B.24}$$

where $\tilde{Z}(s)$ denotes the Fourier transform of the white noise. Taking the product of equation B.24 with its complex conjugate and taking expectations we obtain

$$\left[\sum_{k=0}^{p} a_k(2\pi is_1)^k\right]\left[\sum_{k=0}^{p} a_k(-2\pi is_2)^k\right]\langle\tilde{X}(s_1)\tilde{X}^*(s_2)\rangle = b_0^2\langle\tilde{Z}(s_1)\tilde{Z}^*(s_2)\rangle. \tag{B.25}$$

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

Let $A(z) = \sum_{k=0}^{p} a_k z^k$. Then using eq. (B.22) and the fact that the power spectrum of white noise is 1, we obtain

$$S_{\mathbb{R}}(s) \;=\; \frac{b_0^2}{|A(2\pi i s)|^2}. \tag{B.26}$$

Note that the denominator is a polynomial of order p in s^2. The relationship of stationary solutions of pth-order SDEs to rational spectral densities can be traced back at least as far as Doob [1944].

Above we have assumed that the process is stationary. However, this depends on the coefficients $a_0, \ldots, a_p$. To analyze this issue we assume a solution of the form $X_t \propto e^{\lambda t}$ when the driving term $b_0 = 0$. This leads to the condition for stationarity that the roots of the polynomial $\sum_{k=0}^{p} a_k \lambda^k$ must lie in the left half plane [Arató, 1982, p. 127].

AR(1) process

Example: AR(1) process. In this case we have the SDE

$$X'(t) + a_0 X(t) \;=\; b_0 Z(t), \tag{B.27}$$

where $a_0 > 0$ for stationarity. This gives rise to the power spectrum

$$S(s) \;=\; \frac{b_0^2}{(2\pi i s + a_0)(-2\pi i s + a_0)} \;=\; \frac{b_0^2}{(2\pi s)^2 + a_0^2}. \tag{B.28}$$

Taking the Fourier transform we obtain

$$k(t) \;=\; \frac{b_0^2}{2a_0} e^{-a_0|t|}. \tag{B.29}$$

This process is known as the Ornstein-Uhlenbeck (OU) process [Uhlenbeck and Ornstein, 1930] and was introduced as a mathematical model of the velocity of a particle undergoing Brownian motion. It can be shown that the OU process is the unique stationary first-order Gaussian Markov process.

AR(p) process

Example: AR(p) process. In general the covariance transform corresponding to the power spectrum $S(s) = ([\sum_{k=0}^{p} a_k (2\pi i s)^k][\sum_{k=0}^{p} a_k(-2\pi i s)^k])^{-1}$ can be quite complicated. For example, Papoulis [1991, p. 326] gives three forms of the covariance function for the AR(2) process depending on whether $a_1^2 - 4a_0$ is greater than, equal to or less than 0. However, if the coefficients $a_0, a_1, \ldots, a_p$ are chosen in a particular way then one can obtain

$$S(s) \;=\; \frac{1}{(4\pi^2 s^2 + \alpha^2)^p} \tag{B.30}$$

for some α. It can be shown [Stein, 1999, p. 31] that the corresponding covariance function is of the form $\sum_{k=0}^{p-1} \beta_k |t|^k e^{-\alpha|t|}$ for some coefficients $\beta_0, \ldots, \beta_{p-1}$. For $p = 1$ we have already seen that $k(t) = \frac{1}{2\alpha} e^{-\alpha|t|}$ for the OU process. For $p = 2$ we obtain $k(t) = \frac{1}{4\alpha^3} e^{-\alpha|t|}(1+\alpha|t|)$. These are special cases of the Matérn class of covariance functions described in section 4.2.1.

