16 Linear and Nonlinear Time Series

A Kalman filter, or a Hidden Markov Model, starts with some notion of the
of a system and then seeks to match it to observations. As powerful as these
what if you're given a signal without *a priori* insight into the system that pro
What if your goal is to learn more about the nature of the system, not just wha
is? This is the domain of *time series analysis*. The field is as broad as time i
defined not by any particular tools (it draws on many of the preceeding chap
rather by the intent of their use.

Time series problems arise in almost all disciplines, ranging from studying
in currency exchange rates to variations in heart-rates. Wherever they occur
three recurring tasks:

- *Characterize:* What kind of system produced the signal? How many
freedom does it have? How random is it? Is it linear? How does noise
the system?
- *Forecast:* Based on an estimate of the current state, what will the system
- *Model:* What are the governing equations for the system? What is th
term behavior?

These are closely related but not identical. For example, a model with good
properties may not be the best way to make short-term forecasts and *vice ve*
although it's possible to characterize a system without explicitly writing down
some of the most powerful characterization techniques are based on first building

This chapter will assume that the analyst is an observer, not a participant
modeling comes manipulation. If it is possible to influence a system then th
of descriptions can be used to choose informative inputs (by selecting them
model uncertainty is large), and to drive the system to a desired state (by reve
model to predict inputs based on outputs, the domain of *control theory* [Do
1992, Auerbach *et al.*, 1992]).

Time series originally were analyzed, not surprisingly, in the time domain. C
zation consisted of looking at the series, and the only kind of forecasting or mod
simple extrapolation. A major step was Yule's 1927 analysis of the sunspot cy
1927]. This was perhaps the first time that a model with internal degrees of
(what we would now call a linear autoregressive model) was inferred from meas
of an external observable (the sunspot series). This rapidly bloomed into the
linear time series, which is mature, successful, ubiquitous, and applicable only
systems. It arises in two very different limits: deterministic systems that are so si

can be described by linear governing equations, or systems which are so stochastic that their deviation from ideal randomness is governed by linear random variable equations. In between these two extremes lies the rest of the world, for which nonlinearity does matter. The theories of nonlinear differential equations or stochastic processes in general have no general results, but rather there are many particular tractable cases. However, there is a powerful theory emerging for the characterization and modeling of nonlinear systems without making any linear assumptions. This chapter starts with the linear canon and closes with these newer ideas.

16.1 LINEAR TIME SERIES

The most general linear system produces an output y that is a linear function of external inputs x (sometimes called *innovations*) and its previous outputs:

$$y_t = a_t + \underbrace{\sum_{m=1}^{M} b_m y_{t-m}}_{\text{AR, IIR}} + \underbrace{\sum_{n=0}^{N} c_n x_{t-n}}_{\text{MA, FIR}} \tag{16.1}$$

Typically the a_t term is nonzero only for an initial transient, which imposes the initial conditions on the system. Depending on the side of campus that you are on, the two parts of this equation are called:

- Statistics
  - *Auto-Regressive* (AR): The output is a linear regression of its M previous values.
  - *Moving Average* (MA): The output is an N-point moving average of the input.
  - Taken together, they define an $ARMA(M,N)$ model.
- Engineering
  - *Infinite Impulse Response* (IIR): The output can continue after the input stops.
  - *Finite Impulse Response* (FIR): The output stops after the input stops.

For the statistician these are random variables, while the engineer usually tries to make sure that they are not. *Trending* in a nonstationary signal can be removed by differencing it to some order before model building (first-order time differences remove a linear drift, second-order removes a polynomial trend, and so forth), giving an $ARIMA$ (Auto-Regressive Integrated Moving Average) model.

The z-transform of the output

$$Y(z) \equiv \sum_{n=-\infty}^{n=\infty} y_n z^n \tag{16.2}$$

provides a complete analysis of the system (Chapter 2). Since convolution in the time domain equals multiplication in the z domain, the z-transform can easily be solved:

$$y_t = a_t + \sum_{m=1}^{M} b_m y_{t-m} + \sum_{n=0}^{N} c_n x_{t-n}$$

Given the c's we can calculate the autocorrelation function, or this relationship can be inverted to find a set of c's to match a given autocorrelation function.

Similarly, multiplying both sides of an AR model by $y_{t-\tau}$ and averaging gives

$$\langle y_t y_{t-\tau} \rangle = \sum_{m=1}^{M} b_m \langle y_{t-m} y_{t-\tau} \rangle \quad , \tag{16.8}$$

and then after normalizing by the variance

$$\kappa_\tau = \sum_{m=1}^{M} b_m \kappa_{\tau-m} \quad . \tag{16.9}$$

Unlike the MA case the autocorrelation function need not vanish after M steps. This linear set of equations, called the *Yule–Walker* equations, can be inverted to relate the AR coefficients to the autocorrelation function. *Levinson–Durbin recursion* is an efficient way to do this [Levinson, 1947, Durbin, 1960].

Unlike the simplicity of AR and MA models there is not a unique algorithm to find the best ARMA model to describe a data set (but there is a lot of heated debate about how to do it). The *Box–Jenkins* procedure is a popular recursive solution [Box *et al.*, 1994]. It's always possible to trade off more or less of M versus N in selecting the order of an ARMA model; the *Akaike Information Criteria* (*AIC*) and its Bayesian cousin the *BIC* do this by assigning an informational cost to the number of parameters, to be minimized along with the model error [Akaike, 1979].

16.2 THE BREAKDOWN OF LINEAR SYSTEMS THEORY

The essence of linear systems theory is expressed by the *Wold Decomposition*: any stochastic process can be separated into the sum of two processes – a deterministic one that is a linear function of its past values, and a stochastic one that is a linear function of previous values of an uncorrelated random variable [Priestley, 1981]. Once these two pieces have been found there is nothing more that can be said about the system.

If you limit yourself to linear models, that is. Even simple nonlinearities can be completely misunderstood by a linear analysis. Consider the two simple iterated maps shown in Figure 16.1. The first one,

$$x_{n+1} = 2x_n \text{ (mod 1)} \tag{16.10}$$

is called the *mod map* and it shifts every bit of x (written in a binary fractional expansion) over one place and then discards the most significant bit. This means that the trajectory of the system (for example, which branch it is on) is determined solely by the initial condition. If the initial condition is a real number with digits that appear to be random, say π, then this map will generate a broadband power spectrum. There is a simple truth (equation (16.10)), but a linear model is forced to find the best single straight line to fit what is really two straight lines. The mismatch can only be attributed to stochastic inputs.

The second map

$$x_{n+1} = \lambda x_n (1 - x_n) \tag{16.11}$$

(the logistic map) arises in a variety of systems such as chemical reactions, electrical circuits, hydrodynamic flows, and population dynamics, because of the universal properties

$$\begin{aligned} Y(z) &= A(z) + B(z)Y(z) + C(z)X(z) \\ &= \frac{A(z)}{1-B(z)} + \frac{C(z)}{1-B(z)}X(z) \quad . \end{aligned} \tag{16.3}$$

The z-transform of the output consists of two terms. The first depends on the initial transient, and the second term is equal to the z-transform of the input multiplied by a system *transfer function* that is independent of the input. The output $Y(z)$ consists of a ratio of polynomials $A(z)$ and $C(z)$ divided by $1 - B(z)$ reflecting the system's structure, and a possible non-polynomial part from the input $X(z)$. The numerators, due to the inputs, can have *zeros*, and the denominator, due to the memory of the output, can have *poles*. As we've seen, the location of these poles and zeros determines the system's characteristics (such as stability and oscillation frequencies).

The AR and MA coefficients can be determined from the correlation coefficients. Taking $\langle y \rangle$ to denote the time average expectation value of y (written in the statistics literature as $E[y]$), the *autocorrelation function* is defined to be

$$\begin{aligned} \kappa_\tau &\equiv \frac{\langle (y_t - \langle y_t \rangle)(y_{t-\tau} - \langle y_t \rangle) \rangle}{\langle (y_t - \langle y_t \rangle)^2 \rangle} \\ &= \frac{\langle (y_t - \mu_y)(y_{t-\tau} - \mu_y) \rangle}{\sigma_y^2} \quad . \end{aligned} \tag{16.4}$$

μ_y is the mean and σ_y^2 is the variance. This can also be written as

$$\begin{aligned} \kappa_\tau &= \frac{\langle (y_t - \mu_y)(y_{t-\tau} - \mu_y) \rangle}{\langle (y_t - \mu_y)(y_t - \mu_y) \rangle} \\ &= \frac{\langle y_t y_{t-\tau} \rangle - \mu_y \langle y_{t-\tau} \rangle - \mu_y \langle y_t \rangle + \mu_y \mu_y}{\langle y_t y_t \rangle - \mu_y \langle y_t \rangle - \mu_y \langle y_t \rangle + \mu_y \mu_y} \\ &= \frac{\langle y_t y_{t-\tau} \rangle - \mu_y^2}{\langle y_t y_t \rangle - \mu_y^2} \end{aligned} \tag{16.5}$$

since time averages are independent of the time origin for a stationary process. The autocorrelation function ranges from 1 for perfect correlation between two times, to 0 for uncorrelation, and to -1 for anticorrelation.