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006,
ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

Wiener process

Example: Wiener process. Although our derivations have focussed on *stationary* Gaussian Markov processes, there are also several important non-stationary processes. One of the most important is the Wiener process that satisfies the SDE $X'(t) = Z(t)$ for $t \geq 0$ with the initial condition $X(0) = 0$. This process has covariance function $k(t, s) = \min(t, s)$. An interesting variant of the Wiener process known as the Brownian bridge (or *tied-down* Wiener process) is obtained by conditioning on the Wiener process passing through $X(1) = 0$. This has covariance $k(t, s) = \min(t, s) - st$ for $0 \leq s, t \leq 1$. See e.g. Grimmett and Stirzaker [1992] for further information on these processes.

Markov processes derived from SDEs of order p are $p-1$ times MS differentiable. This is easy to see heuristically from eq. (B.2); given that a process gets rougher the more times it is differentiated, eq. (B.2) tells us that $X^{(p)}(t)$ is like the white noise process, i.e. not MS continuous. So, for example, the OU process (and also the Wiener process) are MS continuous but not MS differentiable.

B.2.2 The Solution of the Corresponding SDE on the Circle

The analogous analysis to that on the real line is carried out on $\mathbb{T}_l$ using

$$X(t) \;=\; \sum_{n=-\infty}^{\infty} \tilde{X}[n] e^{2\pi i n t/l}, \qquad \tilde{X}[n] \;=\; \frac{1}{l}\int_0^l X(t) e^{-2\pi i n t/l} dt. \tag{B.31}$$

As $X(t)$ is assumed stationary we obtain an analogous result to eq. (B.22), i.e. that the Fourier coefficients are independent

$$\langle \tilde{X}[m]\tilde{X}^*[n]\rangle \;=\; \begin{cases} S[n] & \text{if } m = n \\ 0 & \text{otherwise.} \end{cases} \tag{B.32}$$

Similarly, the covariance function on the cirle is given by $k(t-s) = \langle X(t)X^*(s)\rangle = \sum_{n=-\infty}^{\infty} S[n]e^{2\pi i n(t-s)/l}$. Let $\omega_l = 2\pi/l$. Then plugging in the expression $X^{(k)}(t) = \sum_{n=-\infty}^{\infty}(in\omega_l)^k \tilde{X}[n]e^{in\omega_l t}$ into the SDE eq. (B.2) and equating terms in $[n]$ we obtain

$$\sum_{k=0}^{p} a_k (in\omega_l)^k \tilde{X}[n] \;=\; b_0 \tilde{Z}[n]. \tag{B.33}$$

As in the real-line case we form the product of equation B.33 with its complex conjugate and take expectations to give

$$S_{\mathbb{T}}[n] \;=\; \frac{b_0^2}{|A(in\omega_l)|^2}. \tag{B.34}$$

Note that $S_{\mathbb{T}}[n]$ is equal to $S_{\mathbb{R}}\left(\frac{n}{l}\right)$, i.e. that it is a sampling of $S_{\mathbb{R}}$ at intervals $1/l$, where $S_{\mathbb{R}}(s)$ is the power spectrum of the continuous process on the real line given in equation B.26. Let $k_{\mathbb{T}}(h)$ denote the covariance function on the

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

circle and $k_{\mathbb{R}}(h)$ denote the covariance function on the real line for the SDE. Then using eq. (B.15) we find that

$$k_{\mathbb{T}}(t) \;=\; \sum_{m=-\infty}^{\infty} k_{\mathbb{R}}(t - ml). \tag{B.35}$$

1st order SDE

Example: 1st-order SDE. On $\mathbb{R}$ for the OU process we have $k_{\mathbb{R}}(t) = \frac{b_0^2}{2a_0} e^{-a_0|t|}$. By summing the series (two geometric progressions) we obtain

$$k_{\mathbb{T}}(t) \;=\; \frac{b_0^2}{2a_0(1 - e^{-a_0 l})}\Big(e^{-a_0|t|} + e^{-a_0(l-|t|)}\Big) = \frac{b_0^2}{2a_0}\frac{\cosh[a_0(\frac{l}{2} - |t|)]}{\sinh(\frac{a_0 l}{2})} \tag{B.36}$$

for $-l \leq t \leq l$. Eq. (B.36) is also given (up to scaling factors) in Grenander et al. [1991, eq. 2.15], where it is obtained by a limiting argument from the discrete-time GMP on $\mathbb{P}_n$, see section B.3.2.

B.3 Discrete-time Gaussian Markov Processes

We first consider discrete-time Gaussian Markov processes on $\mathbb{Z}$, and then relate the covariance function obtained to that of the stationary solution of the difference equation on $\mathbb{P}_N$. Chatfield [1989] and Diggle [1990] provide good coverage of discrete-time ARMA models on $\mathbb{Z}$.

B.3.1 Discrete-time GMPs on $\mathbb{Z}$

Assuming that the process is stationary the covariance function $k[i]$ denotes $\langle X_t X_{t+i}\rangle \;\; \forall t \in \mathbb{Z}$. (Note that because of stationarity $k[i] = k[-i]$.)

We first use a Fourier approach to derive the power spectrum and hence the covariance function of the AR(p) process. Defining $a_0 = -1$, we can rewrite eq. (B.1) as $\sum_{k=0}^{p} a_k X_{t-k} + b_0 Z_t = 0$. The Fourier pair for $X[t]$ is

$$X[t] \;=\; \int_0^l \tilde{X}(s) e^{2\pi i s t/l}\, ds, \qquad \tilde{X}(s) \;=\; \frac{1}{l}\sum_{t=-\infty}^{\infty} X[t] e^{-2\pi i s t/l}. \tag{B.37}$$

Plugging this into $\sum_{k=0}^{p} a_k X_{t-k} + b_0 Z_t = 0$ we obtain

$$\tilde{X}(s)\Big(\sum_{k=0}^{p} a_k e^{-i\omega_l s k}\Big) + b_0 \tilde{Z}(s) \;=\; 0, \tag{B.38}$$

where $\omega_l = 2\pi/l$. As above, taking the product of eq. (B.38) with its complex conjugate and taking expectations we obtain

$$S_{\mathbb{Z}}(s) \;=\; \frac{b_0^2}{|A(e^{i\omega_l s})|^2}. \tag{B.39}$$

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

Above we have assumed that the process is stationary. However, this depends on the coefficients $a_0, \ldots, a_p$. To analyze this issue we assume a solution of the form $X_t \propto z^t$ when the driving term $b_0 = 0$. This leads to the condition for stationarity that the roots of the polynomial $\sum_{k=0}^{p} a_k z^{p-k}$ must lie *inside* the unit circle. See Hannan [1970, Theorem 5, p. 19] for further details.

As well as deriving the covariance function from the Fourier transform of the power spectrum it can also be obtained by solving a set of linear equations. Our first observation is that X_s is independent of Z_t for $s < t$. Multiplying equation B.1 through by Z_t and taking expectations, we obtain $\langle X_t Z_t \rangle = b_0$ and $\langle X_{t-i} Z_t \rangle = 0$ for $i > 0$. By multiplying equation B.1 through by X_{t-j} for $j = 0, 1, \ldots$ and taking expectations we obtain the *Yule-Walker* equations

Yule-Walker equations

$$k[0] = \sum_{i=1}^{p} a_i k[i] + b_0^2 \tag{B.40}$$

$$k[j] = \sum_{i=1}^{p} a_i k[j-i] \qquad \forall j > 0. \tag{B.41}$$

The first $p+1$ of these equations form a linear system that can be used to solve for $k[0], \ldots, k[p]$ in terms of b_0 and $a_1, \ldots, a_p$, and eq. (B.41) can be used to obtain $k[j]$ for $j > p$ recursively.