For an MA model ($a = b = 0$), if the input is assumed to be zero mean ($\langle x_t \rangle = \mu_x = 0$) then $\mu_y = 0$, and the autocorrelation function becomes

$$\begin{aligned} \kappa_\tau &= \frac{\left\langle \left(\sum_{n=0}^{N} c_n x_{t-n} \right) \left(\sum_{n'=0}^{N} c_{n'} x_{t-\tau-n'} \right) \right\rangle}{\left\langle \left(\sum_{n=0}^{N} c_n x_{t-n} \right) \left(\sum_{n'=0}^{N} c_{n'} x_{t-n'} \right) \right\rangle} \\ &= \frac{\sum_{n=0}^{N} \sum_{n'=0}^{N} c_n c_{n'} \langle x_{t-n} x_{t-\tau-n'} \rangle}{\sum_{n=0}^{N} \sum_{n'=0}^{N} c_n c_{n'} \langle x_{t-n} x_{t-n'} \rangle} \quad . \end{aligned} \tag{16.6}$$

If the input x is an uncorrelated stochastic process ($\langle x_i x_j \rangle = 0$ for $i \neq j$) then the MA coefficients are related to the autocorrelation function by

$$\kappa_\tau = \begin{cases} \dfrac{\sum_{n=\tau}^{N} c_n c_{n-\tau}}{\sum_{n=0}^{N} c_n^2} & (\tau \leq N) \\ 0 & (\tau > N) \end{cases} \quad . \tag{16.7}$$

Given the c's we can calculate the autocorrelation function, or this relationship can be inverted to find a set of c's to match a given autocorrelation function.

Similarly, multiplying both sides of an AR model by $y_{t-\tau}$ and averaging gives

$$\langle y_t y_{t-\tau} \rangle = \sum_{m=1}^{M} b_m \langle y_{t-m} y_{t-\tau} \rangle \quad , \tag{16.8}$$

and then after normalizing by the variance

$$\kappa_\tau = \sum_{m=1}^{M} b_m \kappa_{\tau-m} \quad . \tag{16.9}$$

Unlike the MA case the autocorrelation function need not vanish after M steps. This linear set of equations, called the *Yule–Walker* equations, can be inverted to relate the AR coefficients to the autocorrelation function. *Levinson–Durbin recursion* is an efficient way to do this [Levinson, 1947, Durbin, 1960].

Unlike the simplicity of AR and MA models there is not a unique algorithm to find the best ARMA model to describe a data set (but there is a lot of heated debate about how to do it). The *Box–Jenkins* procedure is a popular recursive solution [Box *et al.*, 1994]. It's always possible to trade off more or less of M versus N in selecting the order of an ARMA model; the *Akaike Information Criteria* (*AIC*) and its Bayesian cousin the *BIC* do this by assigning an informational cost to the number of parameters, to be minimized along with the model error [Akaike, 1979].

16.2 THE BREAKDOWN OF LINEAR SYSTEMS THEORY

The essence of linear systems theory is expressed by the *Wold Decomposition*: any stochastic process can be separated into the sum of two processes – a deterministic one that is a linear function of its past values, and a stochastic one that is a linear function of previous values of an uncorrelated random variable [Priestley, 1981]. Once these two pieces have been found there is nothing more that can be said about the system.

If you limit yourself to linear models, that is. Even simple nonlinearities can be completely misunderstood by a linear analysis. Consider the two simple iterated maps shown in Figure 16.1. The first one,

$$x_{n+1} = 2x_n \pmod 1 \tag{16.10}$$

is called the *mod map* and it shifts every bit of x (written in a binary fractional expansion) over one place and then discards the most significant bit. This means that the trajectory of the system (for example, which branch it is on) is determined solely by the initial condition. If the initial condition is a real number with digits that appear to be random, say π, then this map will generate a broadband power spectrum. There is a simple truth (equation (16.10)), but a linear model is forced to find the best single straight line to fit what is really two straight lines. The mismatch can only be attributed to stochastic inputs.

The second map

$$x_{n+1} = \lambda x_n (1 - x_n) \tag{16.11}$$

(the logistic map) arises in a variety of systems such as chemical reactions, electrical circuits, hydrodynamic flows, and population dynamics, because of the universal properties