Example: AR(1) process. The simplest example of an AR process is the AR(1) process defined as $X_t = a_1 X_{t-1} + b_0 Z_t$. This gives rise to the Yule-Walker equations

AR(1) process

$$k[0] - a_1 k[1] = b_0^2, \quad \text{and} \quad k[1] - a_1 k[0] = 0. \tag{B.42}$$

The linear system for $k[0], k[1]$ can easily be solved to give $k[j] = a_1^{|j|} \sigma_X^2$, where $\sigma_X^2 = b_0^2/(1 - a_1^2)$ is the variance of the process. Notice that for the process to be stationary we require $|a_1| < 1$. The corresponding power spectrum obtained from equation B.39 is

$$S(s) = \frac{b_0^2}{1 - 2a_1 \cos(\omega_l s) + a_1^2}. \tag{B.43}$$

Similarly to the continuous case, the covariance function for the discrete-time AR(2) process has three different forms depending on $a_1^2 + 4a_2$. These are described in Diggle [1990, Example 3.6].

B.3.2 The Solution of the Corresponding Difference Equation on $\mathbb{P}_N$

We now consider variables $\mathbf{X} = X_0, X_1, \ldots, X_{N-1}$ arranged around the circle with $N \geq p$. By appropriately modifying eq. (B.1) we obtain

$$X_t = \sum_{k=1}^{p} a_k X_{\text{mod}(t-k,N)} + b_0 Z_t. \tag{B.44}$$

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

The Z_t's are i.i.d. and $\sim \mathcal{N}(0,1)$. Thus $\mathbf{Z} = Z_0, Z_1, \ldots, Z_{N-1}$ has density $p(\mathbf{Z}) \propto \exp -\frac{1}{2}\sum_{t=0}^{N-1} Z_t^2$. Equation (B.44) shows that $\mathbf{X}$ and $\mathbf{Z}$ are related by a linear transformation and thus

$$p(\mathbf{X}) \ \propto \ \exp\Big(-\frac{1}{2b_0^2}\sum_{t=0}^{N-1}\Big(X_t - \sum_{k=1}^{p} a_k X_{\mathrm{mod}(t-k,N)}\Big)^2\Big). \qquad \text{(B.45)}$$

This is an N-dimensional multivariate Gaussian. For an AR(p) process the inverse covariance matrix has a circulant structure [Davis, 1979] consisting of a diagonal band $(2p+1)$ entries wide and appropriate circulant entries in the corners. Thus $p(X_t|\mathbf{X}\setminus X_t) = p(X_t|X_{\mathrm{mod}(t-1,N)}, \ldots, X_{\mathrm{mod}(t-p,N)}, X_{\mathrm{mod}(t+1,N)}, \ldots, X_{\mathrm{mod}(t+p,N)})$, which Geman and Geman [1984] call the "two-sided" Markov property. Notice that it is the zeros in the *inverse* covariance matrix that indicate the conditional independence structure; see also section B.5.

The properties of eq. (B.44) have been studied by a number of authors, e.g. Whittle [1963] (under the name of circulant processes), Kashyap and Chellappa [1981] (under the name of circular autoregressive models) and Grenander et al. [1991] (as cyclic Markov process).

As above, we define the Fourier transform pair

$$X[n] \ = \ \sum_{m=0}^{N-1} \tilde{X}[m] e^{2\pi i nm/N}, \qquad \tilde{X}[m] \ = \ \frac{1}{N}\sum_{n=0}^{N-1} X[n] e^{-2\pi i nm/N}. \qquad \text{(B.46)}$$

By similar arguments to those above we obtain

$$\sum_{k=0}^{p} a_k \tilde{X}[m](e^{2\pi i m/N})^k + b_0 \tilde{Z}[m] \ = \ 0, \qquad \text{(B.47)}$$

where $a_0 = -1$, and thus

$$S_{\mathbb{P}}[m] \ = \ \frac{b_0^2}{|A(e^{2\pi i m/N})|^2}. \qquad \text{(B.48)}$$

As in the continuous-time case, we see that $S_{\mathbb{P}}[m]$ is obtained by sampling the power spectrum of the corresponding process on the line, so that $S_{\mathbb{P}}[m] = S_{\mathbb{Z}}\left(\frac{ml}{N}\right)$. Thus using eq. (B.16) we have

$$k_{\mathbb{P}}[n] \ = \ \sum_{m=-\infty}^{\infty} k_{\mathbb{Z}}[n + mN]. \qquad \text{(B.49)}$$

AR(1) process

Example: AR(1) process. For this process $X_t = a_1 X_{\mathrm{mod}(t-1,n)} + b_0 Z_t$, the diagonal entries in the inverse covariance are $(1+a_1^2)/b_0^2$ and the non-zero off-diagonal entries are $-a_1/b_0^2$.

By summing the covariance function $k_{\mathbb{Z}}[n] = \sigma_X^2 a_1^{|n|}$ we obtain

$$k_{\mathbb{P}}[n] \ = \ \frac{\sigma_X^2}{(1-a_1^N)}(a_1^{|n|} + a_1^{|N-n|}) \qquad n = 0, \ldots, N-1. \qquad \text{(B.50)}$$

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

We now illustrate this result for $N = 3$. In this case the covariance matrix has diagonal entries of $\frac{\sigma_X^2}{(1-a_1^3)}(1 + a_1^3)$ and off-diagonal entries of $\frac{\sigma_X^2}{(1-a_1^3)}(a_1 + a_1^2)$. The inverse covariance matrix has the structure described above. Multiplying these two matrices together we do indeed obtain the identity matrix.

B.4 The Relationship Between Discrete-time and Sampled Continuous-time GMPs

We now consider the relationship between continuous-time and discrete-time GMPs. In particular we ask the question, is a regular sampling of a continuous-time AR(p) process a discrete-time AR(p) process? It turns out that the answer will, in general, be negative. First we define a generalization of AR processes known as autoregressive moving-average (ARMA) processes.

ARMA processes The AR(p) process defined above is a special case of the more general ARMA(p, q) process which is defined as

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=0}^{q} b_j Z_{t-j}. \tag{B.51}$$

Observe that the AR(p) process is in fact also an ARMA($p, 0$) process. A spectral analysis of equation B.51 similar to that performed in section B.3.1 gives

$$S(s) = \frac{|B(e^{i\omega_l s})|^2}{|A(e^{i\omega_l s})|^2}, \tag{B.52}$$

where $B(z) = \sum_{j=0}^{q} b_j z^j$. In continuous time a process with a rational spectral density of the form

$$S(s) = \frac{|B(2\pi i s)|^2}{|A(2\pi i s)|^2} \tag{B.53}$$

is known as a ARMA(p, q) process. For this to define a valid covariance function we require $q < p$ as $k(0) = \int S(s)ds < \infty$.

Discrete-time observation of a continuous-time process Let $X(t)$ be a continuous-time process having covariance function $k(t)$ and power spectrum $S(s)$. Let X_h be the discrete-time process obtained by sampling $X(t)$ at interval h, so that $X_h[n] = X(nh)$ for $n \in \mathbb{Z}$. Clearly the covariance function of this process is given by $k_h[n] = k(nh)$. By eq. (B.15) this means that

$$S_h(s) = \sum_{m=-\infty}^{\infty} S(s + \frac{m}{h}) \tag{B.54}$$

where $S_h(s)$ is defined using $l = 1/h$.

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

Theorem B.1 *Let X be a continuous-time stationary Gaussian process and X_h be the discretization of this process. If X is an ARMA process then X_h is also an ARMA process. However, if X is an AR process then X_h is not necessarily an AR process.* □

The proof is given in Ihara [1993, Theorem 2.7.1]. It is easy to see using the covariance functions given in sections B.2.1 and B.3.1 that the discretization of a continuous-time AR(1) process is indeed a discrete-time AR(1) process. However, Ihara shows that, in general, the discretization of a continuous-time AR(2) process is not a discrete-time AR(2) process.

B.5 Markov Processes in Higher Dimensions

We have concentrated above on the case where t is one-dimensional. In higher dimensions it is interesting to ask how the Markov property might be generalized. Let ∂S be an infinitely differentiable closed surface separating $\mathbb{R}^D$ into a bounded part S^- and an unbounded part S^+. Loosely speaking[2] a random field $X(\mathbf{t})$ is said to be *quasi-Markovian* if $X(\mathbf{t})$ for $\mathbf{t} \in S^-$ and $X(\mathbf{u})$ for $\mathbf{u} \in S^+$ are independent given $X(\mathbf{s})$ for $\mathbf{s} \in \partial S$. Wong [1971] showed that the only isotropic quasi-Markov Gaussian field with a continuous covariance function is the degenerate case $X(\mathbf{t}) = X(\mathbf{0})$, where $X(\mathbf{0})$ is a Gaussian variate. However, if instead of conditioning on the values that the field takes on in ∂S, one conditions on a somewhat larger set, then Gaussian random fields with non-trivial Markov-type structure can be obtained. For example, random fields with an inverse power spectrum of the form $\sum_{\mathbf{k}} a_{k_1,\ldots,k_D} s_1^{k_1} \cdots s_d^{k_d}$ with $\mathbf{k}^\top \mathbf{1} = \sum_{j=1}^D k_j \leq 2p$ and $C(\mathbf{s} \cdot \mathbf{s})^p \leq \left| \sum_{\mathbf{k}^\top \mathbf{1} = 2p} a_{k_1,\ldots,k_D} s_1^{k_1} \cdots s_D^{k_d} \right|$ for some $C > 0$ are said to be *pseudo-Markovian* of order p. For example, the D-dimensional tensor-product of the OU process $k(\mathbf{t}) = \prod_{i=1}^D e^{-\alpha_i |t_i|}$ is pseudo-Markovian of order D. For further discussion of Markov properties of random fields see the Appendix in Adler [1981].

If instead of $\mathbb{R}^D$ we wish to define a Markov random field (MRF) on a graphical structure (for example the lattice $\mathbb{Z}^D$) things become more straightforward. We follow the presentation in Jordan [2005]. Let $G = (X, E)$ be a graph where X is a set of nodes that are in one-to-one correspondence with a set of random variables, and E be the set of undirected edges of the graph. Let $\mathcal{C}$ be the set of all maximal cliques of G. A potential function $\psi_C(\mathbf{x}_C)$ is a function on the possible realizations $\mathbf{x}_C$ of the maximal clique $\mathbf{X}_C$. Potential functions are assumed to be (strictly) positive, real-valued functions. The probability distribution $p(\mathbf{x})$ corresponding to the Markov random field is given by

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(\mathbf{x}_C), \tag{B.55}$$

where Z is a normalization factor (known in statistical physics as the partition function) obtained by summing/integrating $\prod_{C \in \mathcal{C}} \psi_C(\mathbf{x}_C)$ over all possible as-

[2]For a precise formulation of this definition involving σ-fields see Adler [1981, p. 256].

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006,
ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

signments of values to the nodes X. Under this definition it is easy to show that a local Markov property holds, i.e. that for any variable x the conditional distribution of x given all other variables in X depends only on those variables that are neighbours of x. A useful reference on Markov random fields is Winkler [1995].

A simple example of a Gaussian Markov random field has the form

$$p(\mathbf{x}) \;\propto\; \exp\Big(-\alpha_1 \sum_i x_i^2 - \alpha_2 \sum_{i,j:j\in N(i)} (x_i - x_j)^2\Big), \tag{B.56}$$

where $N(i)$ denotes the set of neighbours of node x_i and $\alpha_1,\ \alpha_2 > 0$. On $\mathbb{Z}^2$ one might choose a four-connected neighbourhood, i.e. those nodes to the north, south, east and west of a given node.

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006,
ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml