The transformation of one-dimensional and two-dimensional autoregressive random fields under coordinate scaling and rotation

by

Ian Douglas Kennedy

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2008

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

A practical problem in computer graphics is that of representing a textured surface at arbitrary scales. I consider the underlying mathematical problem to be that of interpolating autoregressive random fields under arbitrary coordinate transformations. I examine the theoretical basis for the transformations that autoregressive parameters exhibit when the associated stationary random fields are scaled or rotated. The basic result is that the transform takes place in the continuous autocovariance domain, and that the spectral density and associated autoregressive parameters proceed directly from sampling the continuous autocovariance on a transformed grid. I show some real-world applications of these ideas, and explore how they allow us to interpolate into a random field. Along the way, I develop interesting ways to estimate simultaneous autoregressive parameters, to calculate the distorting effects of linear interpolation algorithms, and to interpolate random fields without altering their statistics.

Acknowledgements

I would like to extend my sincerest thanks to David Clausi, for acting both as an instructor and as a thesis supervisor. Dave is a model professor and a great asset to the University of Waterloo. May there be many more faculty members just like him!

Contents

List of Figures

1 Setting the stage

Randomness is essential to our world. It informs the turbulent flows of the world's oceans and atmosphere, which in turn drive climatic cycles. It influences the chromosomal dance that takes place at human conception. In the distant past, small random variations in gaseous density gave rise to the universe's first galaxies. Our daily lives ebb and flow according to the chance encounters and acquaintances that we make. Even our thoughts, which we often view as a bastion of order, have an irrational undercurrent at all times.

So we must celebrate randomness, even when it makes our work challenging. The topic of my thesis is the transformation of autoregressive random fields under scaling and rotation of the coordinate system. The intended application is one in computer graphics: I would like to be able to generate visual textures at arbitrary scales, even when they contain some randomness. My thesis is primarily an exploration of the mathematical underpinnings of the topic, along with illustrative examples, both synthesized and from real life.

So what is visual texture? Along with love and courage, it is easily recognized, can be of crucial importance, but defies definition. Nevertheless, we can say a few things. There must be tonal variation; there must be repetition; the elements that are repeated should not be pre-attentively identifiable as individuals; and texture has a region of support – in other words, it is a property of a region, not a point. Think of scanning a crowd to find a familiar face. The inchoate sea of faces is a texture. To find the person we are searching for, we have to narrow the field of view down to 16-20 people, at which point texture fades and individuals appear.

The Wold decomposition theorem [1] is the starting point. Any one-dimensional (1D) stationary signal divides into a predictable part, and a random process with a moving average (MA) representation. Any two-dimensional (2D) stationary signal divides into a predictable part, an evanescent field (more about that in Section 4), and a moving average process. Rotating and scaling the deterministic part of a signal amounts to transforming a geometric model, which is a solved problem. So I concentrate upon rotating and scaling the random component. In order to accomplish this, we need a parametric model, so that we can generate an autocovariance function under any coordinate transformation. That leads directly to the concept of interpolating a random field, by which I mean the ability to generate field samples in between existing ones, samples which preserve the process statistics.

So why autoregressive fields? An autoregressive (AR) random process is a restriction of a moving average process, so we seem to be losing generality right from the start. I do that because autoregressive models underlie a large number of visual texture analysis techniques, and because behind autoregressive models lies a large body of knowledge extending back many decades.

But to return to our main question: what is texture? It's like explaining where babies come from: we all think we're experts, but it's still hard to describe. We can say a few more things, though. The first-order statistics of a texture derive from illumination and reflectance. Texture intrinsically resides in the second and higher order statistics of an image, in particular in the distribution of information-rich structures such as edges, closures, joins, and crossings. Julesz et al. have demonstrated this quite conclusively [2].

So here is the plan of action: Sections 2-4 define the problem more precisely and describe interesting previous work. Sections 5-8 describe the mathematical nature and

limitations of autoregressive random fields. Sections 9-11 describe how I generate such fields, how I estimate their characteristic parameters, and how accurate those estimates can be. Section 12 develops and illustrates the scaling transform in one dimension. Section 13 shows why an autoregressive model is a poor one if the signal under study is nearly periodic. Section 14 talks about the problems inherent in standard two-dimensional interpolation techniques. Section 15 presents my main argument, namely that rotation and scaling occur in the continuous covariance domain, and do not commute with sampling. Sections 16 and 17 examine the practicality of transform invariants and look at possible basis functions for two-dimensional autocovariances. Sections 18 and 19 look at interpolating one-dimensional random fields. After the wrap-up, appendices A through D collect together various mathematical derivations that the main narrative touches upon.

2 Is transforming a pixel grid always a valid operation?

Well, not really. The whole concept of rotating and scaling a pixel grid presupposes that the random quantity that we are observing has discoverable values "in between" the original pixel grid locations. In some cases, the random quantity of interest is defined at discrete locations and cannot be interpolated. For example, Figure 1 shows a section through the smooth muscle tissue of a monkey's renal calyx. Smooth muscle cells contract when stimulated electrically.

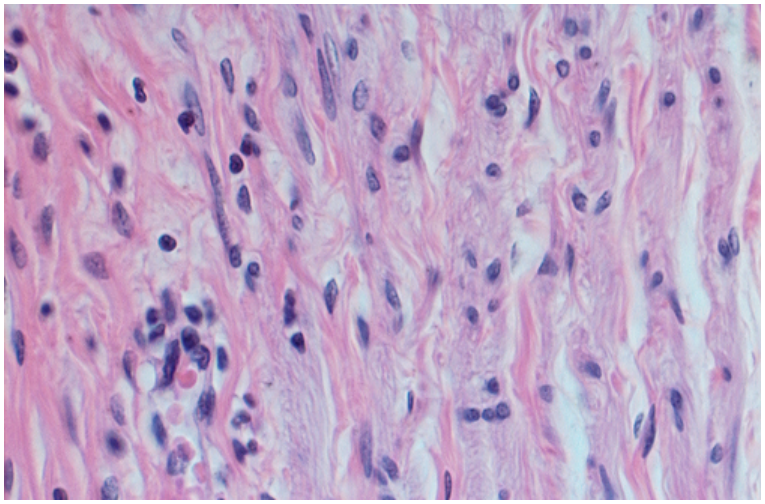

Figure 1: Smooth muscle cells in a monkey's renal calyx. The dark dots are the cell nuclei. Source: Dept of Anatomy and Cell Biology, Indiana University School of Medicine [3]

Suppose we are interested in the pulling strength of each cell under uniform electrical stimulation. There is definitely a random, two-dimensional, highly correlated component to this quantity. Clearly, though, we cannot measure it "in between" existing cells; it exists only for the actual cells. In brief, we can only ask the question about how two-dimensional autoregressive parameters transform under rotation when the resolution at which we are observing the autoregressive field is significantly coarser than the field's intrinsic resolution.

By "intrinsic resolution", I mean the finest possible resolution at which the quantity of interest can be defined. For example, one can consider soil fertility as a correlated random field, and measure it on the scale of kilometers or meters. But we cannot measure soil fertility on a sub-millimeter scale, because the smallest plants that could possibly indicate fertility have millimeter sizes.

Fortunately, when it comes to visual textures, it is nearly always the case that we are imaging a surface at a resolution several orders of magnitude more coarse than its intrinsic resolution. So, in my analysis, I will assume that this is the case.

3 Narrowing the quest: simplifying assumptions

The purpose of the current thesis is to determine how the autoregressive parameters of one-dimensional and two-dimensional random fields transform under coordinate system rotation. This is of interest to me because of an apparent contradiction. In the definition of a discrete Markov random field, the grid on which we make our observations plays a central role. When I observe visual textures, however, I feel intuitively that the position, scale, and orientation of the pixel grid is irrelevant. That leads naturally to the questions of how the autoregressive parameters transform under rotation, and of what invariants – if any! – exist.

So here are the main questions that I want to answer. First, how do autoregressive parameters change when we change scale or rotate the underlying coordinate system? Second, how can we interpolate between the known values of an autoregressive random field without altering its autocovariance structure?

Needless to say, we have to make some simplifying assumptions in order to render the problem tractable. So here they are, in no particular order:

- The random field of interest may be 1D temporal, 1D spatial, 2D spatial, but *not* spatiotemporal.
- The random field is stationary and ergodic. Hence, ensemble averages equal field averages and are constant across the field.
- The field has real, continuous variates and real, continuous autoregressive parameters.
- The field observations are uniformly spaced and noiseless.
- The random fields of interest may be described by a linear ARMA (autoregressive moving average) model, conditional or simultaneous but *not* seasonal (i.e. without a strictly periodic component).
- The model innovations are independent and identically distributed (IID), with a constant variance. I explain this further in Section 5.
- As mentioned in the previous section, the random field can be observed at any scale.

The last two assumptions, in particular, are very restrictive. We exclude heteroskedastic fields, discrete fields, and those random fields that may appear autoregressive at one scale, but not another.

4 Interesting previous work

There is always room in the world for innovation. However, attacking a difficult problem is more bearable when other researchers consider the problem important. Several authors use autoregressive models to analyze and synthesize video sequences, in particular Campbell et al. [4] and Wolfe & Fitzgibbon [5]. Lewis [6] described the general problem of stochastic subdivision (i.e. interpolation into a 2D random field). This technique, sometimes called fractal interpolation, has also been used to generate metallic fracture surfaces [7] and electrical demand sequences [8].

As for the general theory of two-dimensional autoregressive fields, that started with Whittle [9]. Whittle thought that Bessel functions were "natural" basis functions for 2D autocovariances. They suffice for isotropic or nearly-isotropic fields, but for highly directional fields, I find that skew-separable basis functions are a better choice (see Section 16). Of course, one can always use both!

Visual texture analysis has been a topic of interest for more than four decades, and I will not summarize that history here. A very good summary is that of Tuceryan and Jain [10], which divides texture analysis methods into four broad groups: statistical, structural, model-based, and filter-based. As I mentioned earlier, visual textures have a predictable part and a stochastic part. Structural methods focus on the predictable part, while statistical methods focus on the stochastic part. Random mosaic, morphological, and syntactic methods are structural: they regard texture as a collection of placement rules and elements ("textons") to be placed. Models involving Markov random fields, Gibbs random fields, and autoregressive fields all view texture as statistical in nature.

Stochastic models make heavy use of second-order statistics, especially those embodied in the spectral density function. This is a matter of practicality, since texture samples are typically far too small to allow calculation of higher-order image statistics. One observation we can make right away is that second-order statistics alone do not let us distinguish among a deterministic signal with added noise, a constant texton with "jittered" placement rules, and a deterministic signal sampled over a small region; in each case, the spectral density consists of narrow peaks. Figure 2 illustrates this idea.

Several researchers have attempted a direct Wold decomposition of visual texture, in particular Francos et al. [11] and Cadzow et al. [12]. Now, if texture classification were our only goal, then we could ignore the Wold decomposition and focus on structural or stochastic elements exclusively. However, if texture synthesis is our goal – and in my case, it is – then we have to recognize that autoregressive models are poor at synthesizing nearly periodic visual textures. Others have noted this, including Picard [13] and Petrou & Sevilla [14]. In Section 13 I give a mathematical interpretation of this fact.

In two dimensions, the Wold decomposition has three parts (as opposed to two parts in one dimension). A 2D stationary random field divides into a predictable part, a stochastic part with a moving average equivalent, and an evanescent part. An evanescent field is deterministic in one direction and random in the orthogonal direction [11]. Figure 3 illustrates such a field, along with its estimated autocovariance and spectral density. Here, I synthesize the signal as a product of a sinusoid in one direction and an AR(1) series in the orthogonal direction. In my opinion, these functions are not important for 2D visual textures. Even if they were, the diagram clearly shows that the spatial frequency peaks spread out even after extensive averaging, making these field components difficult to identify.

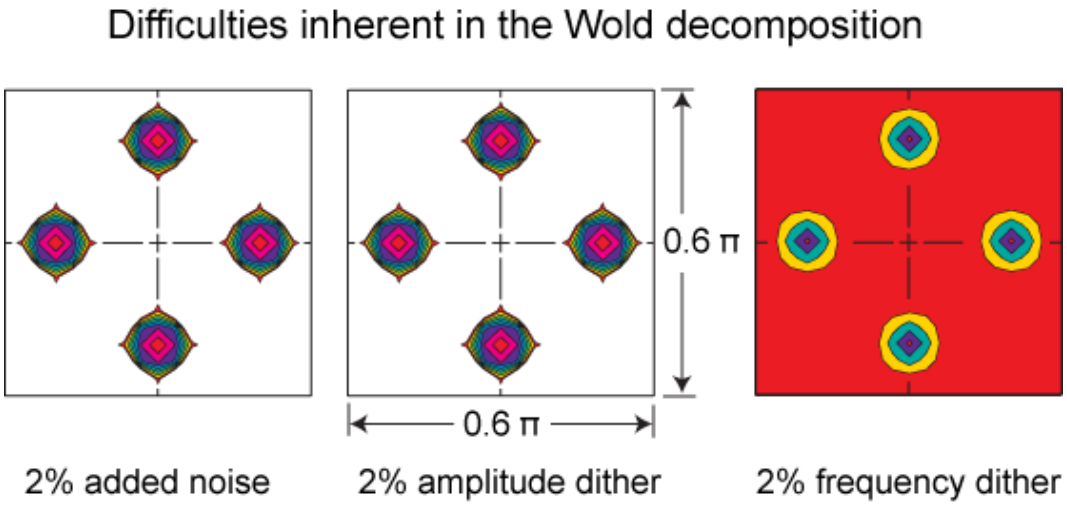

Figure 2: Contour map of the natural logarithm of the estimated spectral density of a nearly periodic signal. Each chart is centered on the origin. Left: a compound 2D sinusoid with 2% added Gaussian noise. Middle: the same 2D sinusoid in which the amplitude has a 2% Gaussian dither. Right: the same 2D sinusoid with a 2% Gaussian frequency dither. In all three cases, the generated field is 1056 x 1056, the contour intervals are one unit apart, and I estimate the spectral density via the Welch method with a Hamming window of size 63 x 63 (see Section 11).

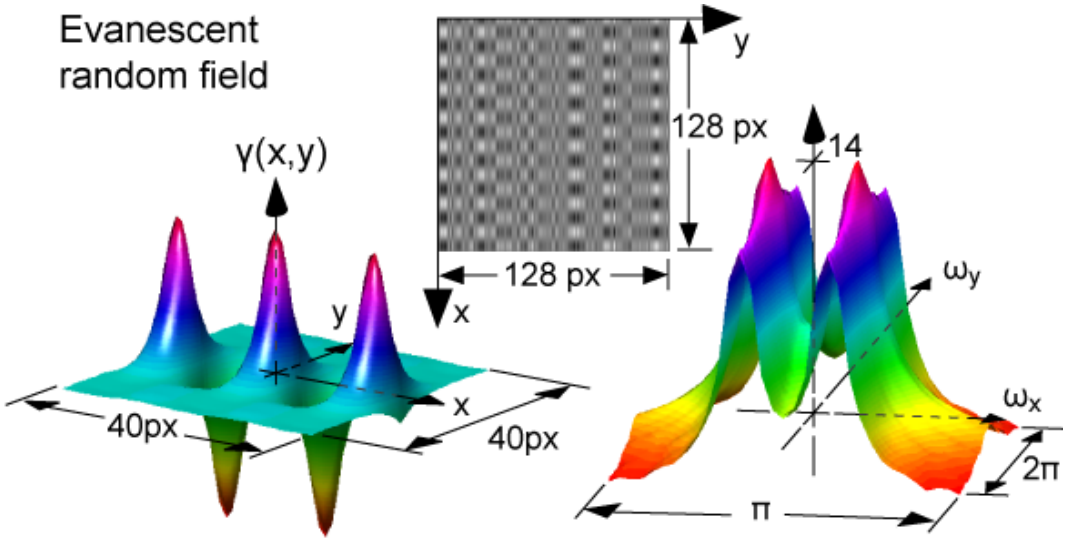

Figure 3: A purely evanescent 2D signal, harmonic in the X direction and AR(1) (with r=0.7) in the Y direction. Left: the estimated autocovariance. Middle: a representative sample. Right: Natural logarithm of the estimated spectral density. I averaged these estimates over 16 instances of a 640 x 640 synthesized random field. I estimate the spectral density via the Welch method with a Hamming window of size 63 x 63 (see Section 11).

The stochastic part of the Wold decomposition admits structural, statistical, and parametric analysis methods. A structural analysis seeks to identify textons and their placement rules, at least one of which must have some randomness. Statistical methods seek to characterize a texture by the statistical distribution of its pixels. Parametric methods posit a model that can reproduce the statistical pixel distribution. In this thesis, I adopt a parametric model, and view the stochastic part of the Wold decomposition as an autoregressive series.

Statistical methods of texture analysis commonly reduce to computing the spectral energy over certain regions of the spatial frequency domain, and using these measures as a texture "signature". Gabor filters, wavelet filters, polarograms, and kriging techniques all fall into this category. I refer to these as "slice and dice" techniques. According to Picard and Elfadel [15], if we assume that third-order statistics are negligible, then Markov random fields, Gibbs random fields, and co-occurrence matrices all reduce to a simultaneous autoregressive model. In other words, they are just different ways of describing a spectral density.

As for fractal generalizations of autoregressive models, they are based on the observation that some naturally occurring textures exhibit self-similarity over a certain range of resolutions. Good introductory treatments are in Hipel & McLeod[16] (for 1D) and Ilow & Leung [17] (for 2D). In independent work, I have found that fractal synthesis gives poor results because the fractional integration stage ruins all the sharp edges that the ARMA model creates. Besides, visual textures are not always self-similar, yet are needed in practical situations such as computer graphics.

The quest for statistical quantities of visual texture that are invariant under rotation is of interest to me because such a quest necessarily involves a notion of how random fields transform under rotation. A good survey of rotation-invariant methods is that of Zhang and Tan [18]. The circular autoregressive model of Kashyap and Khotanzad [19] and the generalized version of Eom [20] are easily shown to be special cases of simultaneous autoregressive fields. Similarly, both the rotation-invariant and multi-resolution simultaneous autoregressive models of Mao and Jain [21] are special cases of simultaneous AR models. I treat the mathematical basis of multi-resolution models in Section 13.

An interesting procedure involving rotational-invariant texture features is described in Lahajnar & Kovacic [22] and Deng & Clausi [23]. The basic idea is to slice and dice the 2D polarogram, and then do a further Fourier transform along the orientation axis. The magnitude of the resulting spectrum is independent of initial texture orientation.

This section's topic is interesting previous work, so I must mention the contributions of Cadzow et al. (for synthesis) and of Cohen, Fan, and Patel (for analysis), since their work aligns most closely with my own view of visual texture. Cadzow et al. [12] view the true visual texture as a filtered excitation with measurement noise, subject to histogram modification:

$$I_{actual} = \phi[H * (E_d + E_r) + N] \tag{4.1}$$

where E_d and E_r are the deterministic and random excitation fields, H is a finite moving average filter, N is additive measurement noise, and ϕ is a histogram shaping operator which models the effect of illumination. In time series analysis, that is known as a transfer function noise model [24].

Cohen, Fan, and Patel [25] attacked this problem in a direct fashion by incorporating both rotation and scaling into their texture classification algorithm, treating visual texture as a Gaussian Markov random field. I comment further upon their work in Section 15.

5 Hello autoregressive world: the conditional discrete AR(1) process

The best way to start our journey is to regard an autoregression as a linear transform between vector spaces. Consider, for example, the simplest 1D discrete conditional autoregressive model, AR(1):

$$[y(t)-\mu] = r\,[y(t-1)-\mu] + e(t) \qquad t>0, \quad y(0), e(0) \text{ given} \tag{5.1}$$

Here, t is an integral index, μ is the mean of the observed random series $y(t)$, r is the sole autoregressive parameter, and $e(t)$ is a stationary, independent and identically distributed (IID) random process with zero mean and constant variance σ^2, commonly known as the innovation sequence. Note that $y(t)$ is an observed series, whereas the innovation sequence is not observed but postulated as part of the model. Let $x(t) = y(t) - \mu$ be the mean-reduced series, which we can split into homogeneous and particular parts $x_h(t)$ and $x_p(t)$. The homogeneous part of $x(t)$ satisfies $x_h(t) - rx_h(t-1) = 0$, and is $x_h(t) = C_0 r^t$. Let $X_p(z)$ and $E(z)$ be the z-transforms of $x_p(t)$ and $e(t)$, respectively. Then the particular part of $x(t)$ is given by

$$\begin{aligned} X_p(z) &= \frac{1}{1-rz^{-1}}E(z) \\ &= (1 + rz^{-1} + r^2z^{-2} + \dots)E(z) \qquad (5.2) \\ \Longrightarrow x_p(t) &= e(t) + re(t-1) + r^2e(t-2) + \cdots + r^te(0) \qquad (5.3) \end{aligned}$$

Imposing the boundary condition at $t = 0$ gives the complete solution

$$x(t) = [x(0) - e(0)]\,r^t + \sum_{j=0}^{t} r^j e(t-j), \quad t \geq 0 \tag{5.4}$$

In order to keep the solution bounded, we must require that $|r| < 1$. Now suppose that $t \gg 0$; then we can neglect the homogeneous part and get

$$\begin{aligned} x(t) &= \sum_{j=0}^{\infty} r^j e(t-j) \qquad (t \gg 0) \qquad (5.5) \\ \Longrightarrow \text{var}\,[x(t)] &= (1 + r^2 + r^4 + \dots)\text{var}\,[e(t)] \\ &= \frac{1}{1-r^2}\sigma^2 \qquad (5.6) \end{aligned}$$

So the marginal series variance is larger than the innovation variance by a factor $1/(1-r^2)$. If we select a group of n successive variates from the series, then the first will have an apparent marginal variance of $\sigma^2/(1-r^2)$. Collecting a mean-reduced sample $\{x(t+1), \dots, x(t+n)\}$ and corresponding innovations $\{e(t+1), \dots, e(t+n)\}$ into vectors $\vec{x}$ and $\vec{e}$ respectively, where $t \gg 0$, their relation is

$$\begin{bmatrix} \sqrt{1-r^2} & 0 & \cdots & 0 & 0 \\ -r & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -r & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \qquad i.e. \quad L\vec{x} = \vec{e} \tag{5.7}$$

So the observation vector $\vec{x}$ and innovation vector $\vec{e}$ are linearly related; in particular, the matrix L defined above is a shear followed by a dilatation.

Now, the innovations are stationary and IID, so the probability density $p(\vec{e})$ should be invariant with respect to all permutations of $\{e_1, \ldots e_n\}$. Thus it must be a function of the quantities $\sum_{j=1}^{n} e_j, \sum_{j=1}^{n} {e_j}^2, \ldots \sum_{j=1}^{n} {e_j}^n$. Suppose that it is a function of $|\vec{e}|^2$ only. In this case,

$$|\vec{e}|^2 = \vec{e}^T\vec{e} = \vec{x}^T(L^TL)\vec{x} = \vec{x}^TQ\vec{x} \tag{5.8}$$

and thus we can observe directly the precision matrix Q, but not the model matrix L. *Any* unitary transform S, applied to $L\vec{x}$, will give equivalent observables:

$$\vec{e} = SL\vec{x} \Longrightarrow \vec{e}^T\vec{e} = \vec{x}^TL^TS^TSL\vec{x} = \vec{x}^TL^TL\vec{x} = \vec{x}^TQ\vec{x} \tag{5.9}$$

This is an important point, so I will illustrate it with some examples. First, consider a set of three successive variates extracted from an AR(1) series in which $t \gg 0$. Their relation to the corresponding innovations is:

$$\begin{bmatrix} \sqrt{1-r^2} & 0 & 0 \\ -r & 1 & 0 \\ 0 & -r & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \qquad i.e. \quad L_f\vec{x} = \vec{e} \tag{5.10}$$

First, perform a reflection in the plane P_1 that contains e_2 and bisects e_1e_3. This takes $\{x_1, x_2, x_3\}$ into $\{x'_1, x'_2, x'_3\}$. Next, perform a further reflection in the plane P_2 that contains the origin, x'_2, and $\frac{1}{2}(x'_1 + x'_3)$. This takes $\{x'_1, x'_2, x'_3\}$ into $\{\bar{x}_1, \bar{x}_2, \bar{x}_3\}$. Figure 4 illustrates these reflections.

The resulting model can be deduced from the diagram by inspection. It is

$$\begin{bmatrix} 1 & -r & 0 \\ 0 & 1 & -r \\ 0 & 0 & \sqrt{1-r^2} \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \qquad i.e. \quad L_b\bar{x} = \vec{e} \tag{5.11}$$

In this new coordinate system, time seems to go backwards! So we will call L_f a causal model, and L_b an anti-causal model. They both lead to the same precision matrix (i.e. ${L_f}^TL_f = {L_b}^TL_b$), so we can't prefer one to the other unless we have additional information. If we know, for example, that $\{x_1, x_2, x_3\}$ is derived from a time series, then we could reject the anti-causal model on physical grounds.

In that last example, the transform that took the causal model into the anti-causal model was the product of two reflections, which is a rotation. Can we rotate half-way and get time to "stand still"? To answer that, consider another three-point conditional AR(1) sample, but this time we'll make a circular approximation:

$$\begin{bmatrix} 1 & 0 & -r \\ -r & 1 & 0 \\ 0 & -r & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = L_f\vec{x} = \vec{e} = L_b\bar{x} = \begin{bmatrix} 1 & -r & 0 \\ 0 & 1 & -r \\ -r & 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} \tag{5.12}$$

The rotation that takes L_f into L_b is clearly around the axis with unit vector $(1/\sqrt{3})(1,1,1)$. The matrix representing a rotation of an angle θ about an axis with unit vector (a_x, a_y, a_z) is [26]:

$$R(\hat{a}, \theta) = \begin{bmatrix} c + (1-c){a_x}^2 & (1-c)a_xa_y - sa_z & (1-c)a_xa_z + sa_y \\ (1-c)a_xa_y + sa_z & c + (1-c){a_y}^2 & (1-c)a_ya_z - sa_x \\ (1-c)a_xa_z - sa_y & (1-c)a_ya_z + sa_x & c + (1-c){a_z}^2 \end{bmatrix} \tag{5.13}$$

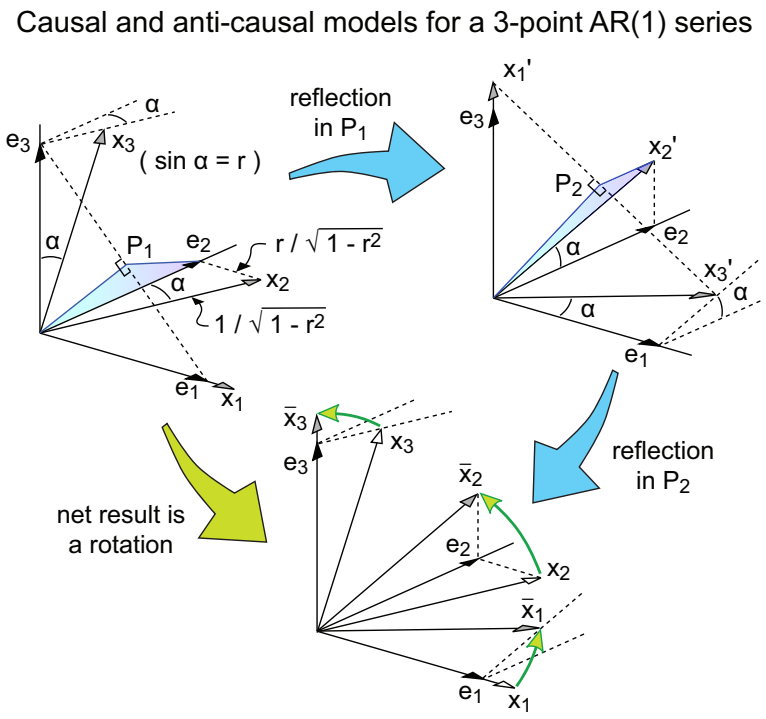

Figure 4: Unitary transformations alter the physical meaning of an autoregressive model, without affecting its observables.

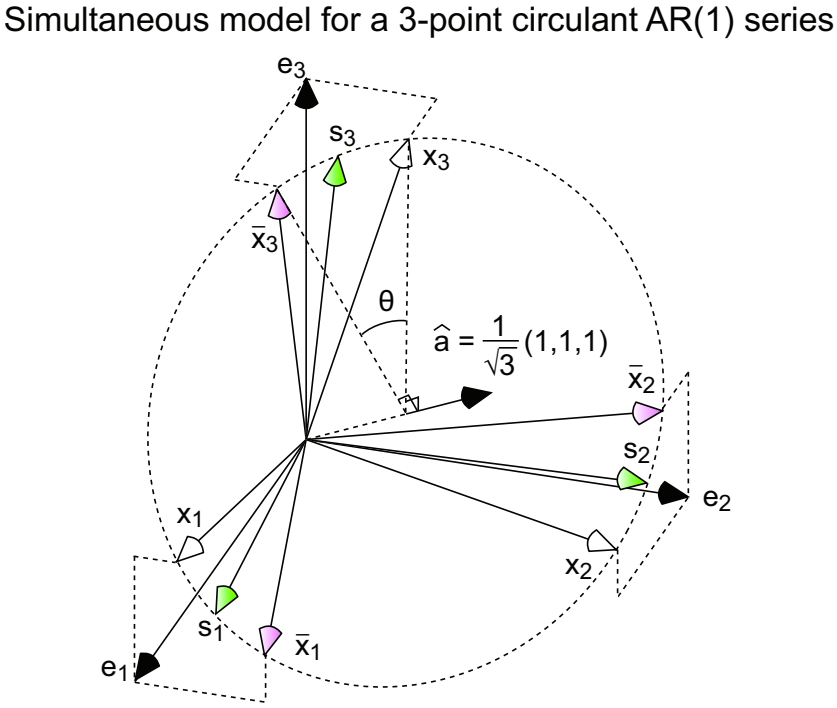

Figure 5: A specific rotation converts a causal autoregressive model into an anticausal one. Rotating half-way about the same axis leads to a simultaneous model.

where $c = \cos\theta$ and $s = \sin\theta$. Solving $R[(1/\sqrt{3})(1,1,1),\theta]L_f = L_b$ yields

$$\cos\theta = \frac{2+2r-r^2}{2(1+r+r^2}, \qquad \sin\theta = \frac{\sqrt{3}r(2+r)}{2(1+r+r^2} \tag{5.14}$$

which in turn gives us

$$\cos\frac{\theta}{2} = \frac{2+r}{2\sqrt{1+r+r^2}}, \qquad \sin\frac{\theta}{2} = \frac{\sqrt{3}r}{2\sqrt{1+r+r^2}} \tag{5.15}$$

which in turn gives us

$$R(\frac{1}{\sqrt{3}}(1,1,1),\frac{\theta}{2})L_f = L_s = \frac{\sqrt{1+r+r^2}}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} + \frac{1-r}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{5.16}$$

and it may be verified that

$$L_f^T L_f = {L_s}^2 = L_b^T L_b = \begin{bmatrix} 1+r^2 & -r & -r \\ -r & 1+r^2 & -r \\ -r & -r & 1+r^2 \end{bmatrix} \tag{5.17}$$

This illustrates that the simultaneous model L_s is "midway" along the rotation that takes L_f into L_b. Figure 5 illustrates this situation. It is important to note that in the absence of any information other than the observed series $\{y_1, y_2, \ldots y_n\}$, we have no reason to prefer a causal model over a simultaneous one, or vice versa. The two are mathematically interconvertible.

6 Simultaneous autoregressive models in 1D

A simultaneous (i.e. bilateral) 1D process is non-causal. Let's start by supposing that

$$x(t) = ax(t-1) + bx(t+1) + e(t), \qquad e(t) \sim \text{I.I.D.} \tag{6.1}$$

where $x(t)$ is a mean-reduced random series, and $e(t)$ is a stationary, IID random series of constant variance. Neglecting the homogeneous part of the solution, we might expect a particular part of the form

$$g(t) = \sum_{k=-\infty}^{t-1} a^{t-k} e(k) + \sum_{k=t+1}^{+\infty} b^{k-t} e(k) + e(t) \tag{6.2}$$

In fact, the equation that $g(t)$ satisfies is not Equation 6.1, but

$$(1+ab)g(t) = ag(t-1) + bg(t+1) + (1-ab)e(t), \qquad e(t) \sim \text{IID} \tag{6.3}$$

But there is a problem here. Since the innovation sequence is IID, we can reflect it about t without altering $g(t)$'s statistics. So replacing $e(k)$ with $e(2t-k)$ in Equation 6.2 gives

$$\begin{aligned} \bar{g}(t) &= \sum_{k=-\infty}^{t-1} a^{t-k} e(2t-k) + \sum_{k=t+1}^{+\infty} b^{k-t} e(2t-k) + e(t) \\ &= \sum_{k=-\infty}^{t-1} b^{t-k} e(k) + \sum_{k=t+1}^{+\infty} a^{k-t} e(k) + e(t) \end{aligned} \tag{6.4}$$

which is the same as $g(t)$, but with the autoregressive parameters a and b interchanged. But $g(t)$ and $\bar{g}(t)$ have identical statistics, so we cannot distinguish a from b. For example, the autocovariance of both $g(t)$ and $\bar{g}(t)$ is

$$\gamma(s) = \left[\frac{a^{|s|}}{1-a^2} + \frac{b^{|s|}}{1-b^2} + \frac{a^{|s|}b - ab^{|s|}}{a-b} \right] \text{var}[e(t)] \tag{6.5}$$

which is invariant under the interchange $a \leftrightarrow b$. Hence a stationary, discrete, simultaneous AR(1) process must have the form

$$(1+r^2)x(t) = rx(t-1) + rx(t+1) + (1-r^2)e(t), \qquad e(t) \sim \text{IID} \tag{6.6}$$

with particular solution

$$x(t) = e(t) + \sum_{k=1}^{\infty} r^k [e(t-k) + e(t+k)] \tag{6.7}$$

To ensure a bounded solution, we must require that $|r| < 1$. The corresponding autocovariance and marginal series variance are

$$\gamma(s) = r^{|s|} \left(\frac{1+r^2}{1-r^2} + |s| \right) \text{var}[e(t)] \tag{6.8}$$

$$\text{var}[x(t)] = \gamma(0) = \frac{1+r^2}{1-r^2} \text{var}[e(t)] \tag{6.9}$$

Now, the preceding section showed that under the very general assumption that the probability distribution of the innovation vector depends only upon its magnitude, conditional and simultaneous models are interconvertible. We can go even further and say that for (nx1)-dimensional observation and innovation vectors $\vec{x}$ and $\vec{e}$, the models L that satisfy $\vec{e}^T\vec{e} = \vec{x}^T L^T L\vec{x} = \vec{x}^T Q\vec{x}$ form a set that includes $\sqrt{Q}$ and all unitary transformations applied to $\sqrt{Q}$ (i.e. all $S\sqrt{Q}$ such that $S^T S = I_n$). But we must remember that although all those $S\sqrt{Q}$'s lead to the same precision matrix Q (i.e. to the same series autocovariance), they represent physically different realities. If we know the nature of the physical reality beforehand, then we may need to convert from a conditional model to a simultaneous one, or vice versa.

The easier conversion is from simultaneous to conditional form. The transfer function corresponding to equation (6.6) is

$$\frac{X(\omega)}{E(\omega)} = \frac{1-r^2}{1+r^2-2r\cos\omega} = \frac{1-r^2}{(1-re^{-j\omega})(1-re^{j\omega})} \tag{6.10}$$

where $X(\omega)$ and $E(\omega)$ are the discrete Fourier transforms of the observation and innovation vectors respectively. The corresponding spectral density is

$$\left|\frac{X(\omega)}{E(\omega)}\right|^2 = \left[\frac{1-r^2}{(1-re^{-j\omega})^2}\right]\left[\frac{1-r^2}{(1-re^{j\omega})^2}\right] \tag{6.11}$$

and it is therefore clear that the equivalent conditional model is

$$\left[\frac{X(\omega)}{E(\omega)}\right]_c = \frac{1-r^2}{(1-re^{-j\omega})^2} \tag{6.12}$$

which would have this form in the time domain:

$$x(t) = 2rx(t-1) - r^2x(t-2) + (1-r^2)e(t), \qquad e(t) \sim \text{IID} \tag{6.13}$$

where $x(t)$ is the mean-reduced series, and $e(t)$ is the innovation sequence. That's a conditional AR(2) model with a double root.

The harder conversion is from conditional to simultaneous form. If our conditional model is

$$x(t) = rx(t-1) + e(t), \qquad e(t) \sim \text{IID} \tag{6.14}$$

then the transfer function and spectral density are

$$\begin{aligned} \frac{X(\omega)}{E(\omega)} &= \frac{1}{1-re^{-j\omega}} \\ \left[\frac{X(\omega)}{E(\omega)}\right]^2 &= \frac{1}{1+r^2-2r\cos\omega} \end{aligned} \tag{6.15}$$

and thus the equivalent simultaneous model has

$$\left[\frac{X(\omega)}{E(\omega)}\right]_s = \frac{1}{\sqrt{1+r^2-2r\cos\omega}} = \frac{1}{\sqrt{1+r^2}}\frac{1}{\sqrt{1-2\beta\cos\omega}} \tag{6.16}$$

where $\beta = r/(1+r^2)$, $|\beta| < 1/2$. To fourth order in β, that would give

$$\sqrt{1+r^2}\left[\frac{X(\omega)}{E(\omega)}\right]_s = \frac{1}{c_0 - c_1\cos\omega - c_2\cos 2\omega - \frac{1}{8}\beta^3\cos 3\omega - \frac{5}{64}\beta^4\cos 4\omega} \tag{6.17}$$

where the first three coefficients are

$$c_0 = 1 - \frac{1}{4}\beta^2 - \frac{15}{64}\beta^4, \qquad c_1 = \beta + \frac{3}{8}\beta^3, \qquad c_2 = \frac{1}{4}\beta^2 + \frac{5}{16}\beta^4 \tag{6.18}$$

The expression in equation (6.16) actually converges pretty quickly; the leading term of order n is $-[(2n-3)!!/(2^{n-1}n!)]\beta^n \cos n\omega$, and $|\beta|$ is always less than one-half. Thus, a conditional AR(1) process has a simultaneous counterpart that is of infinite order. We may approximate its autoregressive roots as follows. Suppose that the equivalent simultaneous model has p autoregressive roots of the form $\{\phi_j = r_j r, 1 \le j \le p\}$. Then we would like to have, as closely as possible, a match-up between the spectral densities:

$$\frac{1}{1 + r^2 - 2r\cos\omega} = \frac{1}{\prod_{j=1}^{p}(1 + {\phi_j}^2 - 2\phi_j \cos\omega)^2} \tag{6.19}$$

Taking natural logarithms of both sides yields

$$\sum_{n=1}^{\infty} \frac{2}{n} r^n \cos n\omega = \sum_{n=1}^{\infty} \frac{4}{n} \left[\sum_{j=1}^{p} {\phi_j}^n \right] \cos n\omega \tag{6.20}$$

and substituting in $\phi_j = r_j r$ for the first p terms gives

$$\sum_{j=1}^{p} {r_j}^n = \frac{1}{2}, \qquad 1 \le n \le p. \tag{6.21}$$

For small values of p, we can solve that system of equations directly. For large values of p, the solution takes the approximate form

$$\begin{aligned} \log r_k &= -a_1(1/p) - \frac{(1/p)\sin(\theta/2)}{a_2 \sin(\theta/2) + a_3(1/p)} + j\theta, \\ \theta &= 2\pi(k-1)/p, \quad 1 \le k \le p, \quad a_1 = 0.83,\ a_2 = 0.15,\ a_3 = 0.455 \end{aligned} \tag{6.22}$$

to first order in $1/p$. Figure 6 shows the spectrum of equivalent autoregressive roots for several values of p. Using the foregoing results, and the methods of partial fraction expansion outlined in Section 21, we can interconvert 1D conditional and simultaneous autoregressions at will.

The two styles of 1D autoregressive series each have their advantages and disadvantages. Under a conditional model of order p, knowledge of p successive mean-reduced variates $\{x(t) \ldots x(t+p-1)\}$ splits the remaining series into two independent parts. Using the causal model

$$x(t) = \sum_{j=1}^{p} a_j x(t-j) + e(t), \qquad e(t) \sim \text{IID} \tag{6.23}$$

we can write the probability distribution for $\{x(t)|t > p\}$ as

$$p[x(t+p+1), x(t+p+2), \ldots] = \prod_{j=t+p}^{\infty} p(x(j)|\{x(j-1) \ldots x(j-p)\}) \tag{6.24}$$

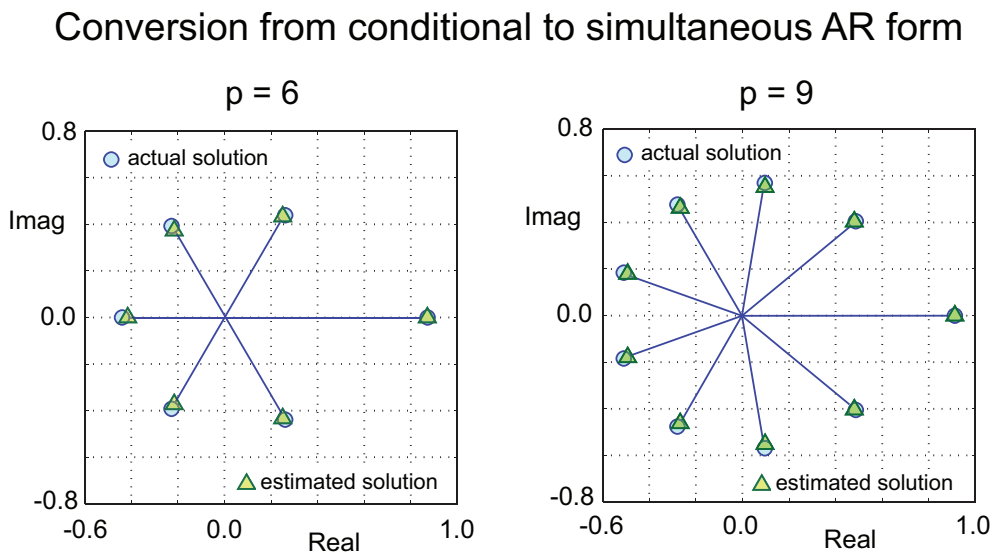

Figure 6: Approximations to the roots of a simulataneous autoregressive process that has the same spectral density as a conditional AR(1) process. The estimated solution is that calculated through Equation 6.22.

and using the equivalent anticausal model

$$x(t) = \sum_{j=1}^{p} a_j x(t+j) + e(t), \qquad e(t) \sim \text{IID} \tag{6.25}$$

we also have

$$p[\ldots x(t-2), x(t-1)] = \prod_{j=-\infty}^{t-1} p(x(j) | \{x(j+1) \ldots x(j+p)\}) \tag{6.26}$$

Those two distributions are independent, conditioned on knowledge of $\{x(t) \ldots x(t+p-1)\}$. That's where the "conditional" name comes from! Under a simultaneous model, however, no such division is possible. For example, if we expand a simultaneous AR(p) model into the form

$$\frac{X(z)}{E(z)} = \sum_{k=1}^{p} \frac{d_k}{1 + {r_k}^2 - r_k(z + z^{-1})} \tag{6.27}$$

via a partial fraction expansion, then the moving average equivalent is

$$x(t) = \sum_{k=1}^{p} \frac{d_k}{1 - {r_k}^2} \left[e(t) + \sum_{j=1}^{\infty} {r_k}^j [e(t+j) + e(t-j)] \right] \tag{6.28}$$

which involves *all* the innovations. That's where the "simultaneous" name comes from; if the series is a temporal one, then the probability distribution of any subset of the variates must necessarily involve all of them, past and future.

From the point of view of practical computations, the most important difference between conditional and simultaneous models has to do with the "bandedness" of the associated model matrices. Suppose that we estimate the covariance of an n-point 1D series,

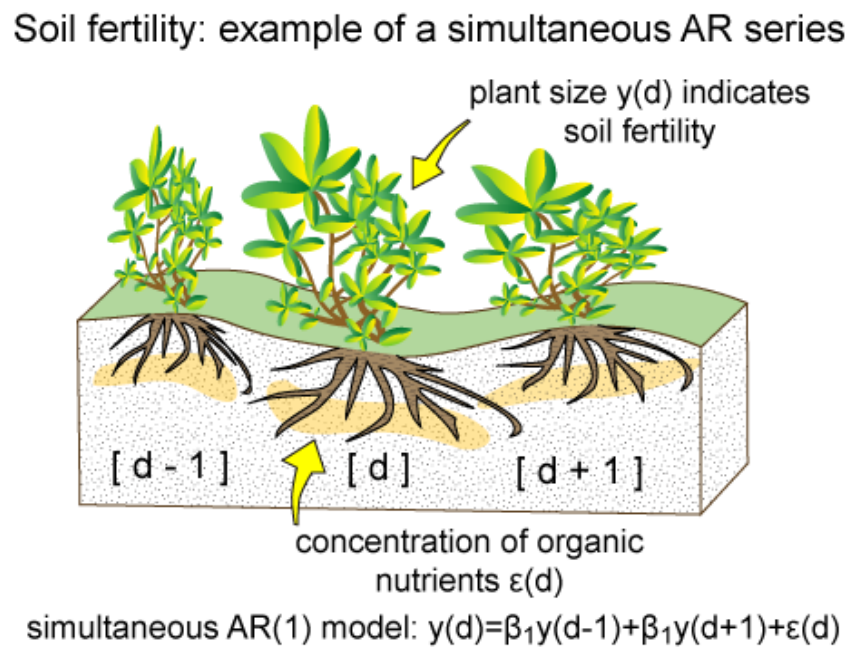

Figure 7: A spatial transect involving soil fertility will most likely be simultaneous in nature, if it is autoregressive at all.

and find that the precision matrix Q is banded with width $2p$. A causal model will have a lower diagonal matrix, banded with width p. A simultaneous model, in the best case, will be banded with width p; in the worst case, it will be dense.

But the worst case is not that bad! Recall from Equation 6.16 above that if we convert a conditional AR(p) model into simultaneous form, the autoregressive coefficients decrease as β^k, where k is the distance from the main diagonal and $|\beta| < 1/2$. Thus we could approximate the simultaneous model very well with a banded matrix of width $p + 7$ or $p + 8$. In the case of a very large series, the fact that the causal model is banded with a small, finite width may override all other considerations for model selection.

On the other hand, we may have a system which is clearly simultaneous in nature. For example, suppose we take a transect through an agricultural field and use plant size as an indication of soil fertility, as I have illustrated in Figure 7. In this figure, the concentration of organic nutrients below each plant plays the role of an innovation. As long as the land is flat, each plant's roots will tend to spread equally to either side, resulting in a simultaneous model.

7 Restrictions on the autoregressive parameters

The bulk of my thesis (Section 12 onwards) involves estimating the parameters for conditional and simultaneous autoregressive fields. This section, and Sections 8-11, lay the mathematical foundations of autoregressive parameter estimation, so that we can know how reliable our estimates are, and whether a particular parametrization is even admissible. Autoregressive models suffer from a truncation problem: if we construct a model with p parameters, and then drop the highest-order one, the resulting model with $p-1$ parameters may be invalid. A reasonable first step, then, is to see what combinations of parameters are possible.

First, we'll examine the 1D models in some detail, and then touch upon 2D models. A conditional AR(p) process has a z-domain transfer function of the form

$$\begin{aligned} \frac{X(z)}{E(z)} &= \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} \ldots - a_p z^{-p}} \\ &= \frac{1}{(1 - r_1 z^{-1})(1 - r_2 z^{-1}) \ldots (1 - r_p z^{-1})} \end{aligned} \tag{7.1}$$

where $X(z)$ is the z-transform of the mean-reduced series, $E(z)$ is the z-transform of the innovation sequence, and $\{r_1, \ldots r_p\}$ are the roots of $z^p - a_1 z^{p-1} \ldots - a_p = 0$. In order to ensure a real series, the autoregressive roots must be real, or occur in conjugate pairs. The spectral density corresponding to equation (7.1) is

$$\begin{aligned} \left|\frac{X(z)}{E(z)}\right|^2 &= \frac{1}{\left[1 + r_1^2 - r_1(z + z^{-1})\right] \ldots \left[1 + r_p^2 - r_p(z + z^{-1})\right]} \\ &= \frac{1}{(1 - a_1 z^{-1} - \ldots - a_p z^{-p})(1 - a_1 z - \ldots - a_p z^{p})} \\ &= \frac{1}{1 + {a_1}^2 + \ldots + {a_p}^2} \; \frac{1}{1 - \beta_1(z + z^{-1}) - \ldots - \beta_p(z^p + z^{-p})} \end{aligned} \tag{7.2}$$

where β_j is given by

$$\beta_j = \frac{a_j - \sum_{k=1}^{p-j} a_k a_{k+j}}{1 + \sum_{k=1}^{p} {a_k}^2} \tag{7.3}$$

The thing to note here is that the spectral density of the conditional AR(p) process is proportional to the transfer function of a simultaneous AR(p) process that has the same roots. Hence, from the known allowable space of the $\{r_1 \ldots r_p\}$, we can calculate the allowable spaces of the $\{a_1 \ldots a_p\}$ and $\{\beta_1 \ldots \beta_p\}$. For example, for a conditional AR(2) process, the root space is composed of two parts:

$$\begin{aligned} \text{Both roots real}: \quad & -1 < r_1 < 1, && -1 < r_2 < r_1 \\ \text{Complex conjugates}: \quad & r_1 = re^{j\theta}, && r_2 = re^{-j\theta} \\ & 0 < \theta < \pi, && 0 \le r < 1 \end{aligned} \tag{7.4}$$

Transforming those spaces through

$$a_1 = r_1 + r_2, \quad a_2 = -r_1 r_2, \quad \beta_1 = \frac{a_1(1 - a_2)}{1 + {a_1}^2 + {a_2}^2}, \quad \beta_2 = \frac{a_2}{1 + {a_1}^2 + {a_2}^2} \tag{7.5}$$

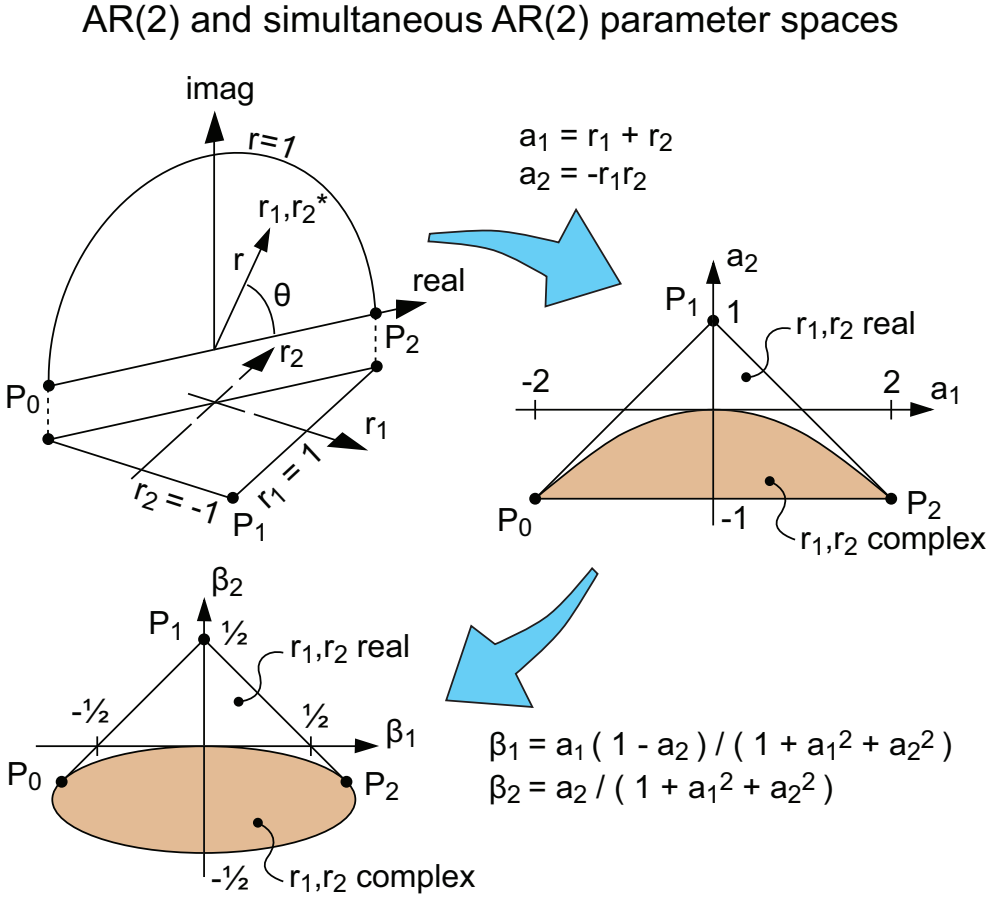

Figure 8: The parameter spaces for conditional AR(2) and simultaneous AR(2) processes. P_0, P_1, and P_2 are the extreme points of the root space, and the figure shows their mappings into the $\{a_1, a_2\}$ and $\{\beta_1, \beta_2\}$ realms.

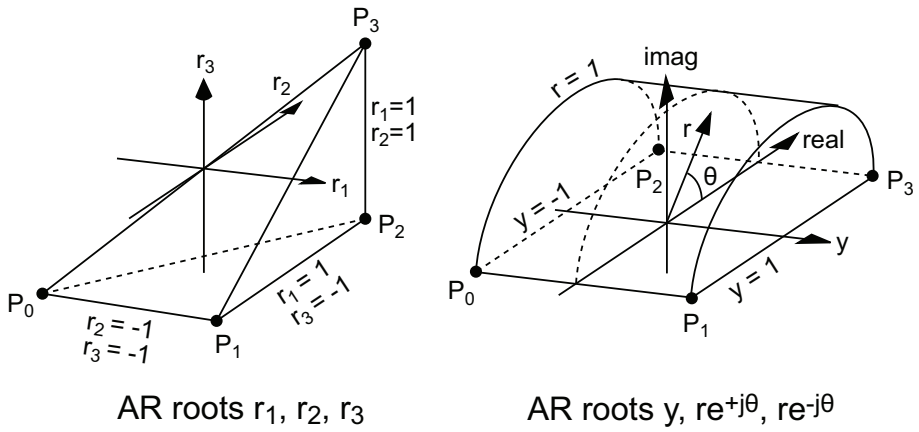

Figure 9: The autoregressive root space for AR(3) processes. There may be three real roots, or one real and two conjugate roots. P_0, P_1, P_2, and P_3 are extreme points of the root space.

gives us the allowable parameter spaces for $\{a_1, a_2\}$ and $\{\beta_1, \beta_2\}$. Figure 8 illustrates these parameter spaces.

For a conditional AR(3) process, the root space is again composed of two parts:

3 real roots:	$-1 < r_1 < 1,$	$-1 < r_2 \leq r_1,$	$-1 < r_3 \leq r_2$
1 real, 2 complex roots:	$-1 < r_1 < 1,$	$r_2 = re^{j\theta},$	$r_3 = re^{-j\theta}$
	$0 < \theta < \pi,$	$0 \leq r < 1$	

Now transform that root space to $\{a_1, a_2, a_3\}$ space through

$$\begin{aligned} a_1 &= r_1 + r_2 + r_3 \\ a_2 &= -(r_1 r_2 + r_2 r_3 + r_1 r_3) \\ a_3 &= r_1 r_2 r_3 \end{aligned} \tag{7.6}$$

Now transform that to $\{\beta_1, \beta_2, \beta_3\}$ space through

$$\begin{aligned} (1 + {a_1}^2 + {a_2}^2 + {a_3}^2)\beta_1 &= a_1 - a_1 a_2 - a_2 a_3 \\ (1 + {a_1}^2 + {a_2}^2 + {a_3}^2)\beta_2 &= a_2 - a_1 a_3 \\ (1 + {a_1}^2 + {a_2}^2 + {a_3}^2)\beta_3 &= a_3 \end{aligned} \tag{7.7}$$

Figures 10 and 11 illustrate the parameter spaces for conditional and simultaneous AR(3) processes in one dimension. Figure 12 illustrates that model truncation is not necessarily possible. By model truncation, I mean dropping the last parameter of an AR(p) model in order to get an AR($p-1$) model. For example, Figure 12 illustrates a case in which the combination $C = (a_1, a_2, a_3) = (0.6, 0.6, -0.4)$ is allowed, but dropping a_3 results in the disallowed combination $\bar{C} = (0.6, 0.6, 0.0)$.

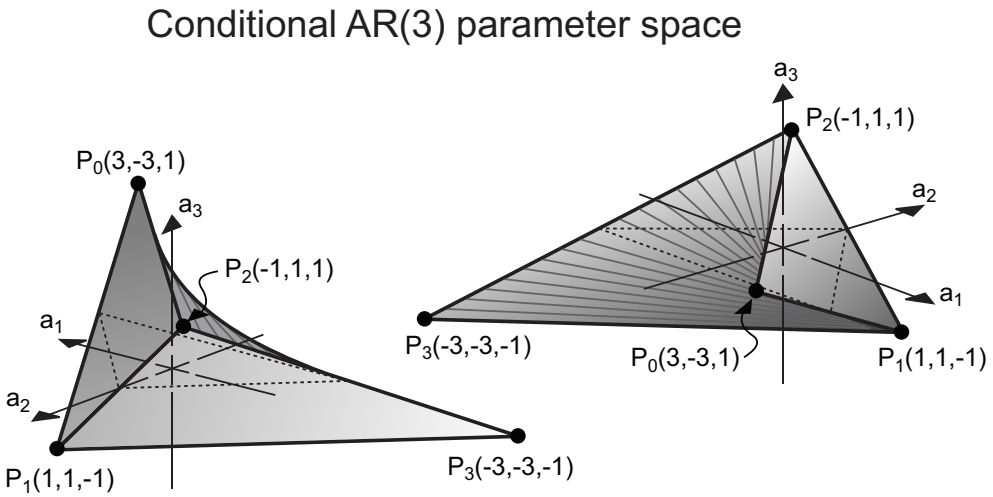

Figure 10: The parameter space for a conditional AR(3) process. P_0, P_1, P_2, and P_3 are the mappings of the extreme points of the root space. The dotted line represents the subspace obtained by setting $a_3 = 0$. The region's boundary has two planar surfaces and one hyberboloid sheet.

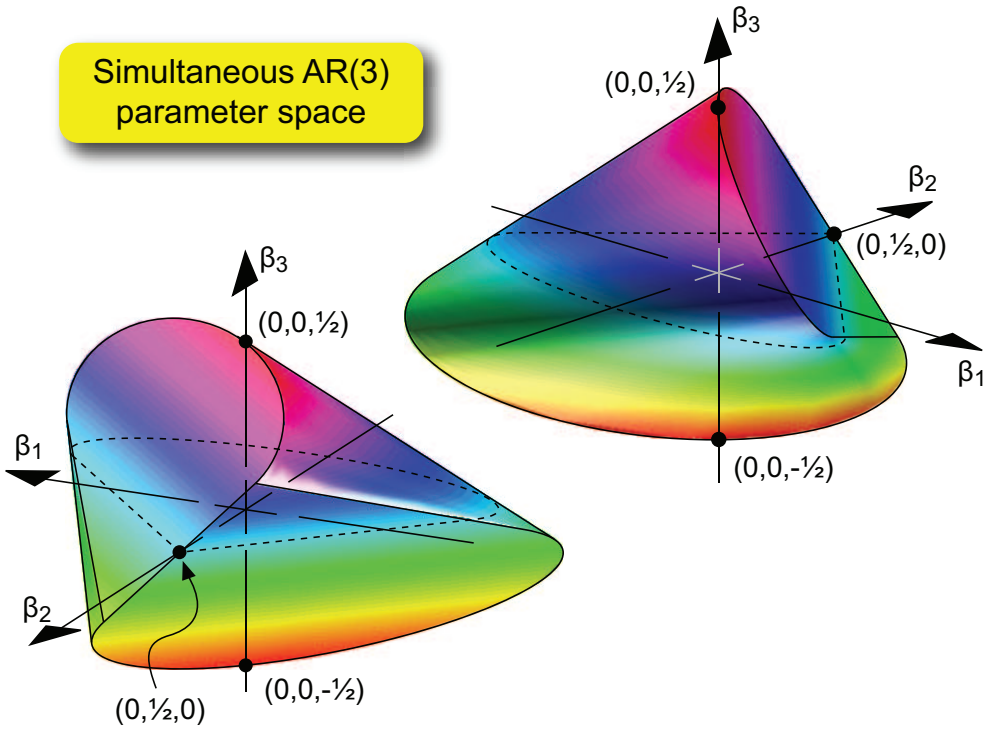

Figure 11: The parameter space for a simultaneous AR(3) process. The dotted line represents the subspace obtained by setting $\beta_3 = 0$.

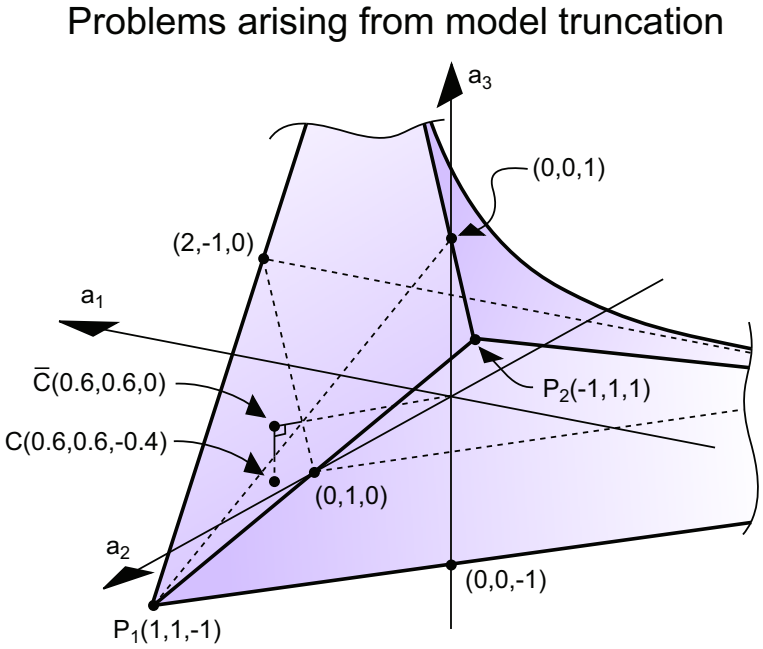

Figure 12: Truncating the last parameter from a conditional AR(p) model may result in an inadmissible AR($p-1$) model. In this example with $p = 3$, point C lies within the allowed region, but point $\bar{C}$ does not.

Autoregressive parameter restrictions in two dimensions are considerably more complicated than in the one-dimensional case. A transfer function such as

$$\begin{aligned} \frac{X(\omega_x,\omega_y)}{E(\omega_x,\omega_y)} &= [1 - 2\beta_{1,0}\cos\omega_x - 2\beta_{0,1}\cos\omega_y \\ &\quad -2\beta_{1,1}\cos(\omega_x+\omega_y) - 2\beta_{-1,1}\cos(-\omega_x+\omega_y) - \ldots]^{-1} \end{aligned} \tag{7.8}$$

(in the simultaneous case) or indeed a spectral density of that form (in the conditional case) is not generally factorable [27]. The general requirement is easy to state: if the transfer function or spectral density takes the form

$$\frac{1}{D(z_x,z_y)} = \left[1 - \sum_{\vec{k}} \beta(\vec{k}) {z_x}^{k_x} {z_y}^{k_y}\right]^{-1} \tag{7.9}$$

where the summation is over all those displacements $\vec{k}$ such that $\vec{k} \neq 0$, $\beta(\vec{k}) \neq 0$, and $\beta(\vec{k}) = \beta(-\vec{k})$, then the locus of the zeros of $D(z_x, z_y)$ must lie outside the unit bicircle [28]. However, working that out in detail can be tricky. For example, suppose we have

$$\begin{aligned} D(\omega_x,\omega_y) &= 1 - 2\beta_{10}\cos\omega_x - 2\beta_{01}\cos\omega_y \\ &\quad -2\beta_{11}\cos(\omega_x+\omega_y) - 2\beta_{11}\cos(-\omega_x+\omega_y) \\ &= 1 - 2\beta_{10}\cos\omega_x - 2\beta_{01}\cos\omega_y - 4\beta_{11}\cos\omega_x\cos\omega_y \end{aligned} \tag{7.10}$$

Then $D(\omega_x,\omega_y) = 0$ is a conic in the $(\cos\omega_x, \cos\omega_y)$ plane, and will touch the square $\{|\cos\omega_x| \leq 1, |\cos\omega_y| \leq 1\}$ at one of its four corners, so the allowed parameter space in

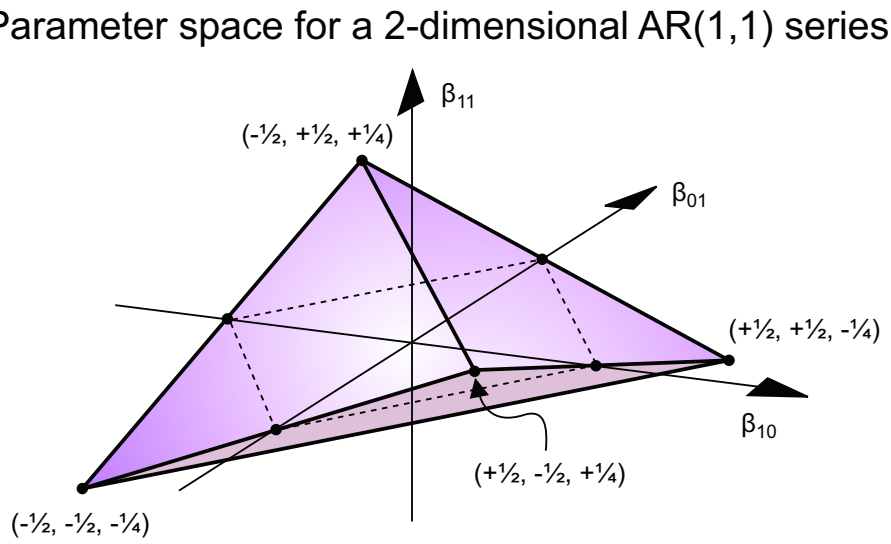

Figure 13: The parameter space for a restricted class of AR(1,1) two-dimensional autoregressive processes in which $\beta(-1,1) = \beta(1,1) = \beta_{11}$. The dotted line represents the subspace obtained by setting $\beta_{11} = 0$.

this case is

$$\begin{aligned} 2\beta_{10} + 2\beta_{01} + 4\beta_{11} &< 1 \\ -2\beta_{10} + 2\beta_{01} - 4\beta_{11} &< 1 \\ 2\beta_{10} - 2\beta_{01} - 4\beta_{11} &< 1 \\ -2\beta_{10} - 2\beta_{01} + 4\beta_{11} &< 1 \end{aligned} \tag{7.11}$$

which I have sketched in Figure 13. The whole topic is explored further in Lakshmanan and Derin [29].

8 Limitations of autoregressive models

All parametric models have their limitations. In this section, I will show that autoregressive models are a poor choice when the random field under scrutiny is nearly periodic. For a stationary conditional autoregressive field, the inverse spectral density is a cosine series. Let's take the example of a one-dimensional AR(p) series, for which the spectral density is

$$\begin{aligned}
\left|\frac{X(z)}{E(z)}\right|^2 &= \frac{1}{(1-a_1z^{-1}-\ldots-a_pz^{-p})(1-a_1z-\ldots-a_pz^p)} \\
&= \frac{1}{1+{a_1}^2+\ldots+{a_p}^2}\,\frac{1}{1-\beta_1(z+z^{-1})-\ldots-\beta_p(z^p+z^{-p})} \\
\Longrightarrow \left|\frac{X(\omega)}{E(\omega)}\right|^2 &= \frac{1}{1+{a_1}^2+\ldots+{a_p}^2}\,\frac{1}{1-2\beta_1\cos\omega-\ldots-2\beta_p\cos p\omega} \qquad (8.1)
\end{aligned}$$

and the general expression for β_j is

$$\beta_j = \frac{a_j - \sum_{k=1}^{p-j} a_k a_{k+j}}{1+\sum_{k=1}^{p}{a_k}^2} \qquad (8.2)$$

So the inverse of the spectral density is

$$\left[\left|\frac{X(\omega)}{E(\omega)}\right|^2\right]^{-1} = (1+{a_1}^2+\ldots+{a_p}^2)(1-2\beta_1\cos\omega-\ldots-2\beta_p\cos p\omega) \qquad (8.3)$$

which is definitely a cosine series. Similarly, for a stationary simultaneous 1D autoregressive field, the inverse square root of the spectral density is a cosine series:

$$\begin{aligned}
\frac{X(z)}{E(z)} &= \frac{1}{1-\beta_1(z+z^{-1})-\ldots-\beta_p(z^p+z^{-p})} \\
\Longrightarrow \left[\left|\frac{X(\omega)}{E(\omega)}\right|^2\right]^{-\frac{1}{2}} &= 1-2\beta_1\cos\omega-\ldots-2\beta_p\cos p\omega \qquad (8.4)
\end{aligned}$$

Thus, autoregressive models are well suited to random fields for which the inverse spectral density, or inverse square root of the spectral density, is well approximated by a cosine series with a small number of terms. When do we *not* get this? Clearly, when the series is truly periodic, or very nearly periodic. In this case, the spectral density has sharp peaks, leading to equally sharp "notches" in the inverse spectral density. Figure 14 illustrates that situation for a nearly periodic two-dimensional random field.

The problem is this: we not only have to reproduce those notches, we have to ensure they do not dip below the zero plane, since spectral density is always positive. So we will need a large number of terms in our cosine series. But that defeats the purpose of modeling: we would like to end up with a small batch of autoregressive parameters, not an extensive one.

I will illustrate this with two random series, both of which are nearly periodic. The first is the well-known Wolfer sunspot series. This series has been discussed at length elsewhere, for example in Hipel and McLeod [30], so I will not dwell on it. The sunspot

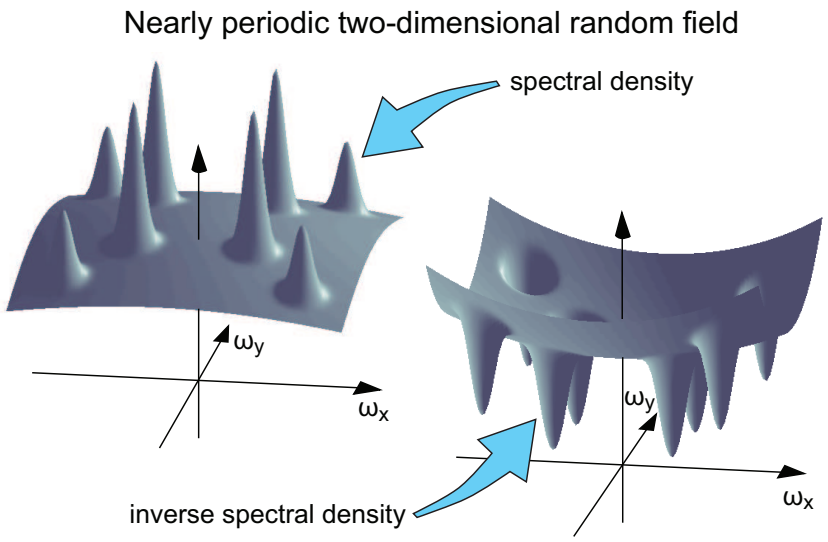

Figure 14: Spectral density and inverse spectral for a nearly periodic two-dimensional autoregressive random field.

series shows approximate cyclic behavior at two periodicities of roughly 11 years and 100 years. Figure 15 shows the series, its estimated spectral density, and the inverse spectral density. I estimate the spectral density by the Welch method, which I describe further in Section 11.

Capturing those first two dips in the inverse spectral density is essential, and the corresponding discrete frequency change is some $2\pi/10$, meaning we will need at least an AR(9) or AR(10) model to capture the nearly-periodic character of the sunspot series. This conclusion remains valid even if we subject the series to a Box-Cox transformation.

My next example is a visual texture, and many visual textures are in fact nearly periodic. Figures 16 and 17 below show the estimated autocovariance and spectral density of the Brodatz raffia texture D84 [31]. I estimated the autocovariance by a discrete FFT method, which I describe further in Section 15.

The thing to note here is that for a human observer, the approximate periodicity of the raffia texture is its most salient feature. For the autocovariance, however, and especially for the spectral density, that near-periodicity is a minor feature. The corresponding peaks in the spectral density are very narrow and tight around the origin, and if our autoregressive model is of a low enough order, we will miss them entirely.

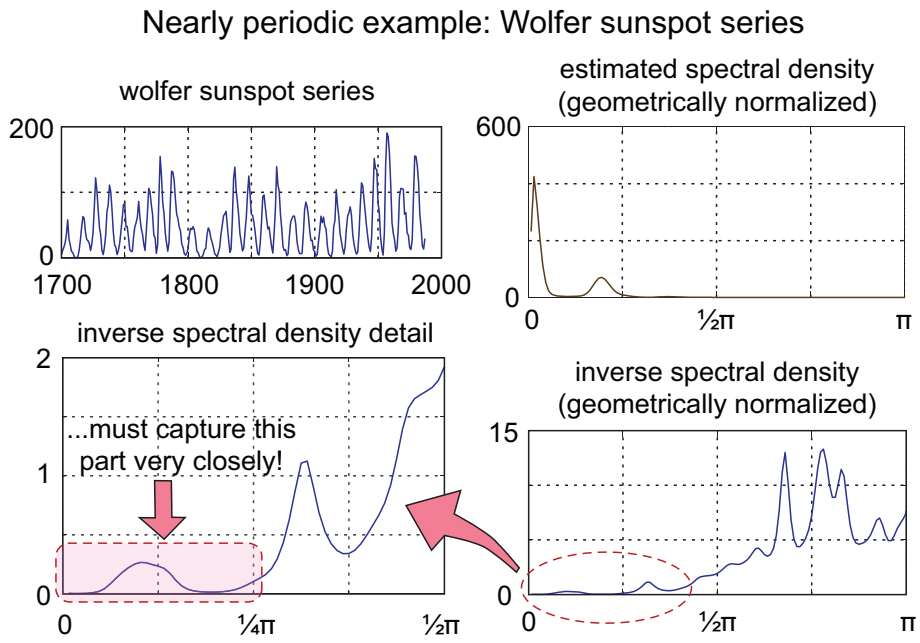

Figure 15: The Wolfer sunspot series illustrates the difficulties of modelling a nearly periodic random series. I call a series 'geometrically normalized' when its average logarithm is zero.

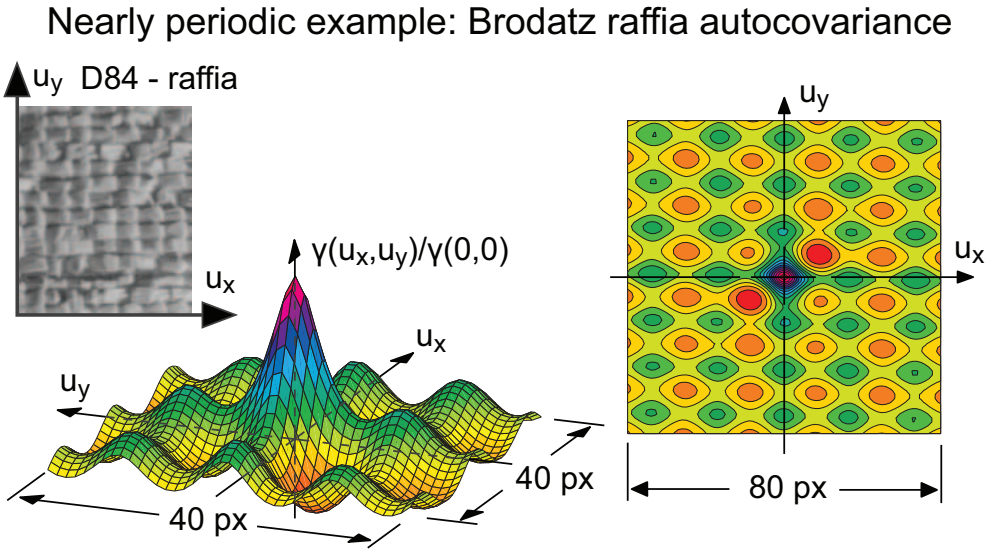

Figure 16: Estimated autocovariance of the Brodatz raffia visual texture at zero degrees rotation. The quantity $\gamma(u_x, u_y)/\gamma(0,0)$ reaches a minimum value of approximately -0.2.

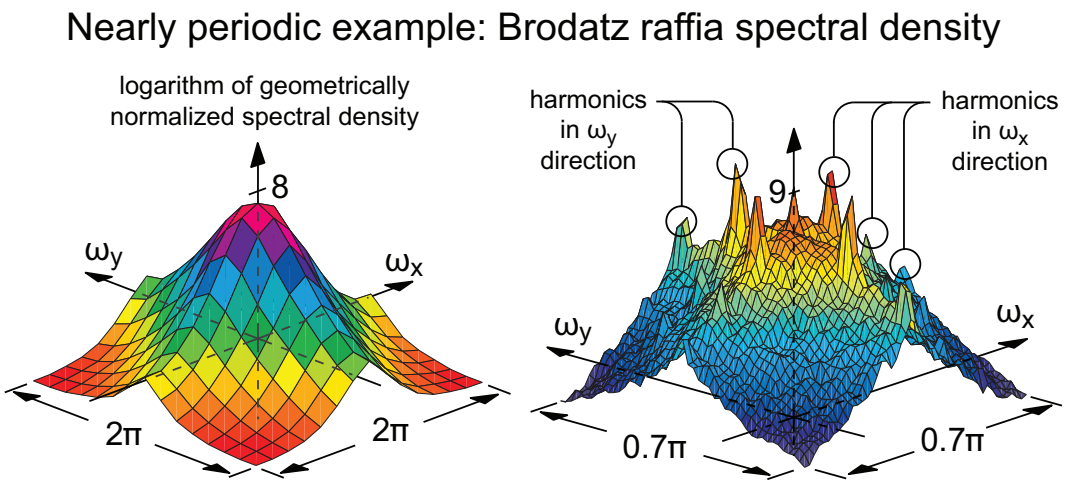

Figure 17: Estimated spectral density of the Brodatz raffia visual texture at zero degrees rotation. The chart on the left uses the Welch method with a 16 x 16 Hamming window, while the one on the right uses a 128 x 128 Hamming window. In each case the vertical scale is logarithmic. The spectral density is scaled so that the average logarithm is zero.

9 Approximations involved in maximum likelihood estimates

We have looked at the mathematical nature and limitations of autoregressive random fields, so a reasonable next step is to look at how we can estimate autoregressive parameters, how reliable those estimates can be, and what effects our approximations have. Before we can talk about maximum likelihood estimates for the various parameters involved in a stationary autoregression, however, we need to choose a probability distribution for the innovations. Now, I said right at the beginning that I will imagine my random fields to be quasi-continuous, in the sense that I can sample them at some resolution of interest, and also at a much smaller intrinsic resolution. For the sake of argument, suppose that we have a stationary, temporal, conditional, mean-reduced autoregressive field $x(t)$ which we examine at a resolution that is a multiple M of the intrinsic resolution. So

$$\begin{aligned} x(t) &= \sum_{k=1}^{p} a_j x(t-k) + e_1(t), \qquad e_1(t) \sim \text{IID} \\ \frac{X(z)}{E(z)} &= \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} \end{aligned} \tag{9.1}$$

Defining $D(z)$ and $\{r_1 \dots r_p\}$ through

$$z^{-p} D(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} = \prod_{k=1}^{p} (1 - r_k z^{-1}) \tag{9.2}$$

the moving average equivalent is (see Section 21)

$$\begin{aligned} x(t_1) &= \sum_{j=0}^{\infty} \sum_{k=1}^{p} \frac{{r_k}^{p-1}}{D'(r_k)} {r_k}^{j} e_1(t_1 - j) \\ &= \sum_{j=-\infty}^{t_1} \left[\sum_{k=1}^{p} \frac{{r_k}^{p-1}}{D'(r_k)} {r_k}^{t_1 - j} \right] e_1(j) \\ &= \sum_{j=-\infty}^{t_1} G_1(j - t_1) e_1(j) \end{aligned} \tag{9.3}$$

When we change to the smaller intrinsic resolution, the transfer function is still a finite partial fraction (this is the subject of Section 12), so we get a similar result at that resolution:

$$x(t_2) = \sum_{j=-\infty}^{t_2} G_2(j - t_2) e_2(j) \tag{9.4}$$

By assumption our random field is stationary, so these two representations must hold at all field points t, and that can only happen if $e_1(t_1)$ is a linear combination of the neighboring $e_2(t_2)$'s when t_1 and t_2 denote the same field point. Figure 18 illustrates this situation.

Specifically, we need

$$\begin{aligned} G_1(0) e_1(t_1) &= \sum_{j=-M+1}^{0} G_2(j) e_2(t_2 + j) \\ G_1(-1) e_1(t_1 - 1) &= \sum_{j=-2M+1}^{-M} G_2(j) e_2(t_2 + j) \quad \text{etc} \end{aligned} \tag{9.5}$$

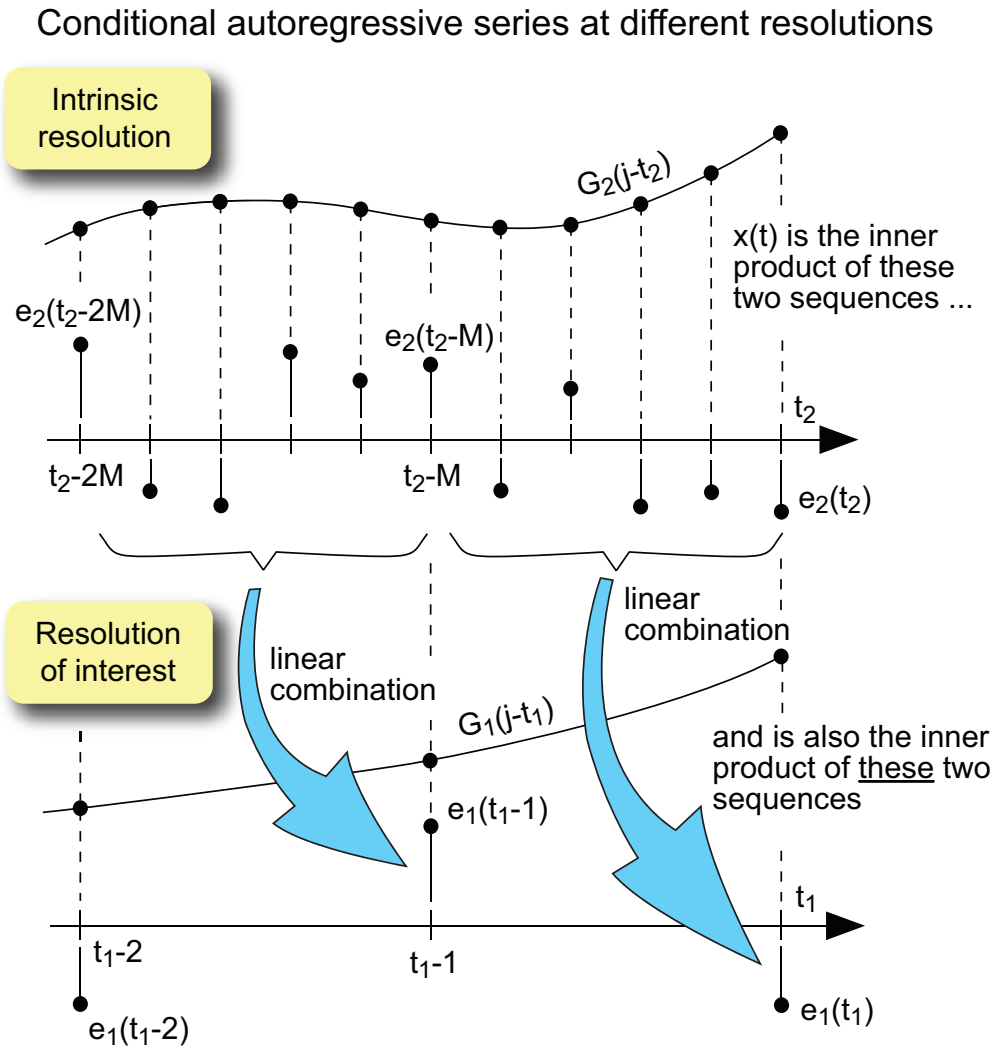

Figure 18: If a random field is autoregressive at two commensurate resolutions, then their innovations must be related.

This shows that $e_1(t_1)$ is a linear combination of $\{e_2(t_2-M+1)\ldots e_2(t_2)\}$ where t_1 and t_2 represent the same physical time. So regardless of the actual probability distribution of the $e_2(t_2)$'s, the central limit theorem assures us that the $e_1(t_1)$'s will have an approximately Gaussian distribution.

For the purposes of this section, then, I will assume a Gaussian distribution for the innovations. We'll consider the approximations inherent in estimating a conditional AR(1) series, and generalize from there. According to Equation 5.7, for a sample of length n, drawn from an unbounded conditional AR(1) series $\{\ldots y(0), y(1), \ldots y(n), y(n+1), \ldots\}$ with mean μ, we have, with $x_j = y_j - \mu$:

$$L\vec{x} = \begin{bmatrix} \sqrt{1-r^2} & 0 & \cdots & 0 & 0 \\ -r & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -r & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \vec{e} \tag{9.6}$$

where r is the autoregressive root and now $e_j \sim N(0, v)$, v being the variance of the innovation sequence. The corresponding probability density is

$$\begin{aligned} p(\vec{x}) &= \frac{1}{(2\pi v)^{n/2}} |\det L| \exp\left[-\frac{\vec{x}^T (L^T L)\vec{x}}{2v}\right] \\ &= \frac{1}{(2\pi v)^{n/2}} \sqrt{1-r^2} \exp\left[-\frac{1}{2v}\left[(1-r^2){x_1}^2 + \sum_{j=2}^{n}(x_j - r x_{j-1})^2\right]\right] \end{aligned} \tag{9.7}$$

and our minimization target will be

$$\begin{aligned} T &= -\frac{2}{n}\log p(\vec{x}) - \log 2\pi \\ &= \log v - \frac{2}{n}\log\sqrt{1-r^2} + \frac{1}{nv}U(\mu, r) \end{aligned} \tag{9.8}$$

where $U(\mu, r)$ is the residual sum of squares:

$$U(\mu, r) = (1-r^2)(y_1-\mu)^2 + \sum_{j=2}^{n}\left[(y_j-\mu) - r(y_{j-1}-\mu)\right]^2 \tag{9.9}$$

Minimizing T with respect to v, μ, and r gives us the normal equations:

$$\begin{aligned} v &= \frac{1}{n}\left[(1-r^2){x_1}^2 + \sum_{j=2}^{n}(x_j - r x_{j-1})^2\right] \\ 0 &= \sum_{j=1}^{n}(y_j - \mu) + \frac{r}{1-r}\left[(y_1-\mu) + (y_n - \mu)\right] \\ 0 &= -\sum_{j=2}^{n} x_j x_{j-1} + r\sum_{j=2}^{n-1}{x_j}^2 + r\frac{v}{1-r^2} \\ 0 &= -\sum_{j=1}^{n-1} x_j x_{j+1} + r\sum_{j=1}^{n-1}{x_j}^2 - r{x_1}^2 + r\frac{v}{1-r^2} \end{aligned} \tag{9.10}$$

Let the zeroth-order estimates for the mean and innovation variance be the following:

$$\begin{aligned} \mu_0 &= \frac{1}{n}\sum_{j=1}^{n} y_j \\ v_0 &= \frac{1}{n-1}\sum_{j=2}^{n}(x_j - r x_{j-1})^2 \end{aligned} \tag{9.11}$$

In terms of these quantities our first two normal equations are

$$\begin{aligned} n\hat{v} &= (1-r^2){x_1}^2 + (n-1)v_0 \\ 0 &= n(\mu_0 - \hat{\mu}) + \frac{r}{1-r}\left[(y_1 - \hat{\mu}) + (y_n - \hat{\mu})\right] \end{aligned} \tag{9.12}$$

These equations demonstrate that both v_0 and μ_0 differ by terms of order $1/n$ from their true values $\hat{v}$ and $\hat{\mu}$. If, however, $|r|$ is near 1, then the edge effects become important, especially for the sample mean. As for the autoregressive root r, we note that $v/(1-r^2)$ is an estimate of the marginal series variance $\bar{x^2} = E[x^2]$, and the zeroth-order estimate of r is

$$r_0 = \frac{\sum_{j=1}^{n-1} x_j x_{j+1}}{\sum_{j=1}^{n-1} {x_j}^2} \tag{9.13}$$

and consequently our third normal equation becomes

$$\begin{aligned} r_0 \sum_{j=1}^{n-1} {x_j}^2 &\approx \hat{r}\sum_{j=1}^{n-1} {x_j}^2 - \hat{r}(\bar{x^2} - {x_1}^2) \\ \Longrightarrow (r_0 - \hat{r})(n-1)\bar{x^2} &\approx \hat{r}({x_1}^2 - \bar{x^2}) \end{aligned} \tag{9.14}$$

Thus the estimate r_0 differs from its true value $\hat{r}$ by a term of order $1/n$, much like our other maximum likelihood estimates.

Now, the precision matrix in the previous example has a Toeplitz structure, as indicated in Equation 9.6:

$$Q = L^T L = \begin{bmatrix} 1 & -r & \cdots & 0 & 0 \\ -r & 1+r^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1+r^2 & -r \\ 0 & 0 & \cdots & -r & 1 \end{bmatrix} \tag{9.15}$$

Suppose we give it a circular structure instead; that is,

$$Q = \begin{bmatrix} 1+r^2 & -r & \cdots & 0 & -r \\ -r & 1+r^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1+r^2 & -r \\ -r & 0 & \cdots & -r & 1+r^2 \end{bmatrix}, \quad L = \begin{bmatrix} 1 & 0 & \cdots & 0 & -r \\ -r & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & -r & 1 \end{bmatrix} \tag{9.16}$$

Here, $\det L$ is now $1 - r^n$, and thus

$$p(\vec{x}) = \frac{1-r^n}{(2\pi v)^{n/2}} \exp -\frac{1}{2v}\left[(x_1 - r x_n)^2 + \sum_{j=2}^{n}(x_j - r x_{j-1})^2\right] \tag{9.17}$$

which results in the following normal equations for maximum likelihood estimates:

$$\begin{aligned} \mu &= \frac{1}{n}\sum_{j=1}^{n} y_j \text{ (no approximation!)} \\ v &= \frac{1}{n}\left[(x_1 - rx_n)^2 + \sum_{j=2}^{n}(x_j - rx_{j-1})^2\right] \\ x_1x_n + \sum_{j=1}^{n-1} x_jx_{j+1} &= r\sum_{j=1}^{n} {x_j}^2 + \frac{nr^{n-1}}{1-r^n}v \end{aligned} \tag{9.18}$$

As for that last equation, we will do no worse than before (i.e. than Equation 9.10) if we maintain

$$\frac{nr^{n-1}}{1-r^n} \le \frac{r}{1-r^2} \quad \Longrightarrow \quad r < 1 - \frac{1}{2(n+1)} \tag{9.19}$$

As long as we have that, then approximating the precision matrix as circulant instead of Toeplitz introduces errors of order $1/n$, which is no worse than what we had before.

So, now we can talk about relative errors. As we have just seen, when estimating a conditional AR(1) series, the circulant approximation and the zeroth-order parameter estimates

$$\begin{aligned} \mu_0 &= \frac{1}{n}\sum_{j=1}^{n} y_j \\ v_0 &= \frac{1}{n-1}\sum_{j=1}^{n-1}(x_{j+1} - rx_j)^2 \qquad (x_j = y_j - \mu_0) \\ r_0 &= \left[\sum_{j=1}^{n-1} x_jx_{j+1}\right] / \left[\sum_{j=1}^{n-1} {x_j}^2\right] \end{aligned} \tag{9.20}$$

introduce biases of approximate magnitude $r/n(1-r)$. And what about the variance of those estimates? I treat this subject further in Section 11, but for my purposes here I will note that if v is the innovation variance, then for a conditional AR(1) process (see Equations 11.8 and 11.9)

$$\begin{aligned} \mathrm{var}(\mu_0 - \mu) &\approx \frac{1}{(1-r)^2}\frac{v}{n} \\ \mathrm{var}(r_0 - r) &\approx \frac{1-r^2}{n} \end{aligned} \tag{9.21}$$

Those results show that for a conditional AR(1) series, the ratio of bias to standard error for μ_0 and r_0 is approximately $1/\sqrt{n}$, and so for sufficiently large series, the approximations we have made so far introduce negligible bias. To be completely sure of that, we should require

$$\frac{1}{1-|r_k|} < \frac{1}{\sqrt{n}} \quad \Longrightarrow \quad |r_k| < 1 - \frac{1}{\sqrt{n}} \tag{9.22}$$

for all of the autoregressive roots $\{r_1 \ldots r_p\}$.

That leaves us with two cases to consider: two-dimensional autoregressions, and simultaneous autoregressions. The effect of the circulant approximation on two-dimensional conditional autoregressive fields is covered in depth by Rue and Held [32], so I won't repeat it all. The basic idea is that for a conditional autoregression of order p x p, and a sample of size N x N, the total edge effect is of order

$$\text{edge effect} \approx \frac{\text{no. of "edge" points}}{\text{total no. of points}} = \frac{2pN}{N^2} = \frac{2p}{N} \tag{9.23}$$

while the relative variance of estimates of the mean and autoregressive coefficients is of order

$$\begin{aligned} \frac{\text{variance of mean}}{\text{variance of innovations}} &\approx \frac{1}{N} \\ \text{variance of AR coeff} &\approx \frac{1}{N} \end{aligned} \tag{9.24}$$

So we're on pretty safe ground, in the sense that the biases introduced by the circulant approximation and by neglecting the Jacobean term, which is the $|\det L|$ in

$$p(\vec{x}) = \frac{1}{(2\pi v)^{-n/2}} |\det L| \exp -\frac{1}{2v} \vec{x}^T (L^T L) \vec{x} \tag{9.25}$$

are of similar size to the standard error of the parameters we wish to estimate.

For simultaneous autoregressive fields in 1D or 2D, the Jacobean term is *not* negligible, so we need another way of obtaining the maximum likelihood estimates. I cover this in Section 11 and Appendix B. The end result is that for conditional autoregressive fields in 1D or 2D, we can approximate the precision matrix as circulant and estimate the model parameters by working directly with the residual sum of squares. For simultaneous autoregressive fields, we can make the circulant approximation and drop edge terms, but we cannot drop the Jacobean term.

10 Series generation via the circulant approximation

The previous section discussed the magnitude of the circulant approximation; the current section discusses how I use the circulant approximation in order to synthesize an autoregressive series. Suppose that we wish to generate a 1D autoregressive series $x(t)$, of length n, whose model is of the form $L\vec{x} = M\vec{e}$, where $\vec{x}$ is the vector of mean-reduced observations and $\vec{e}$ is the innovation vector. Let L be circulant with base $\vec{l}$, and let M be circulant with base $\vec{m}$. Then, both L and M are diagonalized by the discrete Fourier transform [33]:

$$\begin{aligned} [F]_{km} &= \frac{1}{\sqrt{n}} \exp(-2\pi j \frac{km}{n}), \qquad F^H = F^{-1}, \qquad 0 \le k, m < n \\ L &= F\Lambda_L F^H, \qquad M = F\Lambda_M F^H \\ \Lambda_L &= \sqrt{n}\ \mathrm{diag}(F\vec{l}), \qquad \Lambda_M = \sqrt{n}\ \mathrm{diag}(F\vec{m}) \end{aligned} \tag{10.1}$$

Here, F is the forward discrete Fourier transform, and F^H is the inverse transform. Thus, if we have $L\vec{x} = M\vec{e}$, or

$$\begin{bmatrix} l_1 & l_2 & \cdots & l_n \\ l_n & l_1 & \cdots & l_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ l_2 & l_3 & \cdots & l_1 \end{bmatrix} \vec{x} = \begin{bmatrix} m_1 & m_2 & \cdots & m_n \\ m_n & m_1 & \cdots & m_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ m_2 & m_3 & \cdots & m_1 \end{bmatrix} \vec{e} \tag{10.2}$$

where $\vec{l} = [l_1 \dots l_n]$ and $\vec{m} = [m_1 \dots m_n]$ then we can represent an ARMA model of either conditional or simultaneous type. The inverse transform of the reduced observations is

$$\begin{aligned} \vec{x} &= L^{-1}M\vec{e} = (F\Lambda_L F^H)^{-1}(F\Lambda_M F^H)\vec{e} = F{\Lambda_L}^{-1}\Lambda_M F^H \vec{e} \\ \Longrightarrow (F^H\vec{x}) &= {\Lambda_L}^{-1}\Lambda_M (F^H\vec{e}) = \left[\mathrm{diag}(F\vec{m} \oslash F\vec{l})\right](F^H\vec{e}) \\ &= (F\vec{m} \oslash F\vec{l}) \odot (F^H\vec{e}) \end{aligned} \tag{10.3}$$

where $\oslash$ and $\odot$ denote element-by-element division and multiplication, respectively. Hence

$$\vec{x} = \mathrm{FFT}\left[\left[\mathrm{FFT}(\vec{m}) \oslash \mathrm{FFT}(\vec{l})\right] \odot \mathrm{IFFT}(\vec{e})\right] \tag{10.4}$$

Thus, to generate a vector $\vec{x}$, we need only sample the innovation vector $\vec{e}$, and then apply the preceding formula. For a pure autoregressive series, we just set $\vec{m}$ to be the (1 x n)-dimensional row vector $[1, 0, 0, 0...0]$.

In the case of a two-dimensional random field, the matrices L and M are said to be block-circulant, and have two-dimensional bases l and m, which are the same size as the field itself. A two-dimensional simultaneous AR(1,1) model, for example, with defining equation

$$x(s,t) = \beta_{10}[x(s-1,t) + x(s+1,t)] + \beta_{01}[x(s,t-1) + x(s,t+1)] + e(s,t) \tag{10.5}$$

and which is defined over an N x N grid, would lead to

$$l = \begin{bmatrix} 1 & -\beta_{01} & \cdots & -\beta_{01} \\ -\beta_{10} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\beta_{10} & 0 & \cdots & 0 \end{bmatrix}, \qquad m = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \tag{10.6}$$

The formula, then, for generating a two-dimensional ARMA series is

$$x = \text{FFT2}\left[[\text{FFT2}(m) \oslash \text{FFT2}(l)] \odot \text{IFFT2}(e)\right] \tag{10.7}$$

where e is an N x N generated sample of the innovation field.

11 Parameter estimation for autoregressive models

Let's consider the easier case first, namely that of conditional autoregressive models defined in 1D or 2D. The 1D version has been explored at length elsewhere – see for example Luhtkepohl [34] – so I will just indicate the highlights. Suppose that our model is

$$x(t) = \sum_{j=1}^{p} a_j x(t-j) + e(t), \quad x(t) = y(t) - \mu, \quad e(t) \sim \text{IID} \tag{11.1}$$

and that we have both a "preamble" $\{y(-p+1), y(-p+2), \ldots y(0)\}$ and a sample $\{y(1), y(2), \ldots y(n)\}$. Then we'll define the following quantities:

$$\begin{aligned}
P(z) &= z^p - a_1 z^{p-1} - \ldots - a_{p-1} z - a_p \\
&= (z - r_1)(z - r_2) \ldots (z - r_p) \\
\vec{y} &= [y(1) \ldots y(n)] \quad [\text{ 1 x n }] \\
\vec{e} &= [e(1) \ldots e(n)] \quad [\text{ 1 x n }] \\
\vec{a} &= [a_1 \ldots a_p] \quad [\text{ 1 x p }] \\
\vec{1} &= [1 \ldots 1] \quad [\text{ 1 x n }] \\
Z(\mu) &= \begin{bmatrix} y(0) - \mu & \cdots & y(n-1) - \mu \\ \ldots & \ddots & \ldots \\ y(-p+1) - \mu & \cdots & y(n-p) - \mu \end{bmatrix} \quad [\text{ p x n }] \\
\vec{x} &= \vec{y} - \mu\vec{1} = \vec{a}Z + \vec{e}
\end{aligned} \tag{11.2}$$

Our results of Section 9 above allow us to say that if $n \gg p$, and all the autoregressive roots satisfy $|r_k| < 1 - 1/\sqrt{n}$, then we can ignore edge effects and the Jacobean term, and derive our parameter estimates from the residual sum of squares

$$T = \vec{e}\vec{e}^T = (\vec{x} - \vec{a}Z)(\vec{x}^T - Z^T\vec{a}^T) \tag{11.3}$$

Under those assumptions, the parameter estimates satisfy

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} y(j), \quad \text{and} \quad \left(\frac{1}{n}ZZ^T\right)\hat{a}^T = \frac{1}{n}Z\left(\vec{y} - \hat{\mu}\vec{1}\right)^T \tag{11.4}$$

The quantity $(1/n)ZZ^T$ is an estimate of the covariance of p successive variates of the mean-reduced series:

$$\Gamma = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(-1) & \gamma(0) & \cdots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(-p+1) & \gamma(-p+2) & \cdots & \gamma(0) \end{bmatrix} \tag{11.5}$$

where $\gamma(h) = E[x(t)x(t+h)]$. Setting $\hat{\Gamma} = (1/n)ZZ^T$, our estimate of $\vec{a}$ is now

$$\hat{a}^T = \hat{\Gamma}^{-1}\left[\frac{1}{n}Z(\vec{y} - \hat{\mu}\vec{1})^T\right] \tag{11.6}$$

As for the corresponding variances, applying a series averaging operator to the model equation gives

$$\left(1 - \sum_{j=1}^{p} a_j\right)(\hat{\mu} - \mu) = \frac{1}{n}\sum_{j=1}^{n} e(j) \tag{11.7}$$

and taking the variance gives

$$\mathrm{var}(\hat{\mu} - \mu) = \frac{1}{(1 - \sum a_j)^2}\left[\frac{1}{n}\mathrm{var}[e(t)]\right] \tag{11.8}$$

We can be sure that $\sum a_j \neq 1$ because if that were true, then $z = 1$ would be a root of the characteristic polynomial $P(z) = z^p - a_1 z^{p-1} - \ldots - a_{p-1}z - a_p$, but we have already required that all those roots be less than $1 - 1/\sqrt{n}$ in magnitude. A similar calculation yields

$$\mathrm{cov}(\hat{a}^T - a^T) = \hat{\Gamma}^{-1}\left[\frac{1}{n}\mathrm{var}[e(t)]\right] \tag{11.9}$$

and it is clear from the results of Section 9 and this one that, as long as all the autoregressive roots are less than $1 - 1/\sqrt{n}$ in magnitude, our parameter estimates are asymptotically unbiased and consistent. The 2D version is much the same as the 1D version, so I will not go over it here.

The hard case – the one I will dwell on here – is the case of simultaneous regressions in 1D or 2D. In Appendix B, I present an argument that we can form maximum likelihood estimates of the autoregressive parameters of a simultaneous 1D series by fitting a cosine series to the inverse square root of its geometrically normalized periodogram. That makes a whole lot of sense, because for a model such as

$$x(t) = \sum_{k=1}^{p} \beta_k[x(t+j) + x(t-j)] + e(t), \quad x(t) = y(t) - \mu, \quad e(t) \sim \mathrm{IID} \tag{11.10}$$

the corresponding transfer function is

$$\frac{X(\omega)}{E(\omega)} = \frac{1}{1 - 2\sum \beta_k \cos k\omega} \tag{11.11}$$

where $X(\omega)$ and $E(\omega)$ are the discrete Fourier transforms of the mean-reduced series and innovation sequence, respectively. Denoting their respective spectral densities by $P_{xx}(\omega)$ and $E_{xx}(\omega)$, we have

$$\left[\left|\frac{X(\omega)}{E(\omega)}\right|^2\right]^{-\frac{1}{2}} = \left[\frac{P_{xx}(\omega)}{E_{xx}(\omega)}\right]^{-\frac{1}{2}} = 1 - 2\sum_{k=1}^{p} \beta_k \cos k\omega \tag{11.12}$$

which is indeed a cosine series. Now, the periodogram of a series is a poor estimate of its spectral density; Oppenheim and Schafer discuss this topic at length [35]. So we have to do *some* smoothing and averaging. A good starting point is the observation that the periodogram of a colored Gaussian signal is approximately [36]

$$\begin{aligned}
I_m(\omega) &= P_{xx}(\omega) \bullet [I_m(\omega) \text{ for white noise of unit variance }] \\
\Longrightarrow I_m(\omega) &\approx P_{xx}(\omega) \bullet [\chi^2(1) \text{ variate }] \\
\Longrightarrow \mathrm{mean}[I_m(\omega)] &\approx P_{xx}(\omega) \bullet [\text{mean of } \chi^2(1) \text{ variate}] \\
\Longrightarrow \mathrm{var}[I_m(\omega)] &\approx P_{xx}(\omega) \bullet [\text{variance of } \chi^2(1) \text{ variate}]
\end{aligned} \tag{11.13}$$

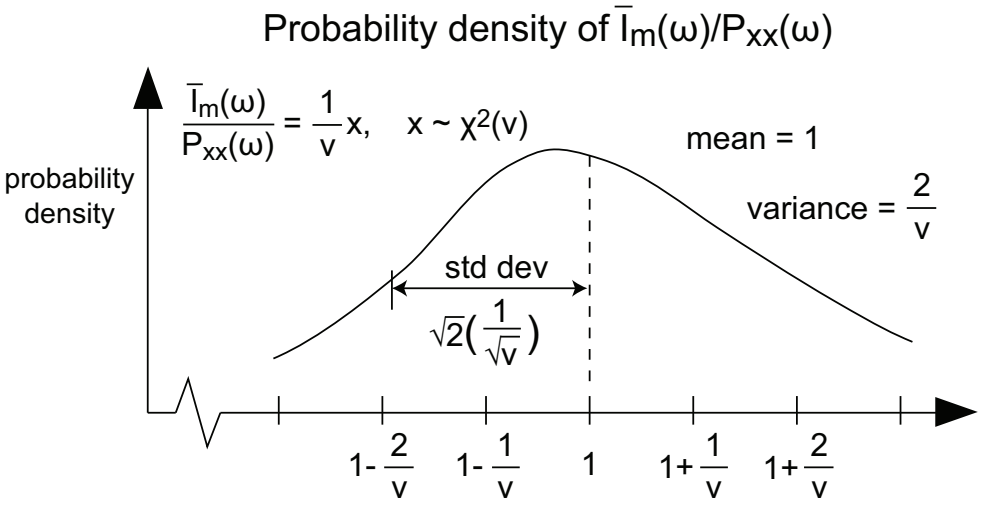

Figure 19: The probability distribution of an averaged periodogram.

where $I_m(\omega)$ is the periodogram of a sample of length m, $P_{xx}(\omega)$ is the true spectral density, each of the m $\chi^2(1)$ variates is independent, and the dot indicates multiplication. Suppose that we average over ν independent periodograms in order to estimate $P_{xx}(\omega)$; then we get

$$\begin{aligned} \bar{I}_m(\omega) &\approx P_{xx}(\omega) \bullet \frac{1}{\nu}[\text{sum of } \nu \quad \chi^2(1) \text{ variates}] \\ &\approx P_{xx}(\omega) \bullet \frac{1}{\nu}[\chi^2(\nu) \text{ variate}] \end{aligned} \tag{11.14}$$

The probability density of a $\chi^2(\nu)$ variate is

$$p(x,\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} \tag{11.15}$$

Using that and Stirling's approximation for $\Gamma(z)$, we find that the mean and variance of $x^{-1/2}$, where x is a $\chi^2(\nu)$ variate, is (to first order in $1/\nu$)

$$\begin{aligned} E[x^{-1/2}] &= \frac{1}{\sqrt{\nu}}(1 + \frac{3}{4\nu}) \\ \text{var}[x^{-1/2}] &= E[x^{-1}] - E^2[x^{-1/2}] = \frac{1}{2\nu^2} \end{aligned} \tag{11.16}$$

which in turn shows that our estimate of $[P_{xx}(\omega)]^{-1/2}$ has a small positive bias. Figures 19 and 20 illustrate this.

What we've got so far is that if we compute ν independent m-point periodograms, and average them, then we get a set $\{F(\omega_1)u_1, F(\omega_2)u_2, \ldots F(\omega_m)u_m\}$ where the u_j's are independent random variables of mean $1 + 3\nu/4$ and variance $1/2\nu$, and

$$\begin{aligned} F(\omega) &= b_0 - 2b_1\cos\omega - 2b_2\cos 2\omega - \ldots - 2b_p\cos p\omega \\ &= [\text{scaling constant}][1 - 2\sum_{k=1}^{p}\beta_k \cos k\omega] \end{aligned} \tag{11.17}$$

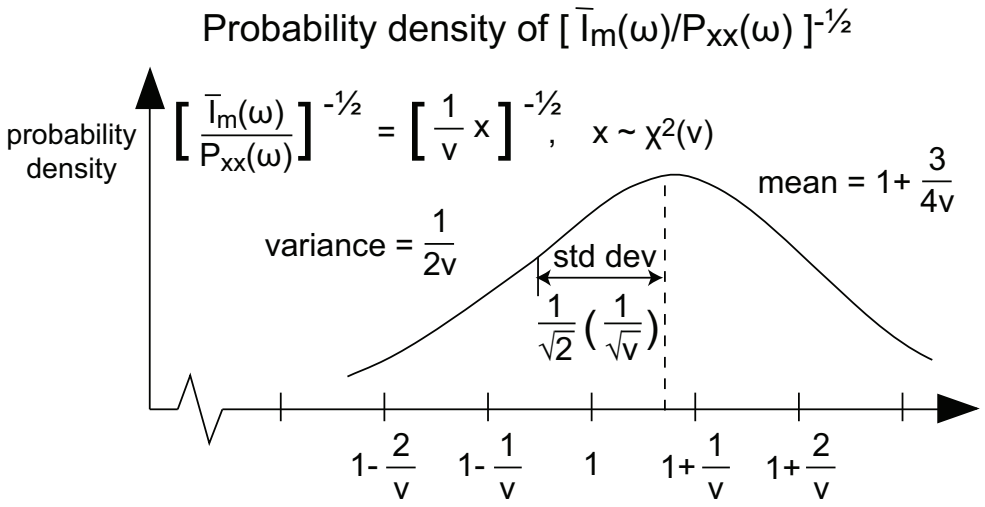

Figure 20: The probability distribution of the inverse square root of an averaged periodogram.

and the ω_j's are equally spaced discrete frequencies. Applying the discrete cosine transform to that set yields biased estimates of the $\{b_j\}$ but unbiased estimates of the $\{\beta_j\}$. As for the estimate variance, I computed that as

$$\begin{aligned} \mathrm{var}[b_0] &= \frac{1}{2\nu m}({b_0}^2 + 2\sum_{k\neq 0}^{p} {b_k}^2) \\ \mathrm{var}[b_j] &= \frac{1}{2\nu m}(\frac{1}{2}{b_0}^2 + \frac{3}{2}{b_j}^2 + \sum_{k\neq j}^{p} {b_k}^2) \end{aligned} \tag{11.18}$$

and consequently

$$\mathrm{var}[\beta_j] \approx \frac{1}{\nu m}(1 + 2{\beta_1}^2 + \ldots + 2{\beta_p}^2) \tag{11.19}$$

The point of all that is simple: if we take a simultaneous autoregressive series of length n, divide it into ν sets of m successive observations, compute the corresponding periodograms, and then fit a cosine series to the inverse square root of the average of those ν periodograms, then the resulting estimates of the autoregressive parameters are unbiased and have a variance of approximately $(1/n)(1 + 2\sum {\beta_k}^2)$.

In order to illustrate the foregoing analysis, I generated simultaneous AR(2) series over the allowable parameter space, and estimated the known parameters β_1 and β_2 two ways: first by using the residual sum of squares (i.e. as if this were a conditional autoregression), and then by computing an averaged periodogram and fitting a cosine series to its inverse square root. In order to estimate the spectral density, I use the Welch method with a Hamming window of size 32. That smoothing method is described further in Oppenheim and Schafer [37].

The results, summarized in Figure 21, show that my method of estimating the autoregressive coefficients of a simultaneous series is pretty good, only really failing when the underlying autoregressive roots r_1 and r_2 are *both* close to 1 or -1.

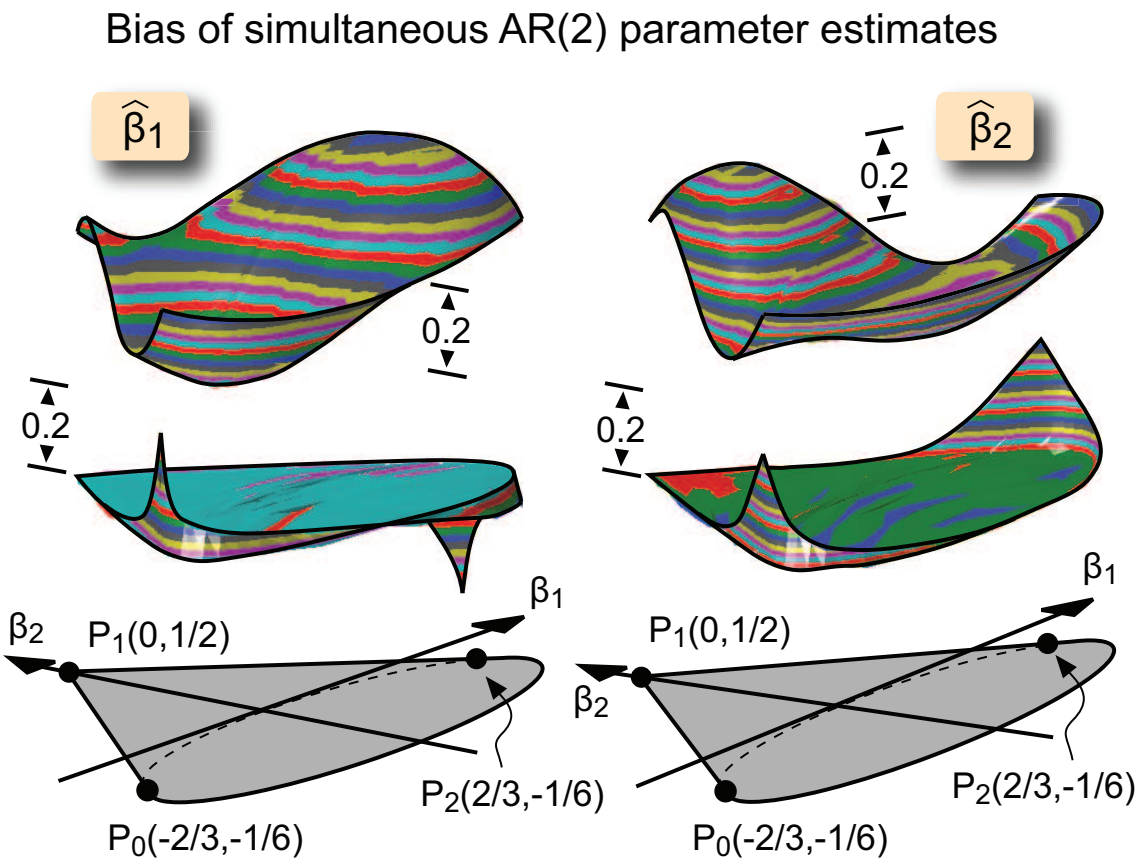

Figure 21: The top row represents the biases obtained by estimating a simultaneous AR(2) series as if it were conditional. The middle row shows the results of my method. The bottom row shows the parameter range under consideration. Each (β_1, β_2) point involves a series of length 2048 and an average over 25 trials.

Another method of estimating the autoregressive coefficients of a series is suggested by the expansion [38]

$$\log\left[\frac{1}{1+r^2-2r\cos\theta}\right] = \sum_{n=1}^{\infty}\frac{2}{n}r^n\cos n\theta \qquad (|r|<1) \tag{11.20}$$

Suppose, for example, that we have a 1D conditional ARMA(2,1) model with discrete transfer function

$$\frac{X(z)}{E(z)} = \frac{1-m_1z^{-1}}{(1-r_1z^{-1})(1-r_2z^{-1})} \tag{11.21}$$

The corresponding spectral density is

$$\left|\frac{X(z)}{E(z)}\right|^2 = \frac{1+{m_1}^2-2m_1\cos\omega}{(1+{r_1}^2-2r_1\cos\omega)(1+r_2^2-2r_2\cos\omega)} \tag{11.22}$$

and taking natural logarithms yields

$$\log\left|\frac{X(z)}{E(z)}\right|^2 = \sum_{n=1}^{\infty}\frac{2}{n}({r_1}^n+{r_2}^n-{m_1}^n)\cos n\omega \tag{11.23}$$

One could thus estimate the spectral density, fit a cosine series to its logarithm, and then choose the autoregressive roots $\{r_1, r_2, m_1\}$ to best reproduce the observed cosine series.

To be more specific, suppose again that we have a simultaneous AR(p) process, and that we compute ν independent periodograms of length m, and average them. Then

$$\bar{I}_m(\omega) \approx P_{xx}(\omega)\bullet\frac{1}{\nu}[\chi^2(\nu)\text{ variate}] \tag{11.24}$$

If x is a $\chi^2(\nu)$ variate, then

$$\begin{aligned} E[\log x] &= \log 2+\Psi(\nu/2) \\ \text{var}[\log x] &= \Psi'(\nu/2) \end{aligned} \tag{11.25}$$

where $\Psi(z)$ and $\Psi'(z)$ are the digamma and trigamma functions, respectively [39]. Using their asymptotic forms for large z gives

$$\begin{aligned} E[\log\bar{I}_m(\omega)] &= \log P_{xx}(\omega)-\frac{1}{\nu} \\ \text{var}[\log\bar{I}_m(\omega)] &= \frac{2}{\nu} \end{aligned} \tag{11.26}$$

and so we see that $\log\bar{I}_m(\omega)$ and $\bar{I}_m(\omega)/P_{xx}(\omega)$ have the same variance. Unfortunately, the negative bias in $E[\log\bar{I}_m(\omega)]$ does not cancel out this time, since we are not taking ratios. Nevertheless, there are some solid advantages to estimating autoregressive roots by this method. First, the target of the discrete cosine transform has uniform variance across its domain. Second, the method estimates the roots directly, as opposed to the quantities $\beta_j = r_j/(1+{r_j}^2)$; this is crucial if any roots are close to one in magnitude. Third, there is no problem with truncation (see Section 7). As long as all the estimated roots satisfy $|r_j|<1$, the product $\prod(1+{r_j}^2-2r_j\cos\omega)$ will be positive definite for all $-\pi\le\omega\le\pi$.

12 The scaling transform in 1D

The basic principle for transformations involving stretch or rotation is that they take place in the continuous autocovariance domain. One of our umbrella assumptions is that the random fields under study are quasi-continuous in nature. Thus, we can posit the existence of a continuous autocovariance function

$$\gamma(\vec{s}) = E[x(\vec{r})x(\vec{r}+\vec{s})], \qquad x(\vec{r}) = y(\vec{r}) - \mu \tag{12.1}$$

and when we choose a particular scale and orientation for our discrete grid, the resulting discrete autocovariance is just a sampling of the continuous one.

As an illustration of this, consider our original example, a 1D conditional AR(1) series, which had

$$\begin{aligned} x(t) &= rx(t-1) + e(t) \qquad x(t) = y(t) - \mu, \quad e(t) \sim \text{IID} \\ \gamma(s) &= E[x(t)x(t+s)] = \frac{r^{|s|}}{1-r^2}\text{var}[e(t)] \qquad (s \text{ integral}) \\ \left|\frac{X(\omega)}{E(\omega)}\right|^2 &= \frac{1}{1+r^2-2r\cos\omega} \end{aligned} \tag{12.2}$$

If we consider the autocovariance as being a sampling from a continuous function $\gamma(u)$, then

$$\frac{1}{\text{var}[e(t)]}\gamma(u) = \frac{r^{|u|}}{1-r^2} \qquad (u \text{ real}) \tag{12.3}$$

and if we sampled that at a new positive spacing h, we would get

$$\frac{1}{\text{var}[e(t)]}\gamma(hs) = \frac{r^{h|s|}}{1-r^2} = \frac{1-r^{2h}}{1-r^2}\left[\frac{r^{h|s|}}{1-r^{2h}}\right] \qquad (s \text{ integral}) \tag{12.4}$$

and the corresponding discrete spectral density is

$$\left|\frac{X(\omega)}{E(\omega)}\right|^2_h = \frac{1-r^{2h}}{1-r^2}\,\frac{1}{1+r^{2h}-2r^h\cos\omega} \tag{12.5}$$

Thus, when we apply a scale change h to a conditional AR(1) process, we get a new AR(1) process, whose autoregressive root is r^h and whose innovation variance is $(1-r^{2h})/(1-r^2)$ times the original one.

Well, the simplicity stops there. Suppose now that we have a conditional AR(2) process with

$$x(t) = (r_1+r_2)x(t-1) - r_1r_2x(t-2) + e(t), \qquad x(t) = y(t) - \mu, \quad e(t) \sim \text{IID} \tag{12.6}$$

then the spectral density is (see Appendix A)

$$\begin{aligned} \left|\frac{X(\omega)}{E(\omega)}\right|^2 &= \frac{1}{(1+{r_1}^2-2r_1\cos\omega)(1+{r_2}^2-2r_2\cos\omega)} \\ &= \frac{1}{(r_1-r_2)(1-r_1r_2)}\left[\frac{r_1}{1+{r_1}^2-2r_1\cos\omega} - \frac{r_2}{1+{r_2}^2-2r_2\cos\omega}\right] \end{aligned} \tag{12.7}$$

and applying a scale change h gives

$$\left|\frac{X(\omega)}{E(\omega)}\right|_h^2 = \frac{1}{(r_1-r_2)(1-r_1r_2)} \bullet \left[\frac{r_1}{1-{r_1}^2}\frac{1-{r_1}^{2h}}{1+{r_1}^{2h}-2r_1^h\cos\omega} - \frac{r_2}{1-{r_2}^2}\frac{1-{r_2}^{2h}}{1+{r_2}^{2h}-2r_2^h\cos\omega}\right] \tag{12.8}$$

Consequently,

$$(r_1-r_2)\quad(1-r_1r_2)\quad \left|\frac{X(\omega)}{E(\omega)}\right|_h^2 = \frac{T_0+T_1\cos\omega}{(1+{r_1}^{2h}-2{r_1}^h\cos\omega)(1+{r_2}^{2h}-2{r_2}^h\cos\omega)} \tag{12.9}$$

where

$$\begin{aligned} T_0 &= r_1\left(\frac{1-r_1^{2h}}{1-{r_1}^2}\right)(1+{r_2}^{2h}) - r_2\left(\frac{1-r_2^{2h}}{1-{r_2}^2}\right)(1+{r_1}^{2h}) \\ \frac{1}{2}T_1 &= r_2{r_1}^h\left(\frac{1-r_2^{2h}}{1-{r_2}^2}\right) - r_1{r_2}^h\left(\frac{1-r_1^{2h}}{1-{r_1}^2}\right) \end{aligned} \tag{12.10}$$

Choosing α so that $2\alpha/(1+\alpha^2) = T_1/T_0$, the spectral density after scaling is

$$\left|\frac{X(\omega)}{E(\omega)}\right|_h^2 = [\text{const}]\frac{1+\alpha^2+2\alpha\cos\omega}{(1+{r_1}^{2h}-2{r_1}^h\cos\omega)(1+{r_2}^{2h}-2{r_2}^h\cos\omega)} \tag{12.11}$$

which shows that the scaled process is no longer AR(2) but ARMA(2,1) with autoregressive roots ${r_1}^h$ and ${r_2}^h$, and a moving average root $-\alpha$. Figure 22 shows the results that I got by trying that case with conjugate roots $r_1 = re^{+j\theta}, r_2 = re^{-j\theta}$.

Clearly, scaling preserves a partial fraction expansion, so we can say that a conditional ARMA(p,$p-1$) process remains an ARMA(p,$p-1$) process under scaling.

What about simultaneous autoregressions? Our basic simultaneous AR(1) process was

$$(1+r^2)x(t) = rx(t-1)+rx(t+1)+(1-r^2)e(t), \quad x(t) = y(t)-\mu, \quad e(t) \sim \text{IID} \tag{12.12}$$

and the corresponding discrete spectral density is

$$\begin{aligned} \left|\frac{X(\omega)}{E(\omega)}\right|^2 &= \left[\frac{1-r^2}{1+r^2-2r\cos\omega}\right]^2 \\ &= (1-r^2)(1+r\frac{\partial}{\partial r})\left[\frac{1}{1+r^2-2r\cos\omega}\right] \end{aligned} \tag{12.13}$$

Subjecting that to a scaling transform gives

$$\frac{1}{1-r^2}\left|\frac{X(\omega)}{E(\omega)}\right|_h^2 = (1+r\frac{\partial}{\partial r})\left[\frac{1-r^{2h}}{1-r^2}\frac{1}{1+r^{2h}-2r^h\cos\omega}\right] \tag{12.14}$$

and working all that out gives

$$\begin{aligned} \left|\frac{X(\omega)}{E(\omega)}\right|_h^2 &= \frac{1}{1-r^2}\frac{T_0+T_1\cos\omega}{(1+r^{2h}-2r^h\cos\omega)^2} \\ T_0 &= 1+r^2-4hr^{2h}+4hr^{2h+2}-r^{4h}-r^{4h+2} \\ \frac{1}{2r^h}T_1 &= (h-1)-(h+1)r^2+(h+1)r^{2h}-(h-1)r^{2h+2} \end{aligned} \tag{12.15}$$

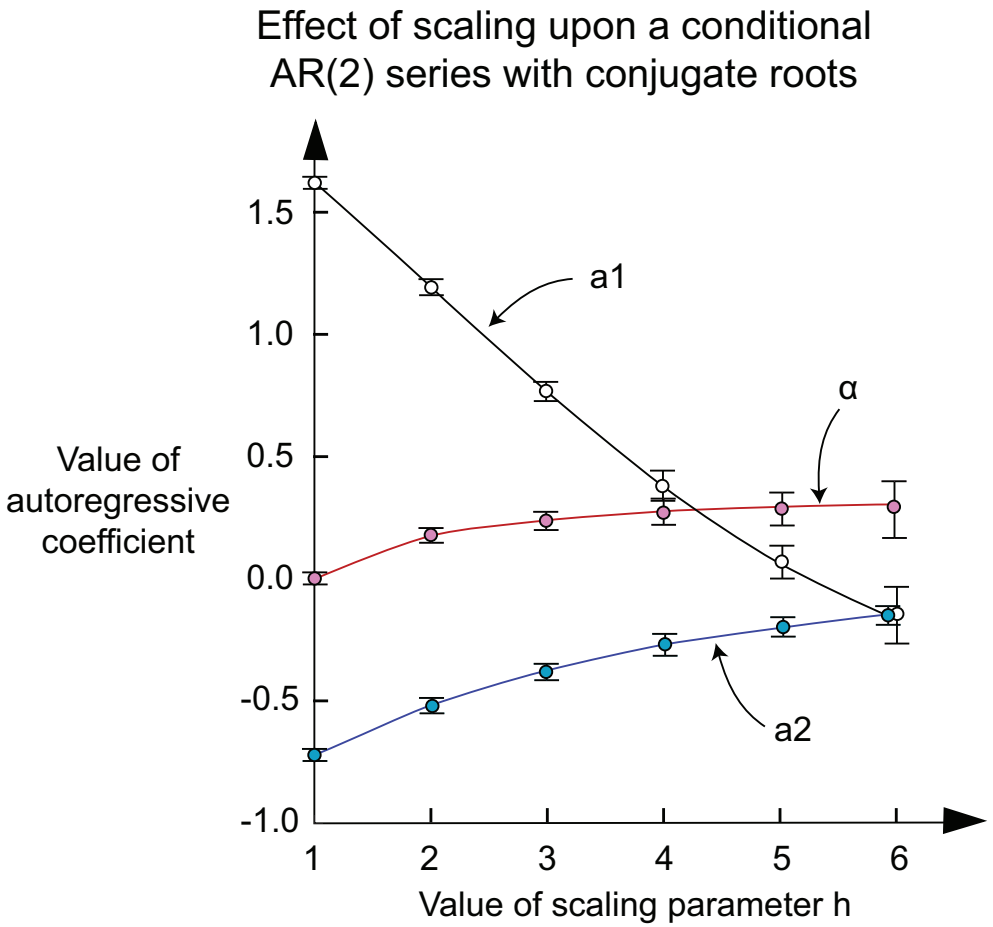

Figure 22: Theoretical and experimental results for the scaling transformation of a conditional AR(2) series that starts out with two conjugate roots defined by $r = 0.85, \theta = 0.3$ at $h = 1$. The series length is 2048, and the chart shows the 95% confidence intervals obtained over 25 trials. The solid curves represent the theoretical results of Equation 12.11.

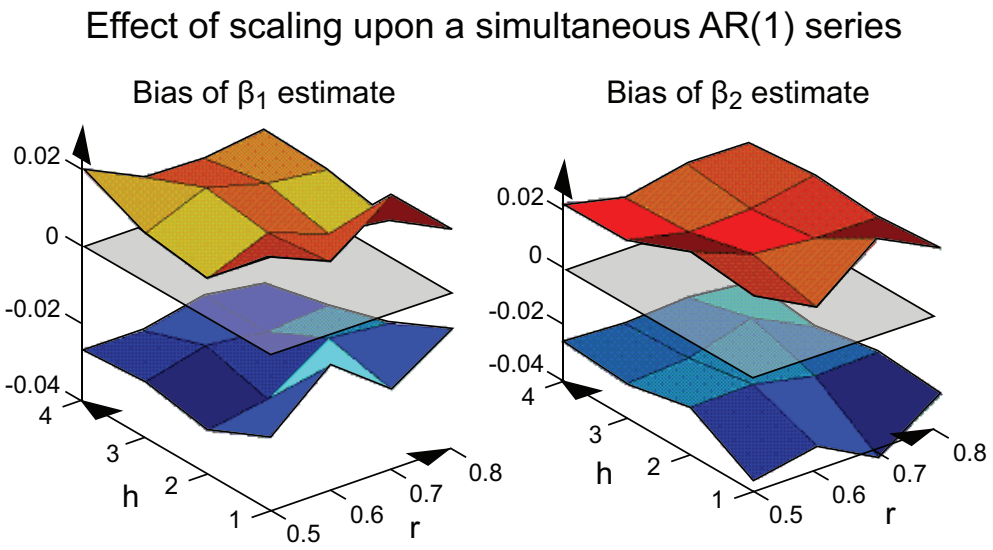

Figure 23: These charts represent the difference between the experimental and theoretical values of β_1 and β_2 for a scaled simultaneous AR(1) series. The theoretical values come from Equation 12.17. The estimates are 95% confidence intervals over 25 trials, each involving a series of length 2048 and a Hamming window of length 32.

The scaled spectral density is no longer a perfect square, so the corresponding series is not a finite simultaneous one. However, if r is not close to 1 and h is not near zero, then T_1/T_0 is pretty small, and we can approximate the scaled series as a simultaneous AR(2) as follows. First, let $a = T_1/(2T_0)$, and $\beta = r^h/(1+r^{2h})$. Then

$$\left[\frac{1}{T_0}(1-r^2)(1+r^{2h})^2\left|\frac{X(\omega)}{E(\omega)}\right|_h^2\right]^{-\frac{1}{2}} = (1-2\beta\cos\omega)(1-2a\cos\omega)^{-\frac{1}{2}} \tag{12.16}$$

which, to third order in a and second order in $\cos\omega$ is

$$\begin{aligned}\left[\left|\frac{X(\omega)}{E(\omega)}\right|_h^2\right]^{-\frac{1}{2}} &= [\text{const}][(1+a\beta+\frac{3}{4}a^2)- \\ &\quad 2(\beta+\frac{1}{2}a+\frac{9}{8}a^2\beta+\frac{15}{16}a^3)\cos\omega+2(\frac{1}{2}a\beta+\frac{3}{8}a^2)\cos 2\omega]\end{aligned} \tag{12.17}$$

Figure 23 shows the results that I got for some typical combinations of r and h. Our theory is definitely holding up so far!

The foregoing analysis suggests that if we sample a conditional AR(p) process at a sufficiently fine scale, then it will appear to be a random walk. Let the partial fraction expansion for its spectral density be

$$\left|\frac{X(\omega)}{E(\omega)}\right|^2 = \sum_{k=1}^{p}\frac{d_k}{1+{r_k}^2-2r_k\cos\omega} \tag{12.18}$$

Then, a scale transformation will result in

$$\left|\frac{X(\omega)}{E(\omega)}\right|_h^2 = \sum_{k=1}^{p} d_k\frac{1-{r_k}^{2h}}{1-{r_k}^2}\frac{1}{1+{r_k}^{2h}-2r_k^h\cos\omega} \tag{12.19}$$

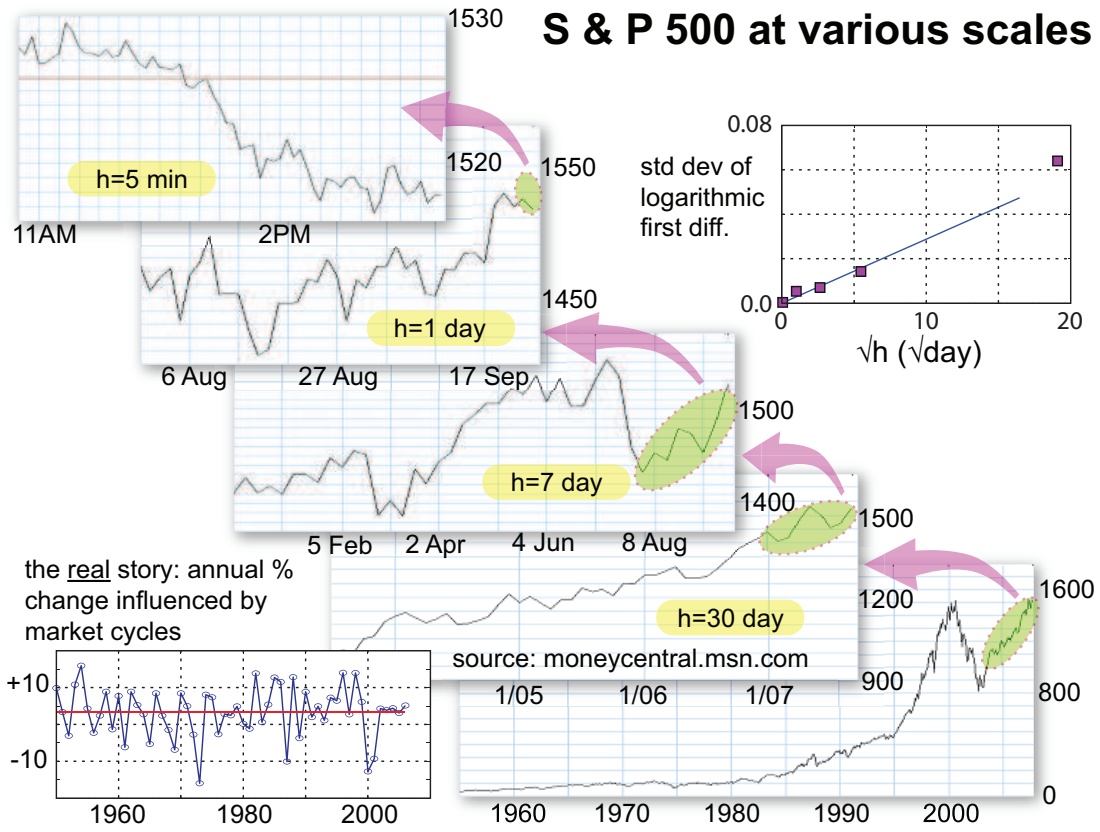

Figure 24: The Standard and Poor 500 index tracks the North American economy. At very fine scales, it looks like a random walk, with an innovation variance that is proportional to the scale. At a very coarse scale, the annual growth would be uncorrelated.

and taking the limit as $h \to 0$ gives

$$\left|\frac{X(\omega)}{E(\omega)}\right|_h^2 = \left[2h \sum_{k=1}^{p} \frac{d_k(-\log r_k)}{1 - {r_k}^2}\right] \left[\frac{1}{2 - 2\cos\omega}\right] \tag{12.20}$$

which is the discrete spectral density of a random walk, whose innovation variance is proportional to the scale h. Figure 24 shows an illustration of that effect, based on the Standard & Poor 500 financial index [40].

Now, the S&P 500 index comprises a broad spectrum of companies with large capitalization, so in effect it tracks the entire North American economy. As such, its autoregressive structure will be at the annual scale, not at daily or hourly scales. I should also point out that this kind of series is more accurately modelled as an ARCH process, that is, one in which the innovation variance is not constant but has a random variation of its own [41]. That makes sense for a financial index since the individual stock transactions that drive it are made by investors with widely diverging tolerance for risk.

One last note about scaling transforms: this theory does allow the concept of a multiresolution texture. Suppose, for example, that we have the situation illustrated in Figure 25. At resolution level 2, for example, r_1 is inactive, we see the structure caused by r_2 and r_3, while r_4 and r_5 create a slowly-varying substrate for the structure that we do see.

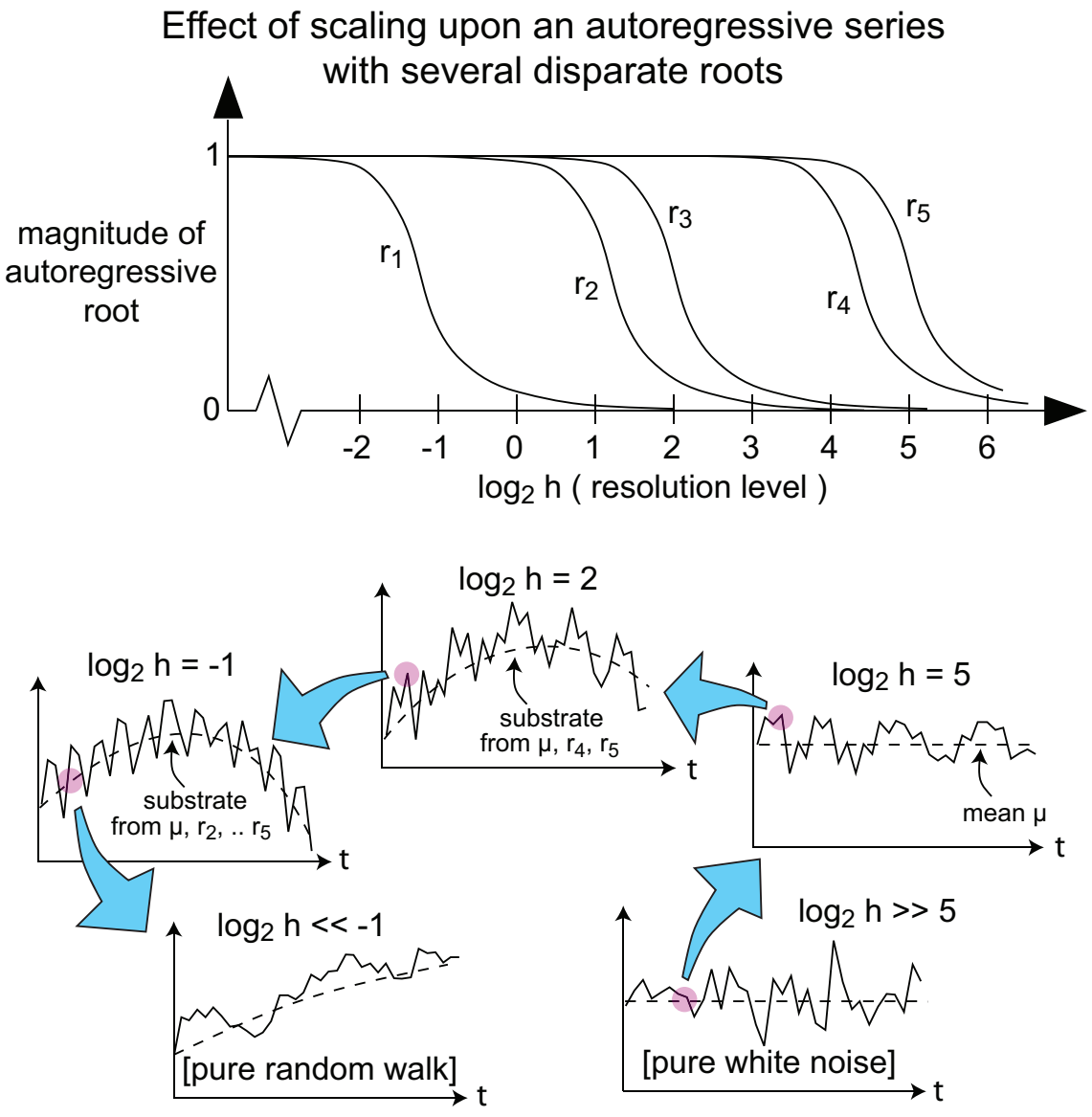

Figure 25: An autoregressive process with several roots can produce different structures at different resolutions.

13 Nature of the continuous domain

Let's review momentarily the moving average representations of one-dimensional AR(1) processes (Sections 5 and 6). For a conditional AR(1) series, we had

$$\begin{aligned} x(t) &= rx(t-1)+e(t), \qquad x(t)=y(t)-\mu, \quad e(t)\sim \text{IID} \\ x(t) &= \sum_{j=0}^{\infty} r^j e(t-j) = \sum_{j=-\infty}^{t} r^{t-j} e(j) \end{aligned} \tag{13.1}$$

whereas for a simultaneous AR(1) series, we had

$$\begin{aligned} (1+r^2)x(t) &= r[x(t-1)+x(t+1)]+(1-r^2)e(t) \\ & \qquad x(t)=y(t)-\mu, \quad e(t)\sim \text{IID} \\ x(t) &= e(t)+\sum_{j=1}^{\infty} r^j[e(t-j)+e(t+j)] = \sum_{j=-\infty}^{+\infty} r^{|t-j|}e(j) \end{aligned} \tag{13.2}$$

In both cases, the operational structure is the same:

$$\begin{bmatrix}\text{discrete}\\ \text{difference}\\ \text{operator}\end{bmatrix} x(t)=e(t); \quad x(t)=\begin{bmatrix}\text{discrete}\\ \text{summation}\\ \text{operator}\end{bmatrix} e(t) \tag{13.3}$$

and this structure continues to hold for AR(p) series. Suppose, for example, that we sample a temporal conditional autoregressive field at an integral multiple of its intrinsic resolution, and that its transfer function at this resolution is

$$\begin{aligned} \frac{X(z)}{E_1(z)} &= (1-\sum_{k=1}^{p} a_k z^{-k}) \qquad \text{with} \\ z^{-p}D(z) &= 1-\sum_{k=1}^{p} a_k z^{-k} = \prod_{k=1}^{p}(1-r_k z^{-1}) \end{aligned} \tag{13.4}$$

Then, the moving average equivalent is

$$x(t_1) = \sum_{j=-\infty}^{t_1} \left[\sum_{k=1}^{p} \frac{{r_k}^{p-1}}{D'(r_k)} {r_k}^{t_1-j}\right] e_1(j) \tag{13.5}$$

which is in the form of a discrete summation operator:

$$x(t_1) = \sum_{j=-\infty}^{t_1} G_1(t_1-j)e_1(j) \quad \text{[resolution of interest]} \tag{13.6}$$

In Section 9, I presented an argument that this form of moving average equivalent is also valid at the smaller intrinsic resolution:

$$x(t_2) = \sum_{j=-\infty}^{t_2} G_2(t_2-j)e_2(j) \quad \text{[intrinsic resolution]} \tag{13.7}$$

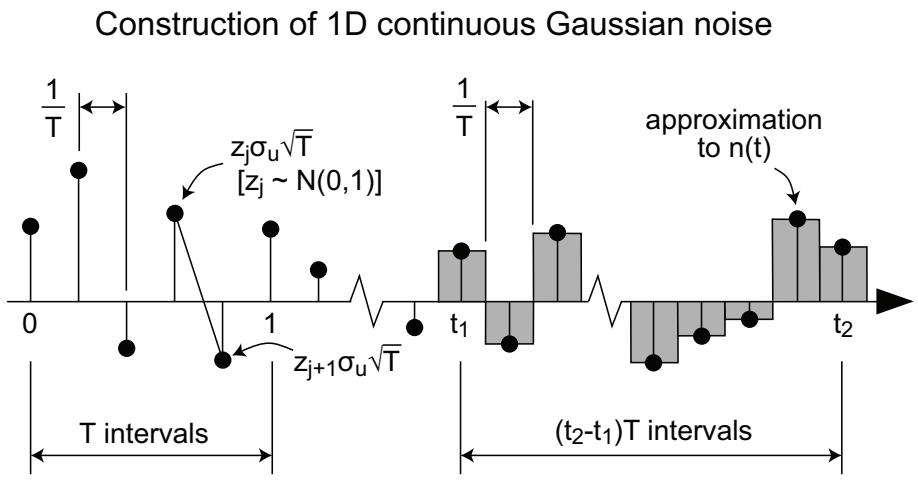

Figure 26: A construction procedure for continuous Gaussian noise.

The requirement of stationarity means that these two representations must both hold when t_1 and t_2 denote the same field point, and so $e_1(t_1)$ must be a linear combination of the neighboring $\{e_2(t_2)\}$. So, regardless of the probability distribution of the $\{e_2(t_2)\}$, the $\{e_1(t_1)\}$ will have an approximately Gaussian probability distribution. If we are going to look at continuous random fields, then, we need the concept of a continuous Gaussian noise field.

This topic has been thoroughly explored elsewhere [42], so I will limit myself to the highlights. To construct quasi-continuous noise in 1D, divide each unit in the interval of interest into T equal subintervals, and place a $N(0, T{\sigma_u}^2)$ Gaussian variate at each point. Figure 26 illustrates this procedure.

The derivative

$$\frac{(z_{j+1} - z_j)\sigma_u\sqrt{T}}{(1/T)} = (z_{j+1} - z_j)\sigma_u T^{\frac{3}{2}} \tag{13.8}$$

has mean zero and variance $2{\sigma_u}^2T^3$, while the "integral"

$$\int_{t_1}^{t_2} n(t)\mathrm{d}t \simeq \sum_{j=1}^{(t_2-t_1)T} (z_j\sigma_u\sqrt{T})(\frac{1}{T}) \tag{13.9}$$

has zero mean and variance $(t_2 - t_1){\sigma_u}^2$. Taking the limit as $T \to \infty$ and setting ${\sigma_u}^2 = 1$ gives us continuous unit Gaussian noise $n(t)$. Although continuous unit Gaussian noise and its derivatives are unbounded, its integral has two properties of interest:

$$\begin{aligned} \mathrm{var}\left[\int_{t_1}^{t_2} n(t)\mathrm{d}t\right] &= (t_2 - t_1) \\ \mathrm{cov}\left[\int_{t_1}^{t_2} n(t)\mathrm{d}t, \int_{t_3}^{t_4} n(t)\mathrm{d}t\right] &= 0 \text{ if } [t_1, t_2] \cap [t_3, t_4] = \emptyset \end{aligned} \tag{13.10}$$

That last equation merely states that distinct increments of the integrated continuous noise are independent.

Now, going back to our conditional AR(1) process, we might expect the discrete difference and discrete summation operators to become differential and integral operators

when the process is viewed as a continuous random field. Indeed, if we have a series

$$\begin{aligned} x(t) &= rx(t-1)+e(t), \\ & x(t)=y(t)-\mu, \quad e(t)\sim \text{IID}, \quad \text{var}[e(t)]=\sigma^2 \end{aligned} \tag{13.11}$$

with autocovariance

$$\gamma(s)=\frac{r^{|s|}}{1-r^2}\sigma^2 \qquad [s \text{ integral}] \tag{13.12}$$

we can imagine that autocovariance to be a sampling from a continuous function, in which case

$$\gamma(u)=\frac{r^{|u|}}{1-r^2}\sigma^2 \qquad [u \text{ real}] \tag{13.13}$$

and the corresponding continuous spectral density is

$$|G(\omega)|^2=\int_{-\infty}^{+\infty}\gamma(u)e^{-j\omega u}\mathrm{d}u=\frac{2\sigma^2}{1-r^2}\frac{p}{p^2+\omega^2} \tag{13.14}$$

where $p=-\log(r)$. At this point, we need to choose a phase for the transfer function $G(\omega)$. I will explore this whole question further in Section 18, but for now let's choose a causal transfer function:

$$G(\omega)=\frac{X(\omega)}{N(\omega)}=\sigma\sqrt{\frac{2p}{1-r^2}}\frac{1}{p+j\omega} \tag{13.15}$$

That then implies a differential equation

$$(p+\frac{\mathrm{d}}{\mathrm{d}t})x(t)=\sigma\sqrt{\frac{2p}{1-r^2}}n(t) \tag{13.16}$$

with corresponding particular solution

$$x(t)=\sigma\sqrt{\frac{2(-\log r)}{1-r^2}}\int_{-\infty}^{t}r^{t-t'}n(t')\mathrm{d}t' \tag{13.17}$$

If we now look at that solution for two values of t, separated by a distance h, we get the situation illustrated in figure 27.

By "envelope" for $x(t)$, I mean the function that multiplies $n(t')$ in the equation

$$x(t)=\int_{-\infty}^{t}g(t-t')n(t')\mathrm{d}t'$$

From the figure, we can see right away that

$$\begin{aligned} x(t) &= (r^h)x(t-h)+e(t,h) \text{ where} \\ e(t,h) &= \sigma\sqrt{\frac{2(-\log r)}{1-r^2}}\int_{t-h}^{t}r^{t-t'}n(t')\mathrm{d}t' \end{aligned} \tag{13.18}$$

and the corresponding innovation variance is

$$\text{var}[e(t,h)]=\sigma^2\frac{2(-\log r)}{1-r^2}\int_{t-h}^{t}r^{2(t-t')}\mathrm{d}t'=\sigma^2\left(\frac{1-r^{2h}}{1-r^2}\right) \tag{13.19}$$

This is exactly what we deduced in Section 12; upon scaling, a conditional AR(1) process with autoregressive root r and innovation variance σ^2 maintains its AR(1) statistics, but the autoregressive root and innovation variance become r^h and $\sigma^2(1-r^{2h})/(1-r^2)$ respectively. The only difference is that in the continuous case, we assume that the innovations $e(t)$ are Gaussian in nature.

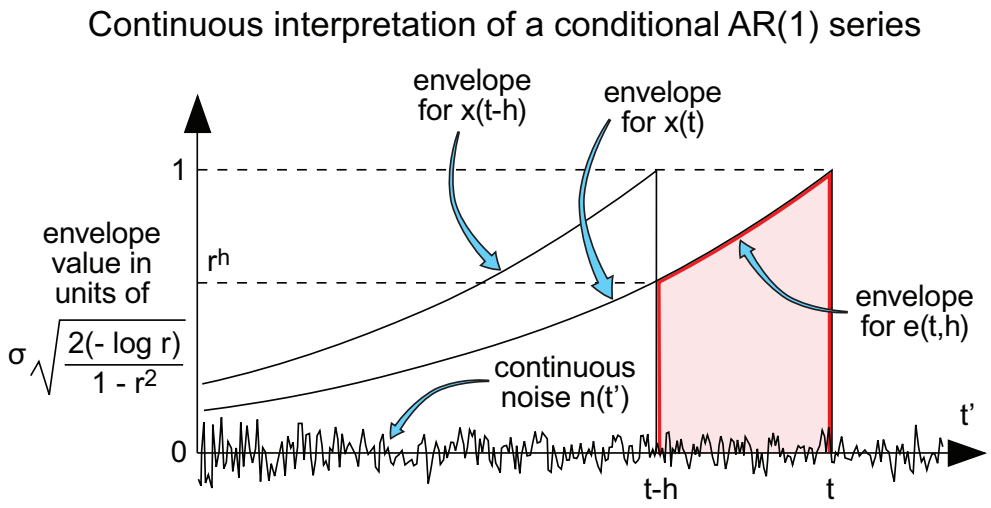

Figure 27: An interpretation of a continuous random field as the convolution of a Gaussian noise function with an envelope.

14 Problems of interpolation

The following few sections are concerned with coordinate transformations that include rotation, and so at times we will be generating discrete random fields and then interpolating between the mesh points. We need to consider what errors, if any, are introduced by the interpolation methods chosen. The plain fact is that interpolation methods can introduce spurious correlations between field points, and can decrease the apparent series variance. The current section is devoted to an exploration of those effects. In Sections 18 and 19, I look at ways to interpolate random fields that preserve their autoregressive characteristics.

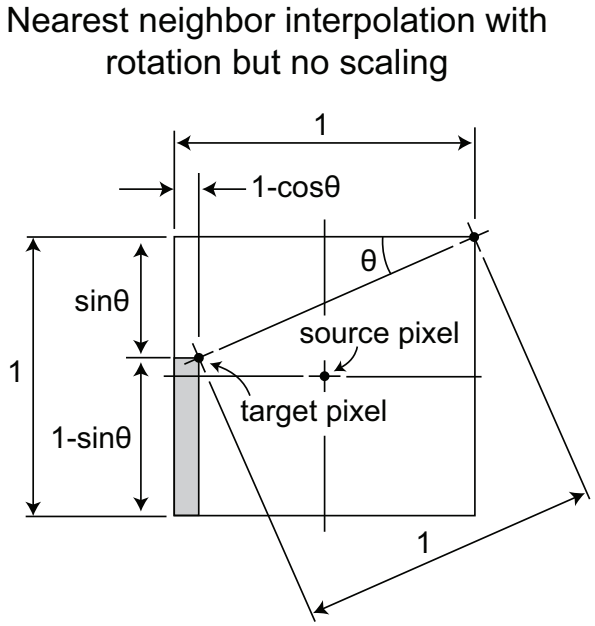

Figure 28: Nearest-neighbor interpolation on a rotated grid will introduce some spurious correlations if there is no scaling.

Suppose that we are doing nearest-neighbor interpolation on a rotated grid, with no scale transform. Although source and target pixels have equal planar densities, some source pixels are used twice and some are unused. Let the inter-pixel distance be h. Then, surrounding each source pixel is a "strike zone" of size h x h, centered on the source pixel. If a target pixel falls within this region, then it assumes the source pixel's value. At rotation angles θ other than zero and ninety degrees, there is a finite probability that two adjacent target pixels will fall within the same "strike zone" and be assigned the same value. Figure 28 illustrates this situation. This, of course, results in a spurious correlation. On a large enough grid, the target points will be approximately evenly distributed, so the spurious correlations introduced will be approximately

$$\frac{\gamma(1,0)}{\gamma(0,0)} = \frac{\gamma(0,1)}{\gamma(0,0)} = (1-\sin\theta)(1-\cos\theta), \quad 0 \le \theta \le \pi/2 \tag{14.1}$$

Figure 29 shows a test of that, in which I generate a two-dimensional Gaussian white noise field, and calculate the $\gamma(1,0)/\gamma(0,0)$ and $\gamma(0,1)/\gamma(0,0)$ correlation coefficients for a rotated grid, using nearest-neighbor interpolation.

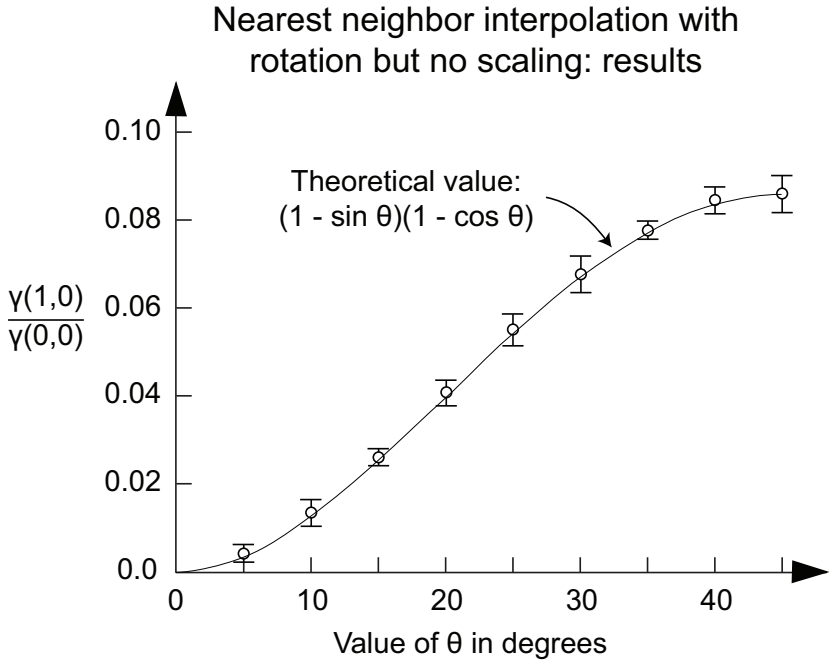

Figure 29: Comparison of theoretical and experimental results for the autocorrelation obtained by nearest-neighbor sampling of a rotated, uncorrelated random field. The original field size is 1024 x 1024, and the sampled field is 512 x 512. The chart shows the 95% confidence intervals over 25 trials.

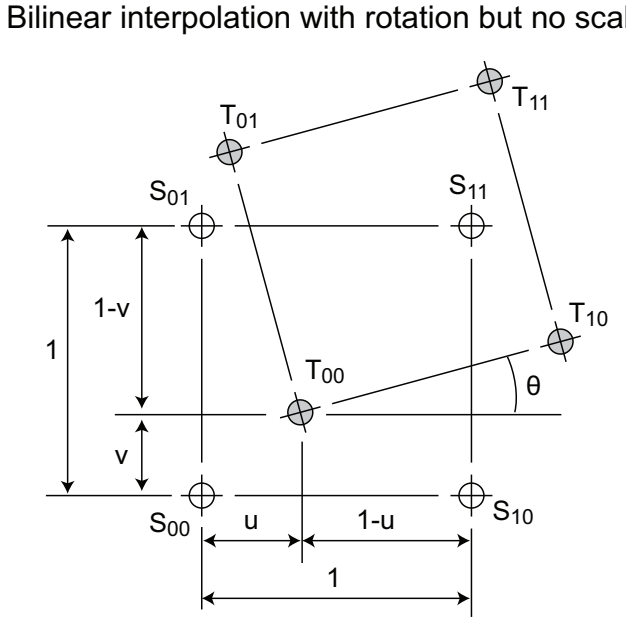

Figure 30: Geometric significance of the quantities involved in bilinear interpolation. The S's are source points, and the T's are target points.

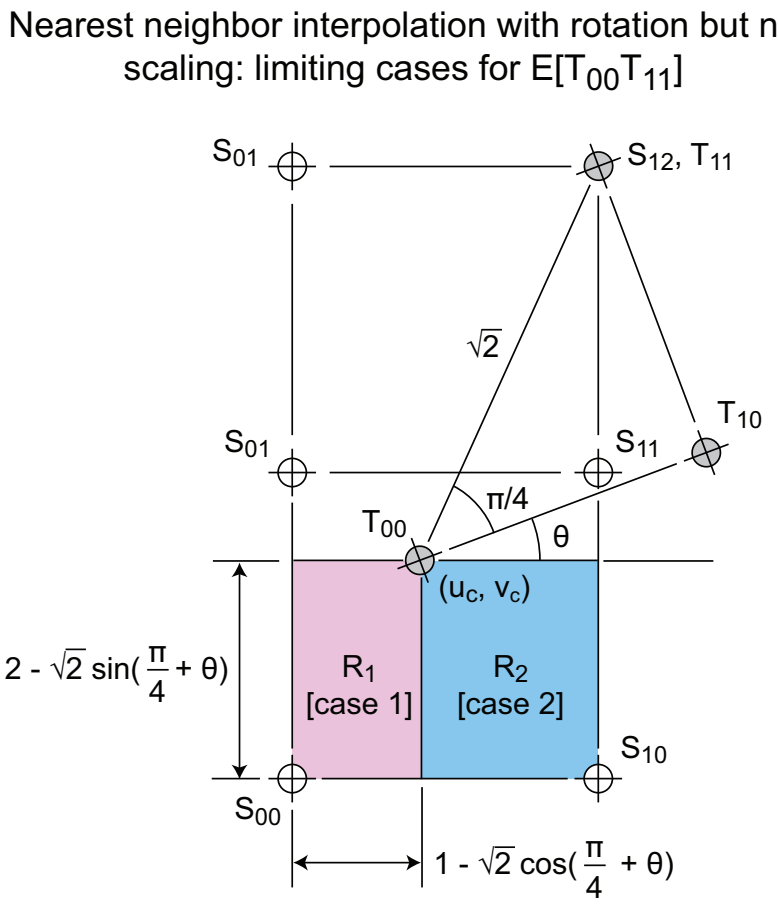

Figure 31: The two cases to consider when calculating $E[T_{00}T_{11}]$

Well, that was an easy case. Suppose now that we are using bilinear interpolation to populate a rotated grid, and that there is no scale transform. Figure 30 illustrates the situation. Here, T_{00} is the field value at the target point, and $\{S_{00}, S_{10}, S_{01}, S_{11}\}$ are the field values at the source points that surround it. Then

$$T_{00} = (1-u)(1-v)S_{00} + (1-u)vS_{01} + u(1-v)S_{10} + uvS_{11} \tag{14.2}$$

If the grid is large enough, then we can average $E[T_{00}T_{00}]$ over the unit square $\{0 \le u, v \le 1\}$ and get

$$E[T_{00}T_{00}] = \frac{4}{9}\gamma(0,0) + \frac{2}{9}[\gamma(1,0) + \gamma(0,1)] + \frac{1}{18}[\gamma(1,1) + \gamma(-1,1)] \tag{14.3}$$

where $\gamma(K,L) = E[S_{00}S_{KL}]$. Thus, if we take an uncorrelated random field and rotate it through bilinear interpolation, we reduce its variance by more than one-half! As for calculations of $E[T_{00}T_{10}]$, $E[T_{00}T_{01}]$ etc, they get pretty complicated, so I will illustrate just one of them – $E[T_{00}T_{11}]$ – in the case where only $\gamma(0,0)$ is non-zero (i.e. the source field is uncorrelated). There are two cases to look at, which I have shown in figure 31.

In the first case, T_{00} and T_{11} have two source points in common, namely S_{01} and S_{11}. In the second case, T_{00} and T_{11} have only one source point in common, S_{11}. We'll look at these two cases in order.

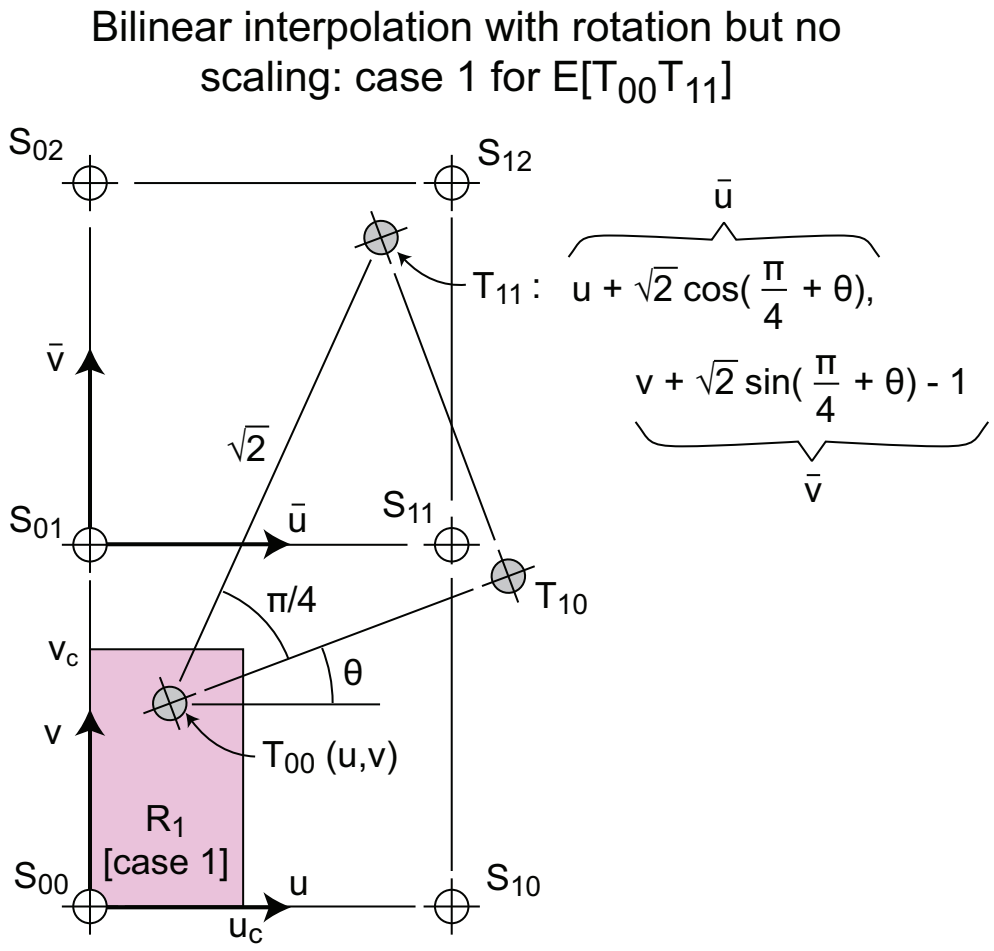

Figure 32: In this case, T_{00} and T_{11} have two source points in common, namely S_{01} and S_{11}.

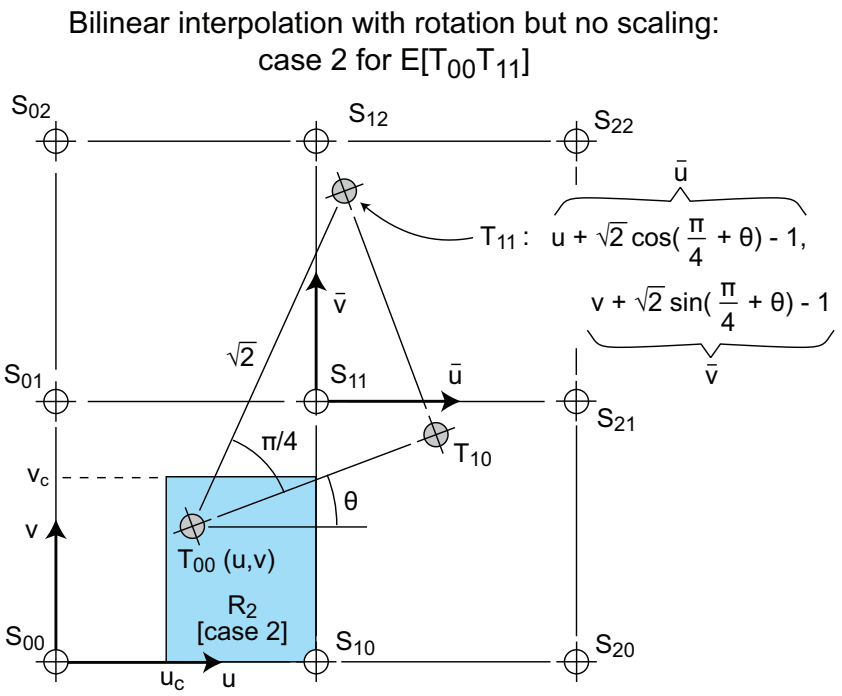

Figure 33: In this case, T_{00} and T_{11} have just one source point in common, namely S_{11}.

In this first case, we have

$$\begin{aligned} T_{00} &= (1-u)(1-v)S_{00} + (1-u)vS_{01} + u(1-v)S_{10} + uvS_{11} \\ T_{11} &= (1-\bar{u})(1-\bar{v})S_{01} + (1-\bar{u})\bar{v}S_{02} + \bar{u}(1-\bar{v})S_{11} + \bar{u}\bar{v}S_{12} \end{aligned} \tag{14.4}$$

If the source field is uncorrelated, so that $E[S_{00}S_{KL}] = \sigma^2\delta_{K0}\delta_{L0}$, then

$$\frac{1}{\sigma^2}E[T_{00}T_{11}] = (1-u)v(1-\bar{u})(1-\bar{v}) + uv\bar{u}(1-\bar{v}) \tag{14.5}$$

The corresponding equations for the second case are:

$$\begin{aligned} T_{00} &= (1-u)(1-v)S_{00} + (1-u)vS_{01} + u(1-v)S_{10} + uvS_{11} \\ T_{11} &= (1-\bar{u})(1-\bar{v})S_{11} + (1-\bar{u})\bar{v}S_{12} + \bar{u}(1-\bar{v})S_{21} + \bar{u}\bar{v}S_{22} \\ \frac{1}{\sigma^2}E[T_{00}T_{11}] &= uv(1-\bar{u})(1-\bar{v}) \end{aligned} \tag{14.6}$$

Integrating over both regions gives

$$\begin{aligned} \frac{1}{\sigma^2}E[T_{00}T_{11}] &= \int_0^{u_c}\int_0^{v_c}(1-u)v[1-u-\sqrt{2}\cos(\tfrac{\pi}{4}+\theta)][2-v-\sqrt{2}\sin(\tfrac{\pi}{4}+\theta)]\mathrm{d}u\mathrm{d}v \\ &+ \int_0^{u_c}\int_0^{v_c} uv[u+\sqrt{2}\cos(\tfrac{\pi}{4}+\theta)-1][2-v-\sqrt{2}\sin(\tfrac{\pi}{4}+\theta)]\mathrm{d}u\mathrm{d}v \\ &+ \int_{u_c}^{1}\int_0^{v_c} uv[2-u-\sqrt{2}\cos(\tfrac{\pi}{4}+\theta)][2-v-\sqrt{2}\sin(\tfrac{\pi}{4}+\theta)]\mathrm{d}u\mathrm{d}v \end{aligned} \tag{14.7}$$

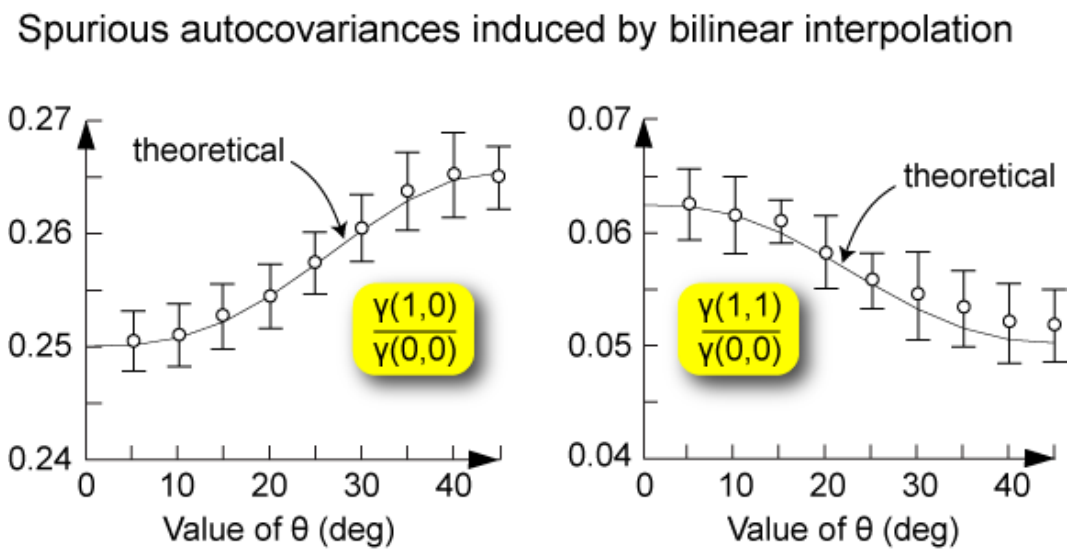

Figure 34: Comparison between theoretical and experimental results for the spurious correlations induced by bilinear interpolation of a rotated uncorrelated random field when there is no scaling. The original field is size 1024 x 1024, the sampled field is size 512 x 512, and the chart shows the 95% confidence intervals over 25 trials.

where

$$u_c = 1 - \sqrt{2}\cos(\frac{\pi}{4} + \theta) \text{ and } v_c = 2 - \sqrt{2}\sin(\frac{\pi}{4} + \theta) \tag{14.8}$$

Similar calculations give $E[T_{00}T_{10}]$, $E[T_{00}T_{01}]$, etc. Figure 34 shows a comparison between those calculations and some experimental results.

So, the bottom line is this: standard methods of interpolation may distort random fields, changing the very correlations that we are trying to study. In Sections 18 and 19, I examine alternative methods of interpolation. For now, however, I will restrict myself to nearest neighbor interpolation.

15 Attempting the rotational transform

A key observation when looking at the transformations of continuous random fields is that the continuous autocovariance function is the closest thing we have to an invariant. Consider the defining equations for the continuous and discrete autocovariance functions:

$$\begin{aligned} \gamma(\vec{u}) &= E[x(\vec{r})x(\vec{r}+\vec{u})], \qquad x(\vec{r}) = y(\vec{r}) - \mu, \qquad \vec{r}, \vec{u} \text{ continuous} \\ \gamma(\vec{s}) &= E[x(\vec{r})x(\vec{r}+\vec{s})], \qquad x(\vec{r}) = y(\vec{r}) - \mu, \qquad \vec{r}, \vec{s} \text{ discrete} \end{aligned} \tag{15.1}$$

where $y(\vec{r})$ is a stationary 2D random field and μ is its mean. As long as the domains of both functions are unbounded, we will get $\gamma(\vec{u}) = \gamma(\vec{s})$ when $\vec{u}$ and $\vec{s}$ coincide. When the discrete domain is bounded, we start to get slight distortion, which comes about in a way which figure 35 illustrates.

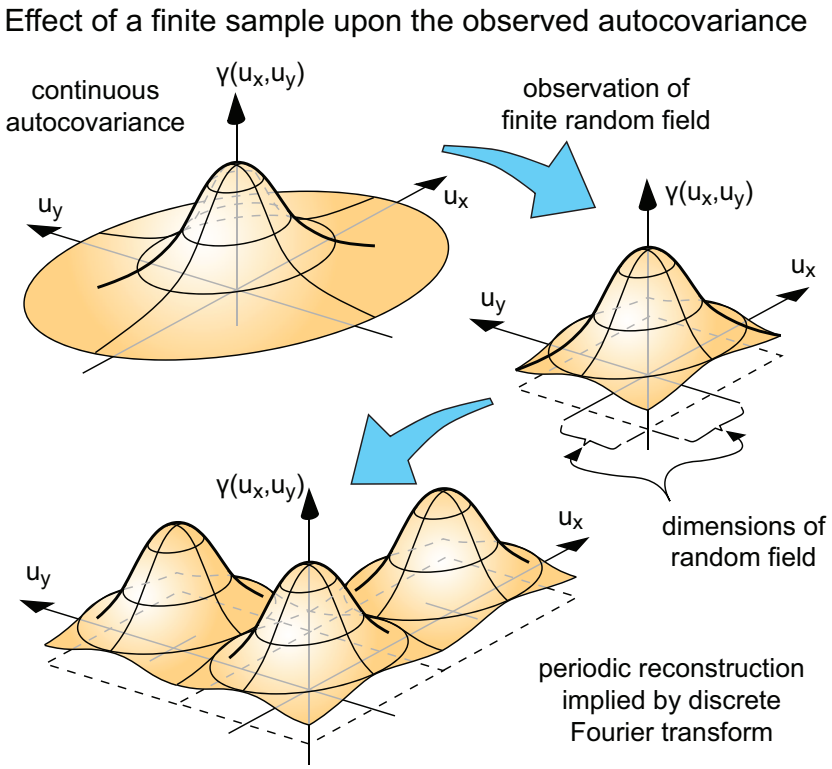

Figure 35: The discrete Fourier transform assumes periodic boundary conditions, which may distort the actual autocovariance function of a random series.

The discrete Fourier transform assumes that the sample has periodic boundary conditions, and the same idea then applies to the autocovariance. This is not an effect of aliasing in the discrete frequency domain. Normally, this is not a big issue because with a large enough sample, the autocovariance drops to insignificant values far inside the sample boundaries.

If, however, the random field is very highly correlated, or if the sample size is small enough, then we lose a substantial portion of the actual autocovariance function. The first situation is rare, especially in the case of visual texture. Our only real concern is when our sample size is small. So let's restrict ourselves to those cases where our resolution

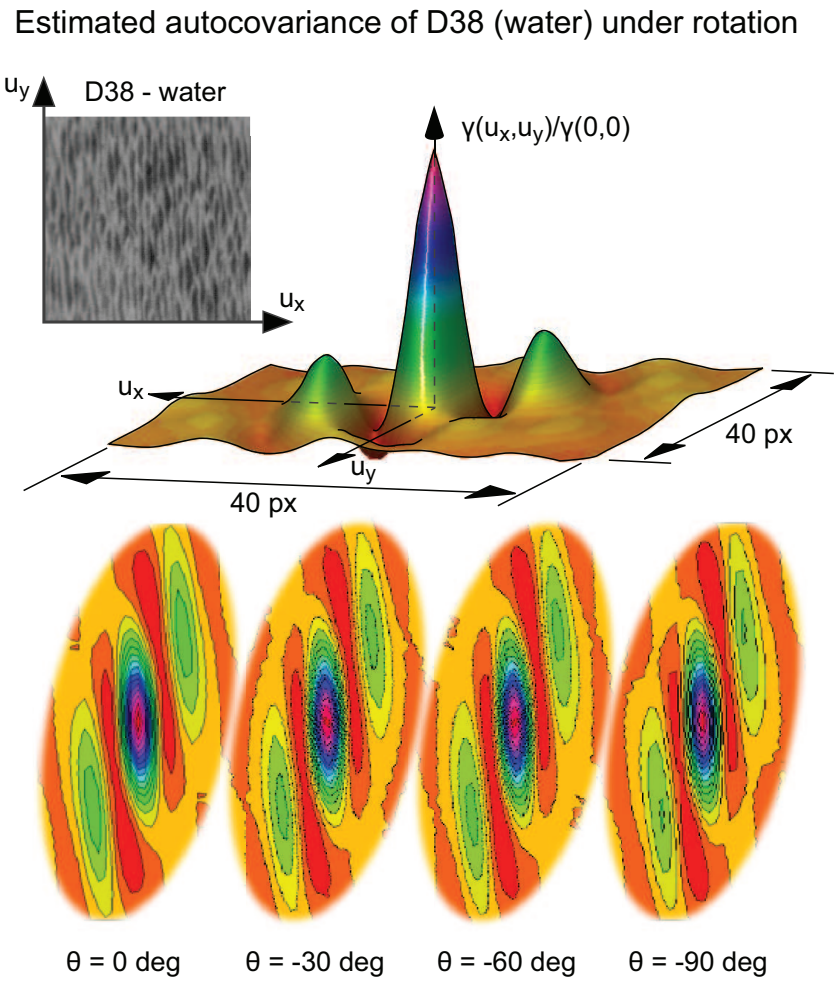

Figure 36: The estimated autocovariance of the Brodatz water texture at several different angles, with no scaling transform. I rotate the results back to the $\theta = 0$ coordinate system for comparison.

and sample size are sufficient to give us an autocovariance that drops to zero well within the sample boundaries. Figure 36, for example, shows the estimated 2D autocorrelation of the Brodatz water texture D38 at several different angles [43]. The source images are not interpolated here; they are digital photographs of the original Brodatz prints, taken at various angles, hence there is no distortion due to interpolation. I use a well-known technique to estimate the autocovariance [44], and rotate the results back to the $\theta = 0$ coordinate system.

These charts illustrate that to a very good approximation, the underlying continuous autocovariance function remains constant. Now consider what happens when we sample a discrete random field, using nearest-neighbor interpolation, in order to estimate the autocovariance of a rotated version of that random field.

If our sample size is large enough, then under coordinate transformation we sample evenly across a unit square centered on the target displacement $\vec{s}$. So the expectation of $\gamma(\vec{s})$ will be the integral of $\gamma(\vec{u})$ over a unit square centered on $\vec{s}$. This will necessarily

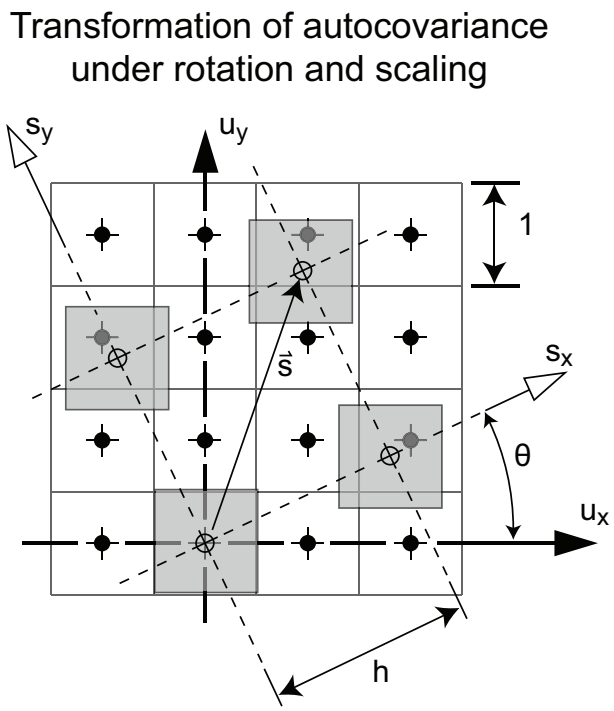

Figure 37: Under nearest neighbor interpolation, the expectation of $\gamma(\vec{s})$ will be the integral of $\gamma(\vec{u})$ over a unit square centered on $\vec{s}$.

cause some distortion if the scale factor h is less than $\sqrt{2}$. Figures 37 and 38 illustrate this idea.

Hence, even when using nearest-neighbor interpolation, we need to introduce a scale factor of at least $\sqrt{2}$ in order to ensure that the autocovariance estimates do not have spurious correlations between them.

As an illustration of these ideas, consider a simulataneous AR(1,1) 2D field with defining equation

$$\begin{aligned} x(s,t) &= \beta_{10}[x(s-1,t)+x(s+1,t)]+\beta_{01}[x(s,t-1)+x(s,t+1)]+e(s,t), \\ & x(s,t)=y(s,t)-\mu, \quad e(s,t)\sim N(0,\sigma^2) \end{aligned} \tag{15.2}$$

which we then sample on a transformed grid via nearest-neighbor interpolation. The theoretical autocovariance is easily calculated from the defining equation. I calculated the theoretical spectral density at various angles and scales through

$$\gamma(\vec{s}) = \int_{u_x-\frac{1}{2}}^{u_x+\frac{1}{2}} \int_{u_y-\frac{1}{2}}^{u_y+\frac{1}{2}} \gamma(u_x,u_y)\mathrm{d}u_x\mathrm{d}u_y \tag{15.3}$$

where (u_x, u_y) is the pre-image of $\vec{s}$, and then taking the discrete Fourier transform of the result. We can compare that to the experimental spectral densities estimated through the methods of Section 11. Figure 39 shows some typical results.

Now, we'll get more adventurous and try this theory on some real data. Two estimates of the spectral density for the Brodatz water texture D38, for a transform with $\theta = -\pi/6$ and $h = 1.4$, appear in Figure 40.

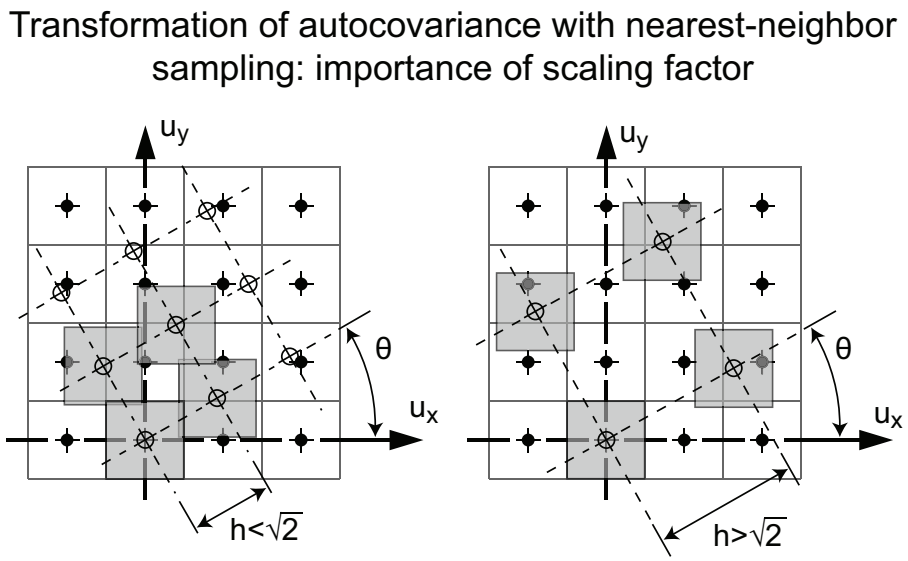

Figure 38: Under nearest neighbor interpolation, there can be significant distortion of the autocovariance if the scale factor is less than $\sqrt{2}$.

Here's how I got those two estimates. For the one on the left in Figure 40, I estimated the autocovariance at $\theta = -\pi/6$ and $h = 1.4$ on a 248 x 248 grid at each of the seven angles for which I have rotated Brodatz texture images. Then, I took the average, took the 2D discrete Fourier transform, and collapsed the result into a 31 x 31 array. For the estimate on the right in Figure 40, I estimated the spectral density directly from the rotated Brodatz image at $\theta = -\pi/6$, sampling with $h = 1.4$ via nearest-neighbor interpolation on a 358 x 358 grid, and applying the Welch method with a Hamming window size of $M = 31$.

If my theory is correct, then these two estimates should match up. They do show remarkable agreement, even though the spectral density varies by more than three orders of magnitude over its range. So our theory looks pretty solid so far. A good question to ask here is, when calculating the theoretical discrete spectral density from the continuous autocovariance, do we transform the coordinate system and then take the discrete Fourier transform, or can we reverse those operations? In other words, do the operations of coordinate transformation and discrete Fourier transform commute? That question may seem inconsequential, but in fact it does matter, and the difference is measurable. So let's take a closer look.

First, suppose we do sampling before rotation. Then we get a spectral density of the form

$$|X(\vec{\omega})|^2 = \frac{1}{\sum \beta(\vec{k}) \exp(j\vec{k}^T\vec{\omega})}|E(\vec{\omega})|^2 = \frac{\sigma^2}{\sum \beta(\vec{k}) \exp(j\vec{k}^T\vec{\omega})} \tag{15.4}$$

The corresponding autocovariance is

$$\gamma(\vec{s}) = \frac{\sigma^2}{4\pi^2}\int_{A1} \frac{\exp(+j\vec{s}^T\vec{\omega})}{\sum \beta(\vec{k}) \exp(j\vec{k}^T\vec{\omega})} \mathrm{d}^2\vec{\omega} \tag{15.5}$$

where $A1$ is the square region over which discrete frequencies are defined in the original coordinate system. This only equals the continuous autocovariance $\gamma(\vec{u})$ at the discrete

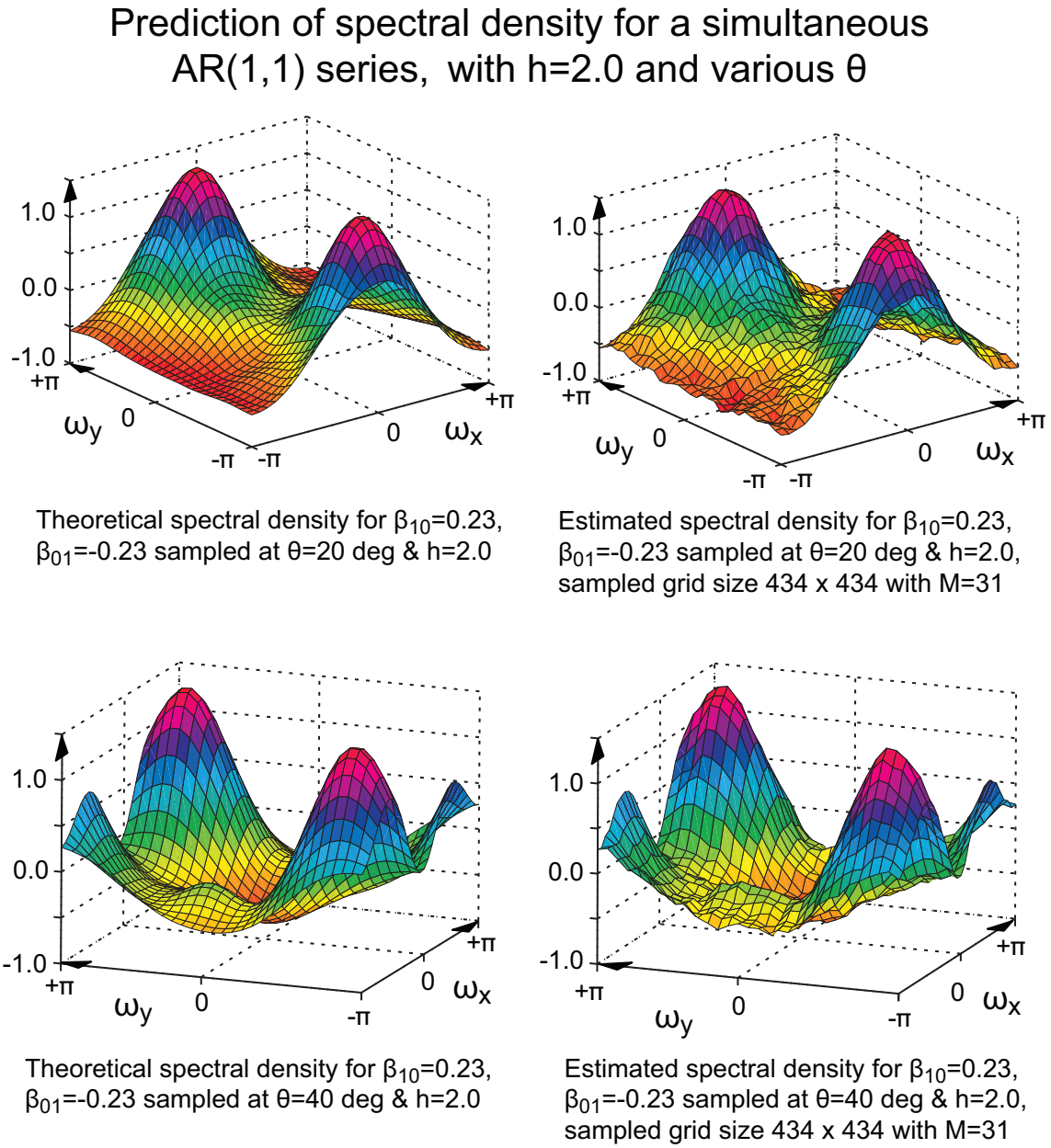

Figure 39: Comparison between experimental and theoretical spectral densities for a synthesized simultaneous AR(1,1) series that is subject to both rotation and scaling. The original field is of size 1300 x 1300. M is the size of the Hamming window used in estimating the spectral density. The spectral densities are scaled so that their average logarithm is zero.

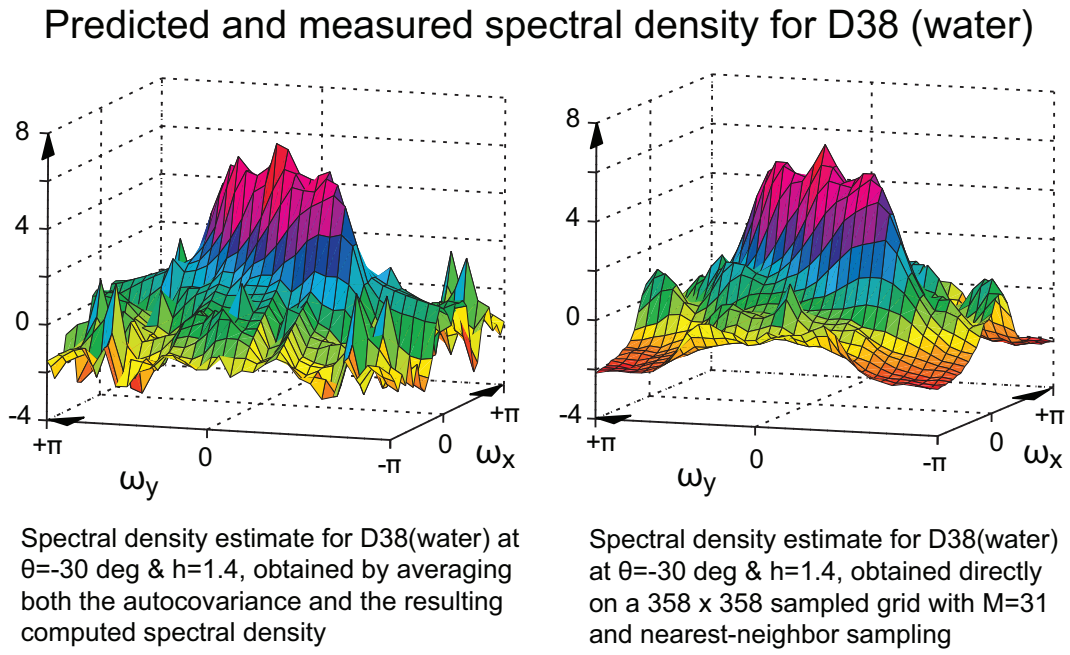

Figure 40: Two different ways of estimating the spectral density of the Brodatz water texture at $\theta = -\pi/6$ and $h = 1.4$. The vertical scale is logarithmic, and the estimates are scaled so that their average logarithm is zero. The text has further details.

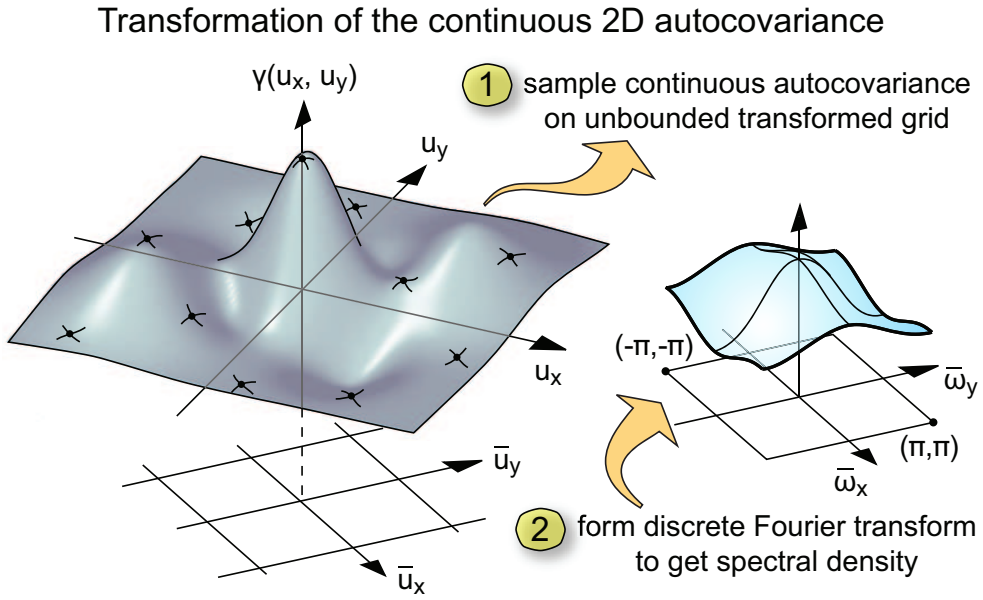

Figure 41: In this representation of a random field transform, we perform the rotation first, then the sampling. The order of operations is important.

grid points. In between, it is a blending of all the autocovariance samples [45]:

$$\gamma_s(\vec{u}) = \sum_{\text{grid}} \gamma(\vec{s})\text{sinc}\pi(s_x - u_x)\text{sinc}\pi(s_y - u_y) \tag{15.6}$$

Clearly, then, when we perform a rotation, we get a possibly distorted version of $\gamma(\vec{u})$:

$$\begin{aligned}\gamma_s(R\vec{u}) &= \sum_{\text{grid}} \gamma(\vec{s})\text{sinc}\pi[s_x - h(u_x\cos\theta + u_y\sin\theta)] \bullet \\ &\qquad \text{sinc}\pi[s_y - h(u_y\cos\theta - u_x\sin\theta)]\end{aligned} \tag{15.7}$$

where I have separated the coordinate transformation R into a scale factor h and a rotation angle θ.

An example will make this idea clear. Suppose that we have a 2D stationary random field that is AR(1) with innovation variance σ^2 and autoregressive parameter r along the x-direction, and is uncorrelated along the y-direction. Then, the continuous 2D autocovariance will be

$$\gamma(u_x, u_y) = \frac{\sigma^2}{1-r^2} r^{|u_x|}\delta(u_y) \tag{15.8}$$

Sampling this continuous autocovariance at integral grid points gives

$$\gamma(n_x, n_y) = \frac{\sigma^2}{1-r^2} r^{|n_x|}\delta_{0,n_y} \tag{15.9}$$

The corresponding discrete spectral density is

$$\begin{aligned}X(\omega_x, \omega_y) &= \frac{\sigma^2}{1 + r^2 - 2r\cos\omega_x} \\ &= \frac{\sigma^2}{1-r^2}\sum_{n_x} r^{|n_x|}e^{-jn_x\omega_x}\end{aligned} \tag{15.10}$$

and the continuous reconstruction of the autocovariance would be

$$\begin{aligned}\gamma_s(u_x, u_y) &= \frac{1}{4\pi^2}\int_{-\pi}^{+\pi}\int_{-\pi}^{+\pi}\frac{\sigma^2}{1-r^2}e^{j(\omega_x u_x + \omega_y u_y)}\sum_{n_x} r^{|n_x|}e^{-jn_x\omega_x}\text{d}\omega_x\text{d}\omega_y \\ &= \sum_{n_x}\frac{\sigma^2}{1-r^2}r^{|n_x|}\text{sinc}\pi u_y\text{sinc}\pi(u_x - n_x)\end{aligned} \tag{15.11}$$

which can look quite different from the original autocovariance, as Figure 42 shows.

What would we predict for the autoregressive coefficients in a case like this? Well, we had

$$\gamma(\vec{u}) = \frac{\sigma^2}{4\pi^2}\int_{A1}\frac{\exp(+j\vec{u}^T\vec{\omega})}{\sum\beta(\vec{k})\exp(j\vec{k}^T\vec{\omega})}\text{d}^2\vec{\omega} \tag{15.12}$$

where the summation is over all lags $\vec{k}$ for which $\beta(\vec{k})$ is non-zero. Applying a rotation R and a scaling h results in

$$\begin{aligned}\gamma(Rh\vec{u}) &= \frac{\sigma^2}{4\pi^2}\int_{A1}\frac{\exp(+jRh\vec{u}^T\vec{\omega})}{\sum\beta(\vec{k})\exp(j\vec{k}^T\vec{\omega})}\text{d}^2\vec{\omega} \\ &= \frac{\sigma^2}{4\pi^2h^2}\int_{A2}\frac{\exp(+j\vec{u}^T\bar{\omega})}{\sum\beta(\vec{k})\exp[j(\frac{R^T}{h}\vec{k})^T\bar{\omega}]}\text{d}^2\bar{\omega}\end{aligned} \tag{15.13}$$

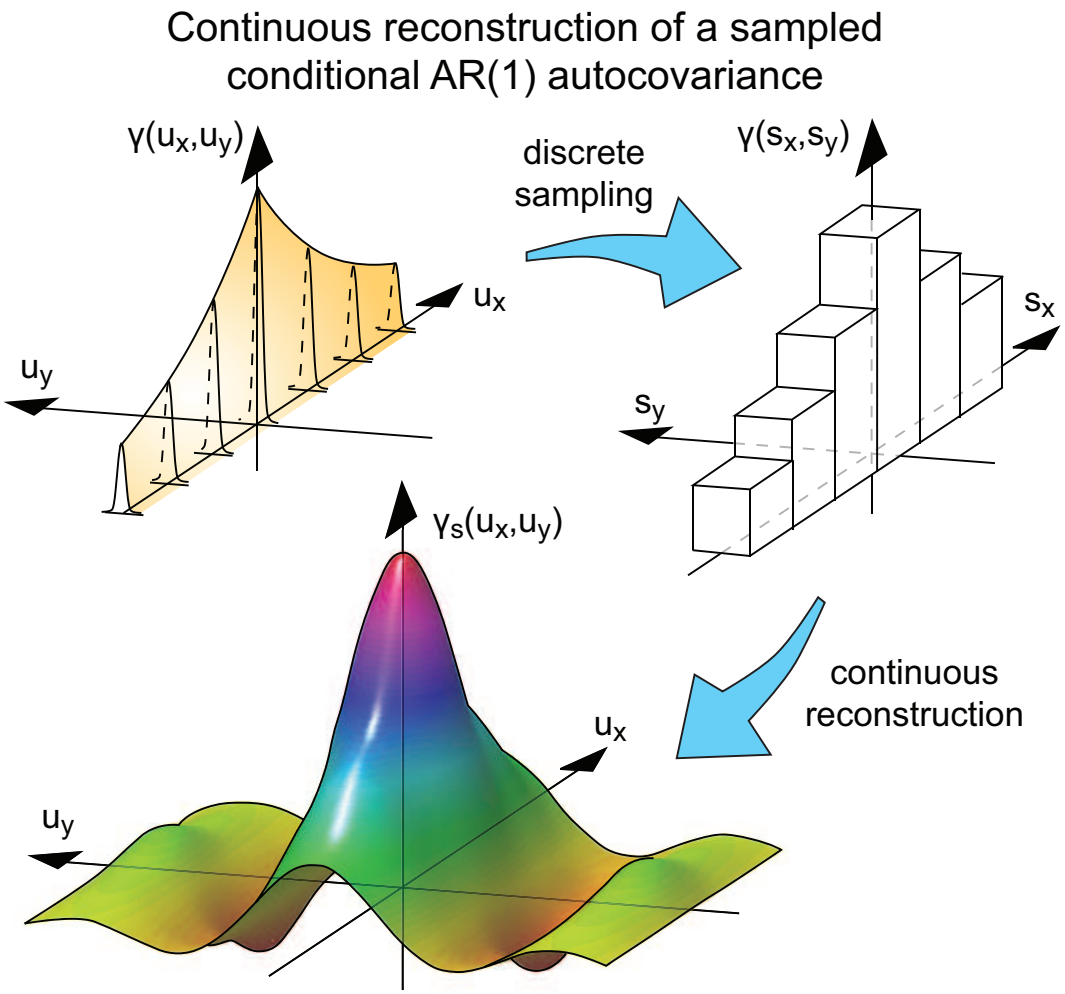

Figure 42: The reconstruction of an autocovariance function from its samples can look quite different from the original.

under the change of variables $\bar{\omega} = hR^T\vec{\omega}$. To put this in standard autoregressive form, we need

$$\bar{\gamma}(\vec{u}) = \frac{\bar{\sigma}^2}{4\pi^2}\int_{A1}\frac{\exp(+j\vec{u}^T\bar{\omega})}{\sum\bar{\beta}(\vec{k})\exp(j\vec{k}^T\bar{\omega})}\mathrm{d}^2\bar{\omega} \tag{15.14}$$

and so, at a minimum, we need to solve

$$\bar{\beta}(\vec{k})\exp(j\vec{k}^T\bar{\omega}) = \beta(\vec{k})\exp[j(\frac{R^T}{h}\vec{k})^T\bar{\omega}] \tag{15.15}$$

Also, in order to proceed further along this line of reasoning, we must make the additional assumption that the spectral density is band-limited, so that we can assert that integrations over regions A1 and A2 are the same. In this case, we can expand $\exp[j(R^T/h)\vec{k}^T\vec{\omega}]$ in terms of $\exp(j\vec{k}^T\vec{\omega})$. Setting

$$\vec{L} = \frac{R^T}{h}\vec{k} = \frac{1}{h}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}k\\ l\end{bmatrix}, \qquad \vec{k} = \begin{bmatrix}m\\ n\end{bmatrix} \tag{15.16}$$

the projection of $\exp(j\vec{L}^T\vec{\omega})$ onto $\exp(j\vec{k}^T\vec{\omega})$ is

$$\begin{aligned}\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} & \exp j(\vec{L}-\vec{k})^T\vec{\omega}\mathrm{d}^2\vec{\omega}\\ &= \mathrm{sinc}_2\pi(\vec{L}-\vec{k})\\ &= \mathrm{sinc}\pi\left[\frac{k\cos\theta - l\sin\theta}{h} - m\right]\mathrm{sinc}\pi\left[\frac{k\sin\theta + l\cos\theta}{h} - n\right]\end{aligned} \tag{15.17}$$

and hence the total coefficient of $\exp(j\vec{k}^T\vec{\omega})$ would be

$$\bar{\beta}(m,n) = \sum_{k,l}\beta(k,l)\mathrm{sinc}\pi\left[\frac{k\cos\theta - l\sin\theta}{h} - m\right]\mathrm{sinc}\pi\left[\frac{k\sin\theta + l\cos\theta}{h} - n\right] \tag{15.18}$$

This is similar to Equation 3.15 in Cohen, Fan, and Patel [25]. It may look seductive, but it isn't right! The operations of rotation and discrete Fourier transform do *not* commute. Returning to our previous example, suppose we have a conditional AR(1) field along the x-direction, so that

$$\gamma(\vec{u}) = \frac{\sigma^2}{1-r^2}r^{|u_x|}\delta(u_y) \tag{15.19}$$

The corresponding discrete spectral density is

$$|X(\omega)|^2 = \frac{\sigma^2}{1+r^2-2r\cos\omega_x} \tag{15.20}$$

and, using the notation of equation (15.4) above, we can say that $\beta(0,0) = 1+r^2$ and $\beta(1,0) = \beta(-1,0) = -r$. Now subject this system to a transform with $\theta = -\pi/4$ and $h = \sqrt{2}$. If we sample first and then rotate, then Equation 15.18 above would yield

$$\begin{aligned}\bar{\beta}(0,0) &= 1 + r^2 - r[\mathrm{sinc}^2(\frac{\pi}{2}) + \mathrm{sinc}^2(-\frac{\pi}{2})]\\ \bar{\beta}(m,n) &= -r[\mathrm{sinc}\pi(-\frac{1}{2}-m)\mathrm{sinc}\pi(-\frac{1}{2}-n) + \mathrm{sinc}\pi(\frac{1}{2}-m)\mathrm{sinc}\pi(\frac{1}{2}-n)]\end{aligned} \tag{15.21}$$

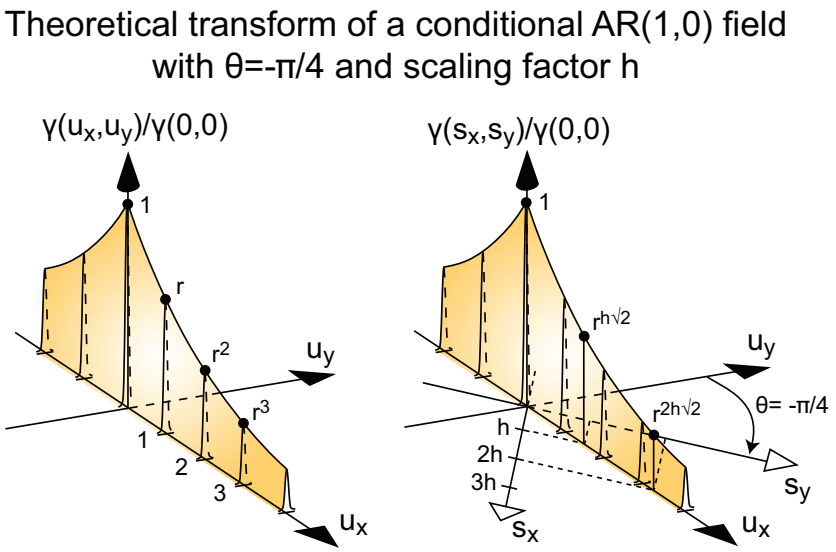

Figure 43: If we subject a conditional AR(1,0) series to a rotation $\theta = -\pi/4$ and scale factor h, we should get a pretty simple AR series with only $\bar{\beta}(1,1)$ showing up.

This predicts a non-zero value for *all* $\bar{\beta}(m,n)$. Now suppose instead that we rotate and scale the autocovariance first, and then sample it, as illustrated in Figure 43. After sampling, we find that the system is still a conditional AR(1) autoregressive field, but with spectral density

$$|X(\omega)|^2 = \frac{[\text{const}]}{1 + r^{2h\sqrt{2}} - 2r^{h\sqrt{2}}\cos(\omega_x + \omega_y)} \tag{15.22}$$

that is, the only non-zero coefficients are $\bar{\beta}(0,0)$, $\bar{\beta}(-1,-1)$ and $\bar{\beta}(1,1)$. Figure 44 charts some actual results, which clearly demonstrate that we must transform the random field before sampling it.

The bottom line, then, is that in order to calculate the discrete spectral density under coordinate transformation, we need to:

(1) estimate (or calculate) the continuous autocovariance

(2) rotate and scale the continuous autocovariance

(3) sample the transformed autocovariance, averaging if necessary
according to the mode of physical sampling

(4) calculate the resulting spectral density and autoregressive coefficients

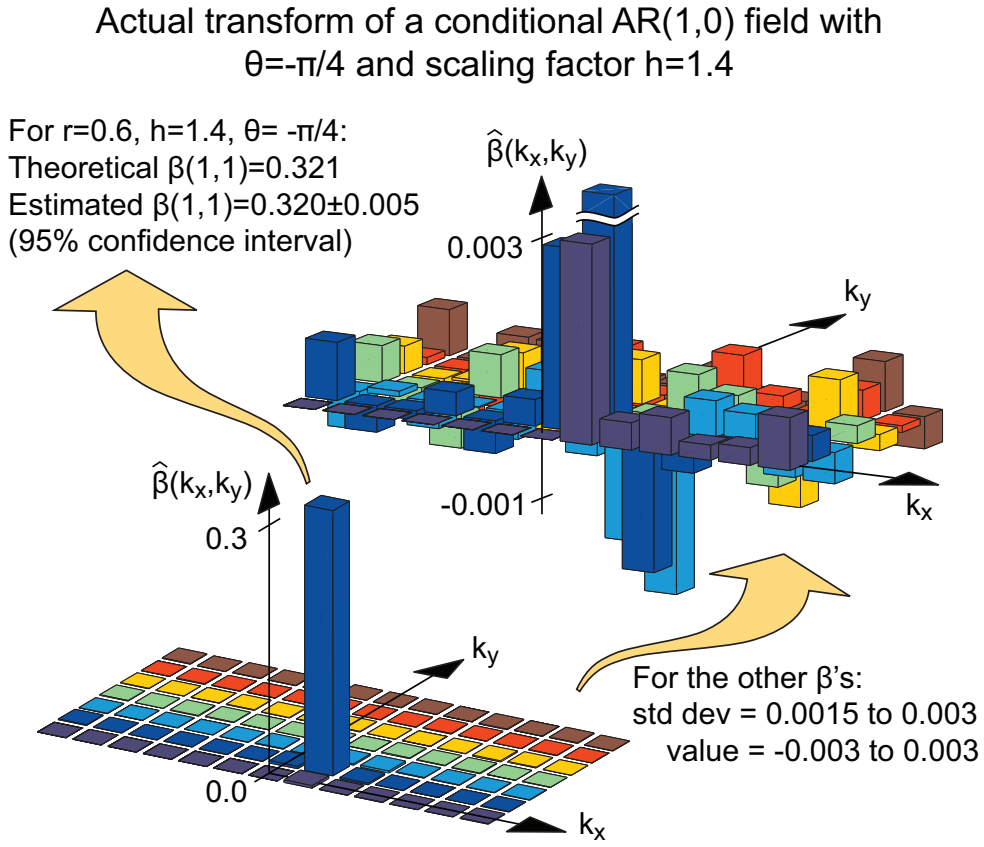

Figure 44: Here I subject a synthesized conditional AR(1,0) series with $r = 0.6$ to a transform with $\theta = -\pi/4$ and $h = 1.4$. The resulting spectral density and estimated autoregressive parameters are consistent with the idea that we must apply the transform to the continuous autocovariance first, and then sample the result. The original field is of size 1200 x 1200, the sampled field is of size 558 x 558, the Hamming window is of size 31 x 31, and the chart shows the average autoregressive parameters obtained over 16 trials.

16 Skew-separable spectral density

In Appendix D, I demonstrate that the two-dimensional Fourier transform retains its form even under a shear transformation of the coordinate axes. What this means, in practice, is that if an autocovariance term is separable along two directions, then the corresponding spectral density term is also separable along two directions, perpendicular to the original pair.

Why might this be of interest? Well, in the last section I talked about estimating the continuous autocovariance of a two-dimensional random field, and it would be nice to have some basis functions to use for that purpose. The aim of the current section is to suggest what kind of basis functions would make sense for 2D autoregressive fields.

We can always estimate a two-dimensional conditional autoregressive random field's spectral density as

$$\left|\frac{X(\vec{\omega})}{E(\vec{\omega})}\right|^2 = \frac{[\text{const}]}{\prod[1 + {r_j}^2 - 2r_j \cos {\vec{k}_j}^T \vec{\omega}]} \tag{16.1}$$

by fitting a cosine series to the logarithm of the spectral density, as outlined in Section 11. We might attempt a partial fraction expansion of the form

$$\left|\frac{X(\vec{\omega})}{E(\vec{\omega})}\right|^2 = \sum_j \frac{D_j}{1 + {r_j}^2 - 2r_j \cos {\vec{k}_j}^T \vec{\omega}} \tag{16.2}$$

but that implies a continuous autocovariance

$$\frac{1}{\sigma^2}\gamma(\vec{u}) = \sum_j \frac{D_j}{1 - {r_j}^2} {r_j}^{|\hat{k}_j{}^T \vec{u}|/|\vec{k}_j|} \delta(\hat{k}_j{}^T \hat{u} - 1) \tag{16.3}$$

where σ^2 is the innovation variance, $\hat{k}_j = \vec{k}_j/|\vec{k}_j|$, and $\hat{u} = \vec{u}/|\vec{u}|$. Each individual term in equation (16.3) is what I call a "line autocovariance", because it looks like a thin sheet when you plot it out (see Figure 45).

So an expansion such as Equation 16.2 above may look reasonable, but actually represents a non-physical autocovariance. What to do? Well, the next simplest representation would be a summation of the form

$$\left|\frac{X(\vec{\omega})}{E(\vec{\omega})}\right|^2 = \sum_j \frac{D_j}{1 + {r_{j1}}^2 - 2r_{j1} \cos {\vec{k}_{j1}}^T \vec{\omega}} \quad \frac{1}{1 + {r_{j2}}^2 - 2r_{j2} \cos {\vec{k}_{j2}}^T \vec{\omega}} \tag{16.4}$$

in which each term involves *two* lags, namely $\vec{k}_{j1}$ and $\vec{k}_{j2}$. I call such terms "skew-separable". Even if these two lags are not perpendicular to each other, we know from Appendix D that the corresponding autocovariance term will also be separable. If need be, we can get the various possible $\vec{k}_j$'s from a cosine transform of the representation

$$\left|\frac{X(\vec{\omega})}{E(\vec{\omega})}\right|^2 = \frac{[\text{const}]}{\prod[1 + {r_j}^2 - 2r_j \cos {\vec{k}_j}^T \vec{\omega}]} \tag{16.5}$$

as noted above. The individual discrete spectral density terms

$$\frac{1}{1 + {r_1}^2 - 2r_1 \cos {\vec{k}_1}^T \vec{\omega}} \quad \frac{1}{1 + {r_2}^2 - 2r_2 \cos {\vec{k}_2}^T \vec{\omega}} \tag{16.6}$$

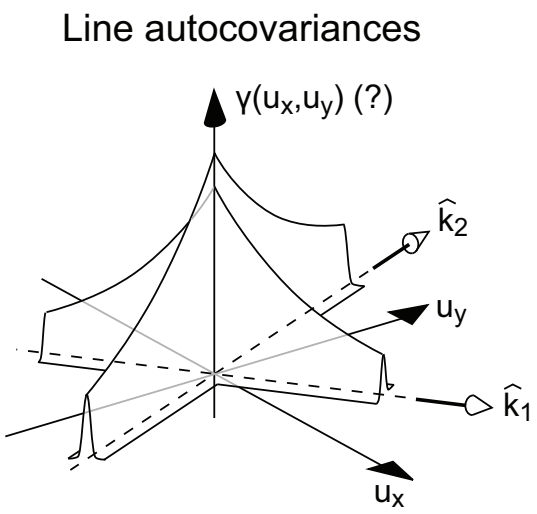

Figure 45: A spectral density such as that in Equation 16.2 would lead to a non-physical autocovariance.

have corresponding continuous autocovariance terms

$$r_1{}^{|\hat{q}_1{}^T \vec{u}|/|\vec{q}_1|} r_2{}^{|\hat{q}_2{}^T \vec{u}|/|\vec{q}_2|} \tag{16.7}$$

where $\vec{k}_2{}^T \vec{q}_1 = 0$, $\vec{k}_1{}^T \vec{q}_2 = 0$, $\hat{q}_1 = \vec{q}_1/|\vec{q}_1|$, and $\hat{q}_2 = \vec{q}_2/|\vec{q}_2|$. These autocovariance terms now have a two-dimensional shape and could form possible basis functions for a real autocovariance. In fact, we can group them together by allowing one or both of the factors of

$$\frac{1}{1 + r_1{}^2 - 2r_1 \cos \vec{k}_1{}^T \vec{\omega}} \quad \frac{1}{1 + r_2{}^2 - 2r_2 \cos \vec{k}_2{}^T \vec{\omega}} \tag{16.8}$$

to assume complex conjugate AR(2) form, in which case the corresponding autocovariance term would look like

$$\begin{aligned} & r_1{}^{|u_1|/q_1} r_2{}^{|u_2|/q_2} \cos(\theta_2 u_2 - \Psi_2) \\ \text{or} \quad & r_1{}^{|u_1|/q_1} \cos(\theta_1 u_1 - \Psi_1) r_2{}^{|u_2|/q_2} \cos(\theta_2 u_2 - \Psi_2) \end{aligned} \tag{16.9}$$

where $u_1 = \hat{q}_1{}^T \vec{u}$, $u_2 = \hat{q}_2{}^T \vec{u}$, $q_1 = |\vec{q}_1|$, $q_2 = |\vec{q}_2|$, and the remaining variables have the meanings assigned to them in Section 23, where I consider continuous AR(2) propagators. In particular, Ψ_1 is a function of (r_1, θ_1) and Ψ_2 is a function of (r_2, θ_2). So our AR(2) basis function

$$r_1{}^{|u_1|/q_1} \cos(\theta_1 u_1 - \Psi_1) r_2{}^{|u_2|/q_2} \cos(\theta_2 u_2 - \Psi_2) \tag{16.10}$$

is really a six-parameter function, those parameters being r_1, θ_1, r_2, θ_2, and the azimuthal directions of $\hat{q}_1$ and $\hat{q}_2$. These functions are quite versatile. Figure 46 shows some typical functional shapes, and Figure 47 shows a fit of four such terms onto the average covariance of the Brodatz D38 water texture, with no rotation. Even though there are only four terms, and I performed the fit by hand, those four terms capture the principal features of the discrete autocovariance.

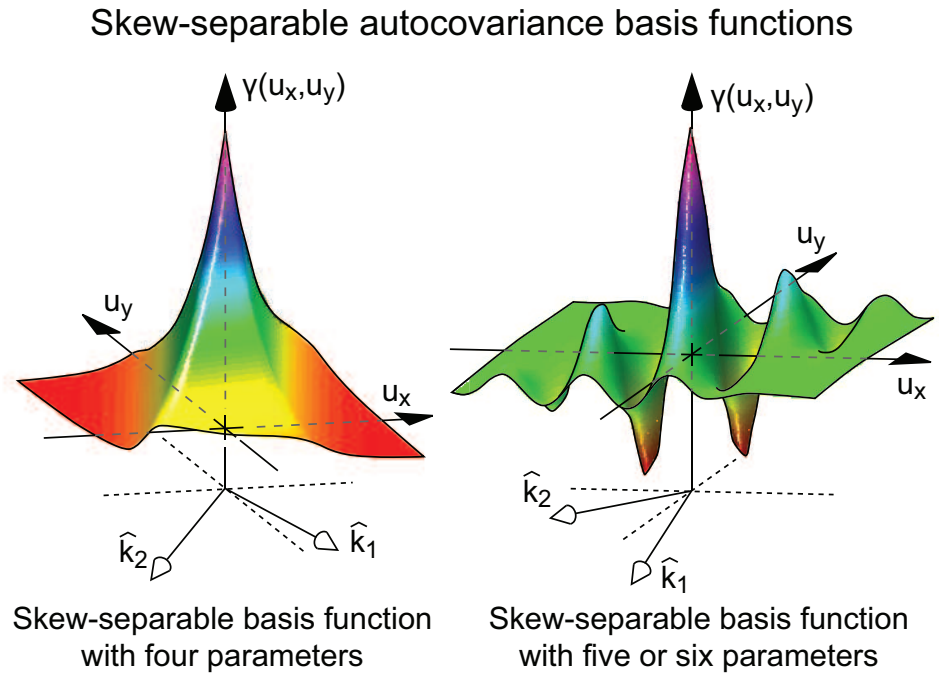

Figure 46: Some typical skew-separable basis function shapes.

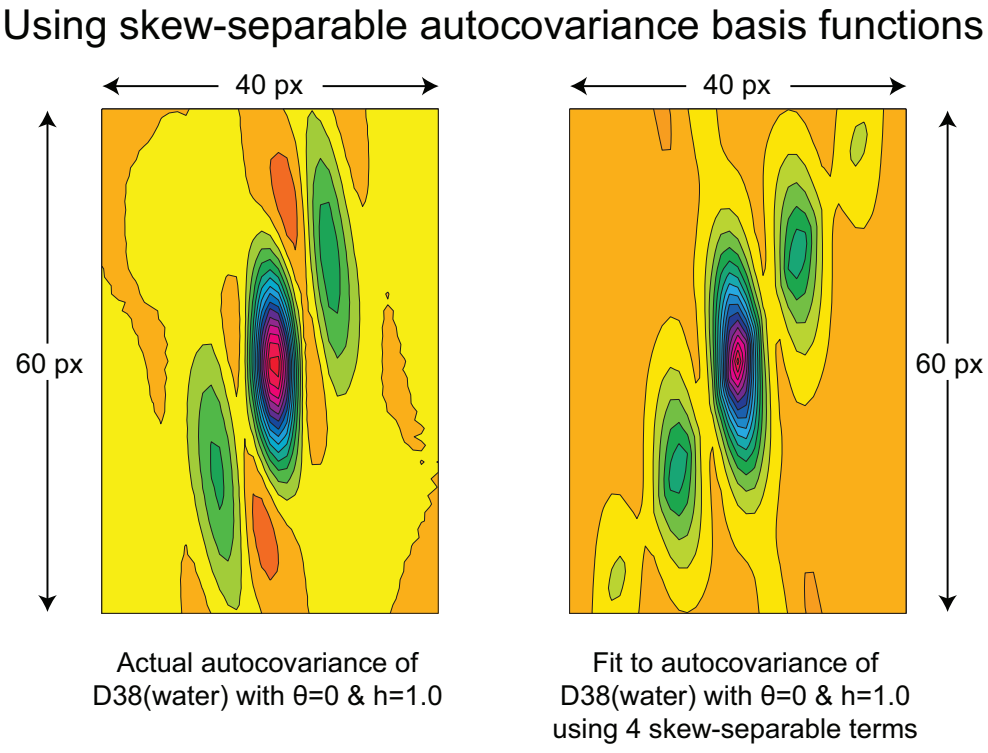

Figure 47: A manual fit of four skew-separable basis functions to the estimated autocovariance of the Brodatz water texture at zero degrees rotation.

17 Rotational invariants

In the last section, I suggested that a logarithmic expansion of a 2D discrete spectral density such as

$$\left|\frac{X(\vec{\omega})}{E(\vec{\omega})}\right|^2 = \frac{[\text{const}]}{\prod_j [1 + {r_j}^2 - 2r_j \cos(\vec{\omega}^T \vec{k}_j)]} \tag{17.1}$$

could lead to a representation of the corresponding continuous autocovariance as

$$\gamma(\vec{u}) = \sum_j {r_{j1}}^{|\vec{u}^T \hat{k}_{j1}|} {r_{j2}}^{|\vec{u}^T \hat{k}_{j2}|} \quad (\hat{k}_{j1} \neq \hat{k}_{j2}) \tag{17.2}$$

and that this representation can be a practical one for visual texture. Are there any rotational invariants in there? Well, the r_j's are clearly invariant under pure rotation, but if we have say n distinct directions $\hat{k}_1 \ldots \hat{k}_n$, then one is arbitrary and the remaining $n-1$ directions may be measured with respect to the arbitrary one.

In fact we can say a bit more. Under the transform $\bar{u} \rightarrow Rh\vec{u}$ where R is a rigid rotation and h is a scale factor, we have

$$\gamma(\vec{u}) = \sum_j [{r_{j1}}^{(1/h)}]^{|\bar{u}^T R \hat{k}_{j1}|} [{r_{j2}}^{(1/h)}]^{|\bar{u}^T R \hat{k}_{j2}|} \quad (\hat{k}_{j1} \neq \hat{k}_{j2}) \tag{17.3}$$

which shows that the quantities $(\log r_j)/h$ and the angular differences between the $\hat{k}_j$'s are preserved.

We might consider looking at the Taylor series for the continuous autocovariance for possible rotational invariants, but that idea has little merit since the second-order curvatures are undefined at the origin. A more promising avenue of attack is to look at the inverse of the spectral density. There is some theoretical justification for this. For a "pseudo-Markovian" continuous random field, for example, the inverse spectral density is a finite polynomial in $\vec{\omega}$ [46] :

$$\left|\frac{X(\vec{\omega})}{E(\vec{\omega})}\right|^{-2} = \sum_{\vec{k}} a_{k_1 \ldots k_N} {\omega_1}^{k_1} \ldots {\omega_N}^{k_N} \tag{17.4}$$

where $k_1 + \cdots + k_N \leq 2p$. Here the field is N-dimensional, $\vec{\omega}$ is an N-dimensional continuous spatial frequency, and $2p$ is the order of the expansion. Might that work for us? Well, as noted above, by exploiting the basic relation

$$\log \frac{1}{1 + r^2 - 2r\cos\theta} = \sum_{n=1}^{\infty} \frac{2}{n} r^n \cos n\theta \tag{17.5}$$

we can represent any conditional autoregressive spectral density as

$$\left|\frac{X(\vec{\omega})}{E(\vec{\omega})}\right|^2 = \frac{[\text{const}]}{\prod_j [1 + {r_j}^2 - 2r_j \cos(\vec{\omega}^T \vec{k}_j)]} \tag{17.6}$$

If the scale of our observations is *already* small with respect to the r_j's, then we can derive an approximate version as follows:

$$\begin{aligned}\left|\frac{X(\vec{\omega})}{E(\vec{\omega})}\right|^2 &\simeq [\text{const}]\prod_j\left[\frac{(-2\log r_j)}{(-\log r_j)^2+(\vec{\omega}^T\vec{k}_j)^2}\right]\\ \left|\frac{X(\vec{\omega})}{E(\vec{\omega})}\right|^{-2} &\simeq [\text{const}]\prod_j\left[(-\log r_j)^2+(\vec{\omega}^T\vec{k}_j)^2\right]\\ &\simeq [\text{const}]\prod_j\left[1+\frac{1}{(-\log r_j)^2}(\vec{\omega}^T\vec{k}_j)^2\right]\end{aligned} \tag{17.7}$$

That is clearly a polynomial in the discrete spatial frequency $\vec{\omega}$. We know that the spectral density is invariant under coordinate inversion, so we can expand it as

$$\left|\frac{X(\vec{\omega})}{E(\vec{\omega})}\right|^{-2} = [\text{const}][1+\frac{1}{2!}\vec{\omega}^T A_2\vec{\omega}+\frac{1}{4!}\vec{\omega}\vec{\omega}^T A_4\vec{\omega}\vec{\omega}^T+\ldots] \tag{17.8}$$

where $A_2, A_4, \ldots$ are matrices related to the partial derivatives of $|X(\vec{\omega})/E(\vec{\omega})|^{-2}$ at the origin. Suppose, for example, that f is the normalized inverse spectral density:

$$f(\vec{\omega}) = \frac{|X(\vec{\omega})/E(\vec{\omega})|^{-2}}{|X(\vec{0})/E(\vec{0})|^{-2}} = 1+\frac{1}{2!}\vec{\omega}^T A_2\vec{\omega}+\frac{1}{4!}\vec{\omega}\vec{\omega}^T A_4\vec{\omega}\vec{\omega}^T+\ldots \tag{17.9}$$

Then A_2 is given by

$$A_2 = \begin{bmatrix}\partial^2/\partial x^2 & \partial^2/\partial x\partial y\\ \partial^2/\partial x\partial y & \partial^2/\partial y^2\end{bmatrix} f = \begin{bmatrix}\nabla_{xx} & \nabla_{xy}\\ \nabla_{xy} & \nabla_{yy}\end{bmatrix} f \tag{17.10}$$

The transforms of partial derivatives under rotation are pretty straightforward. Here are the second order ones:

$$\begin{gathered}\frac{1}{2}(\nabla_{xx}+\nabla_{yy}) = \frac{1}{2}(\bar{\nabla}_{xx}+\bar{\nabla}_{yy})\\ \begin{bmatrix}\frac{1}{2}(\nabla_{xx}-\nabla_{yy})\\ \nabla_{xy}\end{bmatrix} = \begin{bmatrix}\cos 2\theta & -\sin 2\theta\\ \sin 2\theta & \cos 2\theta\end{bmatrix}\begin{bmatrix}\frac{1}{2}(\bar{\nabla}_{xx}-\bar{\nabla}_{yy})\\ \bar{\nabla}_{xy}\end{bmatrix}\end{gathered} \tag{17.11}$$

In this last set of equations, the coordinate system rotates through a counterclockwise angle θ, and the barred derivatives are measured after rotation. The fourth order transforms are:

$$\begin{gathered}\frac{1}{2}(\nabla_{xxxx}+2\nabla_{xxyy}+\nabla_{yyyy}) = \frac{1}{2}(\bar{\nabla}_{xxxx}+2\bar{\nabla}_{xxyy}+\bar{\nabla}_{yyyy})\\ \begin{bmatrix}\frac{1}{2}(\nabla_{xxxx}-\nabla_{yyyy})\\ \nabla_{xxxy}+\nabla_{xyyy}\end{bmatrix} = \begin{bmatrix}\cos 2\theta & -\sin 2\theta\\ \sin 2\theta & \cos 2\theta\end{bmatrix}\begin{bmatrix}\frac{1}{2}(\bar{\nabla}_{xxxx}-\bar{\nabla}_{yyyy})\\ \bar{\nabla}_{xxxy}+\bar{\nabla}_{xyyy}\end{bmatrix}\\ \begin{bmatrix}\nabla_{xxxy}-\nabla_{xyyy}\\ \frac{1}{4}(\nabla_{xxxx}-6\nabla_{xxyy}+\nabla_{yyyy})\end{bmatrix} = \begin{bmatrix}\cos 4\theta & \sin 4\theta\\ -\sin 4\theta & \cos 4\theta\end{bmatrix}\begin{bmatrix}\bar{\nabla}_{xxxy}-\bar{\nabla}_{xyyy}\\ \frac{1}{4}(\bar{\nabla}_{xxxx}-6\bar{\nabla}_{xxyy}+\bar{\nabla}_{yyyy})\end{bmatrix}\end{gathered} \tag{17.12}$$

Thus, for example, in the expansion

$$\begin{aligned}\left|\frac{X(\vec{\omega})}{E(\vec{\omega})}\right|^{-2} &\simeq [\text{const}]\prod_j\left[1+\frac{1}{(-\log r_j)^2}(\vec{\omega}^T\vec{k}_j)^2\right]\\ &\simeq [\text{const}]\left[1+\sum_j\frac{1}{(-\log r_j)^2}\vec{\omega}^T\vec{k}\vec{k}^T\vec{\omega}+\ldots\right]\end{aligned} \tag{17.13}$$

we can identify

$$\mathrm{tr}A_2 = \nabla^2 f = \sum_j \frac{|\vec{k}_j|^2}{(-\log r_j)^2} \tag{17.14}$$

We already know from equation (12.5) that the r_j's transform exponentially under scaling, so we would expect the quantity

$$h^2 \mathrm{tr}A_2 = h^2 \nabla^2 f = \sum_j \left[\frac{h}{-\log r_j}\right]^2 |\vec{k}_j|^2 \tag{17.15}$$

to be an invariant under rotation and scaling, where h is the scale factor. It is proportional to the Laplacian of the inverse of the continuous spectral density.

Figure 48 shows a test of that idea, applied to a synthetic conditional AR(1,1) series. What the figure tells us is that calculating rotational invariants in this way is a fool's errand. As far as samples of visual texture go, 480 x 480 is a pretty large size. But even at this sample size, our ability to calculate even the simplest rotational invariant is poor. At 35 degrees, for example, the smallest significant contributor to $\sum |\vec{k}_j|^2/(-\log r_j)^2$ is at lag (7,5), and that term contributes more than one-quarter of the total. So calculating rotational invariants to any reasonable precision at smaller sample sizes is just not on.

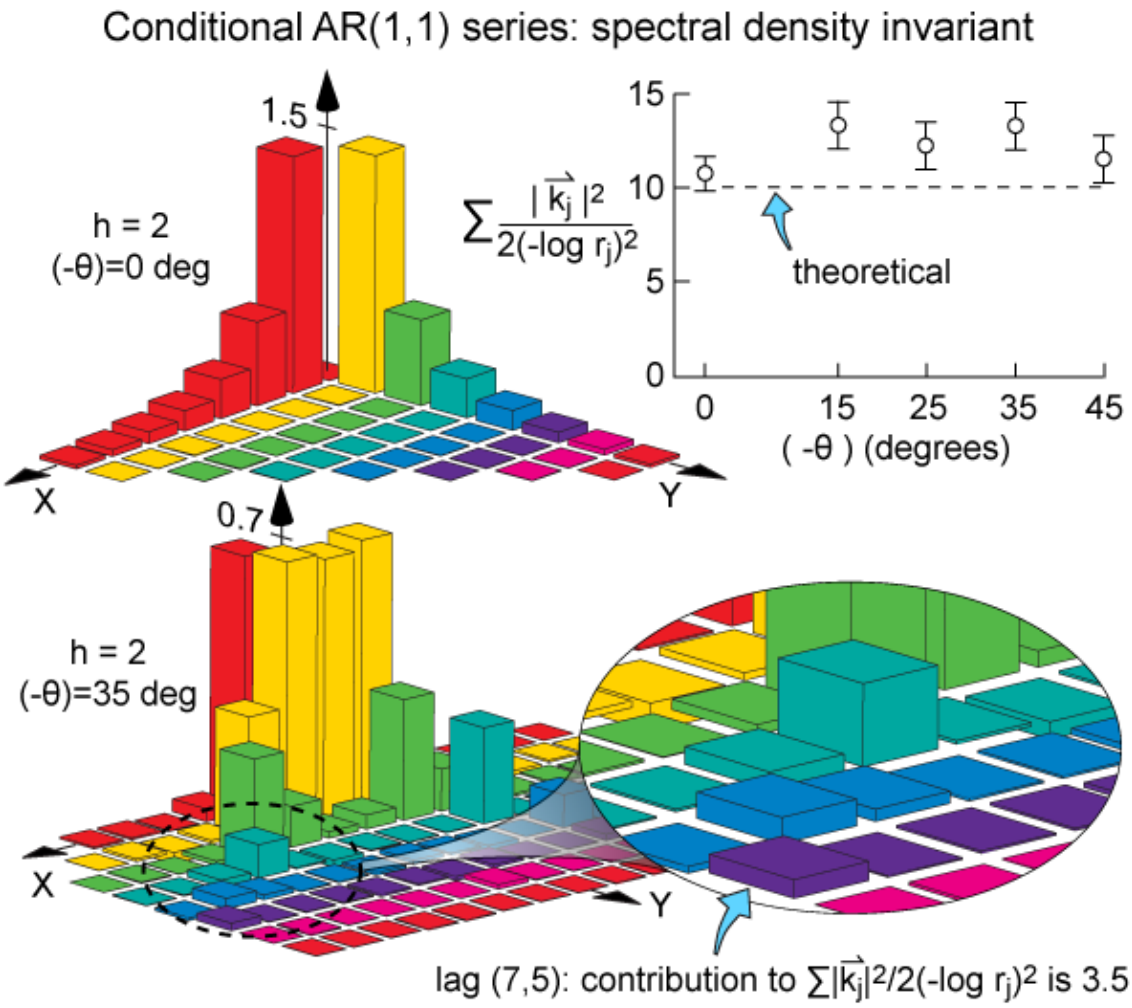

Figure 48: Calculation of a rotational invariant, in this case the Laplacian of the inverse of the continuous spectral density. The base series is a synthesized conditional AR(1,1) random field of size 1400 x 1400 with $r_x = r_y = 0.854$. The sampled series are of size 480 x 480 with $h = 2$. The bar charts show the discrete cosine transform of the logarithm of the estimated spectral density.

18 Interpolation into a conditional AR(1) field

As we've seen before, standard methods of interpolation can cause noticeable distortion when applied to correlated random fields. In particular, they can reduce the series variance and introduce spurious correlations. So, a natural question is: are there better ways to interpolate into a random field?

The first thing we must do is distinguish between interpolation and forecasting. The point of forecasting is to *minimize* the target point's variance, while the point of interpolation is to *preserve* the target point's variance (i.e. make it "fit in" with the rest of the series). For example, suppose that we have a conditional AR(2) series, defined by

$$x(t) = a_1 x(t-1) + a_2 x(t-2) + e(t), \quad x(t) = y(t) - \mu, \quad e(t) \simeq N(0, \sigma^2) \tag{18.1}$$

and a sample $\{x(1) \ldots x(n)\}$. The best *forecast* of $x(n+1)$ would be $a_1 x(n) + a_2 x(n-1)$, because that is the expected value of $x(n+1)$, conditioned on the sample. However, the best *extrapolation* of the series would be $a_1 x(n) + a_2 x(n-1) + z\sigma$ where z is a standard $N(0,1)$ variate.

Let's start by being more precise about what we want to accomplish. I imagine a discrete sequence (or grid) of source points, whose autoregressive character I either know, or have estimated. I wish to calculate a field value at a target point, that is in between some source points. A target point's value will consist of two parts: a linear combination of nearby source points, and a random part. The linear combination will be related to the autoregressive coefficients of the source sequence. The variance of the random part will be zero when the target point coincides with a source point, and will be greatest when the target point is equidistant from neighboring source points.

Now, let's put those ideas into practice. For a conditional AR(1) series, defined by

$$x(t) = rx(t-1) + e(t), \qquad x(t) = y(t) - \mu, \qquad e(t) \simeq N(0, \sigma^2) \tag{18.2}$$

the transfer function and autocovariance are

$$\begin{aligned} \left| \frac{X(\omega)}{E(\omega)} \right|^2 &= \frac{1}{1 + r^2 - 2r\cos\omega} \\ \frac{1}{\sigma^2}\gamma(s) &= \frac{r^{|s|}}{1 - r^2} \qquad [s \text{ integral}] \end{aligned} \tag{18.3}$$

If we regard the series $\{x(t)\}$ as being a sampling from a continuous field, then the continuous autocovariance will be

$$\frac{1}{\sigma^2}\gamma(u) = \frac{r^{|u|}}{1 - r^2} \qquad [u \text{ real}] \tag{18.4}$$

and the corresponding continuous spectral density is

$$|G(\omega)|^2 = \int_{-\infty}^{+\infty} \gamma(u) e^{-j\omega u} \mathrm{d}u = \frac{2\sigma^2}{1 - r^2} \frac{p}{p^2 + \omega^2} \tag{18.5}$$

where $p = -\log r$. In order to create a continuous moving average equivalent, or "propagator", we need to choose a phase function for $G(\omega)$. The only constraint is that the phase be antisymmetric, so that the propagator (and hence the continuous random field) will be

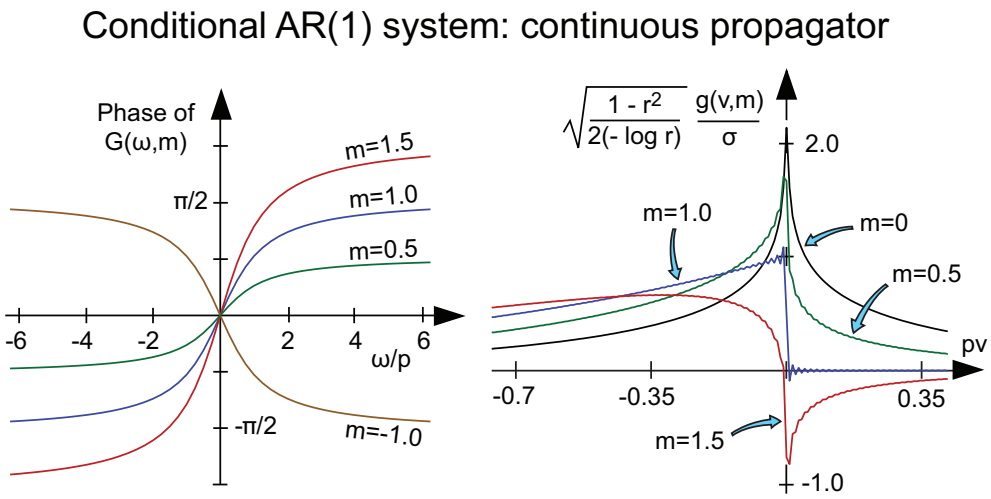

Figure 49: Knowledge of the spectral density function does not completely constrain the phase of the corresponding propagator function.

real-valued. This situation is the continuous analogue of the situation that I noted back in Section 5: if a mean-reduced sample of an autoregressive series and the corresponding innovations are in a linear relation $L\vec{x} = \vec{e}$, and the probability density of $\vec{e}$ depends on $\vec{e}^T\vec{e}$ only, then *any* unitary transform applied to L yields an equivalent model. If we choose

$$\begin{aligned} G(\omega) &= \sigma\sqrt{\frac{2p}{1-r^2}}\frac{1}{p+j\omega} \\ &= \sigma\sqrt{\frac{2p}{1-r^2}}\frac{1}{\sqrt{p^2+\omega^2}}\exp\left(-j\tan^{-1}\frac{\omega}{p}\right) \end{aligned} \tag{18.6}$$

then we get a causal model, in which

$$x(t) = \sigma\sqrt{\frac{2(-\log r)}{1-r^2}}\int_{-\infty}^{t} r^{t-t'}n(t')\mathrm{d}t' \tag{18.7}$$

and $n(t')$ is continuous unit Gaussian noise (see Section 13). But that's not the only possibility! Suppose that we choose instead

$$\begin{aligned} G(\omega,m) &= \sigma\sqrt{\frac{2p}{1-r^2}}\frac{1}{\sqrt{p^2+\omega^2}}\exp\left(-jm\tan^{-1}\frac{\omega}{p}\right) \\ g(v,m) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} G(\omega,m)e^{+j\omega v}\mathrm{d}\omega \end{aligned} \tag{18.8}$$

where m is now a parameter. I calculated $g(v,m)$ numerically, and Figure 49 shows what the resulting propagator looks like.

The propagator for $m = 1$ is the usual causal one. The propagator for $m = 0$ is the simultaneous version,

$$\frac{1}{\sigma}\sqrt{\frac{1-r^2}{2p}}g(v,0) = \frac{1}{\pi}K_0(pv) \tag{18.9}$$

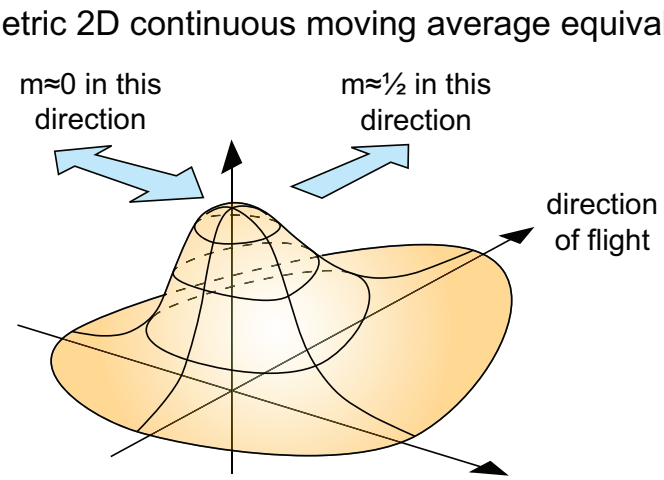

Figure 50: Birds that fly in flocks probably have a moving average function that is skewed toward the forward direction.

where $K_0(z)$ is a modified Bessel function of the second kind [47]. The other cases illustrated are "partially causal". A good question to ask here is: are there are any physical realizations of a partially causal propagator? Well here's one: birds in a flock take their acceleration cues from their spatial neighbors. Any one specific bird will have its own random component of acceleration, but it will be most heavily influenced by the birds in front of it, and will be the least influenced by birds behind it (note that most flocking birds have a wide visual field). So their moving average function will be something like the one illustrated in Figure 50.

For now, though, we will restrict ourselves to a causal continuous propagator:

$$g(t,t') = \begin{cases} \sigma\sqrt{2p/(1-r^2)}\, r^{t-t'} & t' < t, p = -\log r \\ 0 & t' > t \end{cases} \tag{18.10}$$

with corresponding moving average representation

$$x(t) = \sigma\sqrt{\frac{2p}{1-r^2}} \int_{-\infty}^{t} r^{t-t'} n(t')\mathrm{d}t', \quad n(t') = \text{unit Gaussian noise} \tag{18.11}$$

In a case like this, where $x(t) = \int f(t-t')n(t')\mathrm{d}t'$, I say that $f(t-t')$ is the "envelope" of $x(t)$. Figure 51 illustrates the envelopes of $x(t)$, $x(t+\alpha)$, and $x(t+1)$.

We'll attempt an interpolant of the form

$$x(t+\alpha) = g_0 x(t) + g_1 x(t+1) + e(t,\alpha) \tag{18.12}$$

in which case the envelope of $e(t,\alpha)$ is that of the shaded area in figure 51. The variance of $e(t,\alpha)$ works out to

$$\begin{aligned} \frac{1-r^2}{\sigma^2}\mathrm{var}[e(t,\alpha)] &= (g_0 + rg_1 - r^\alpha)^2 + (r^\alpha - g_1 r)^2(r^{-2\alpha} - 1) \\ &\quad + (g_1 r)^2(r^{-2} - r^{-2\alpha}) \end{aligned} \tag{18.13}$$

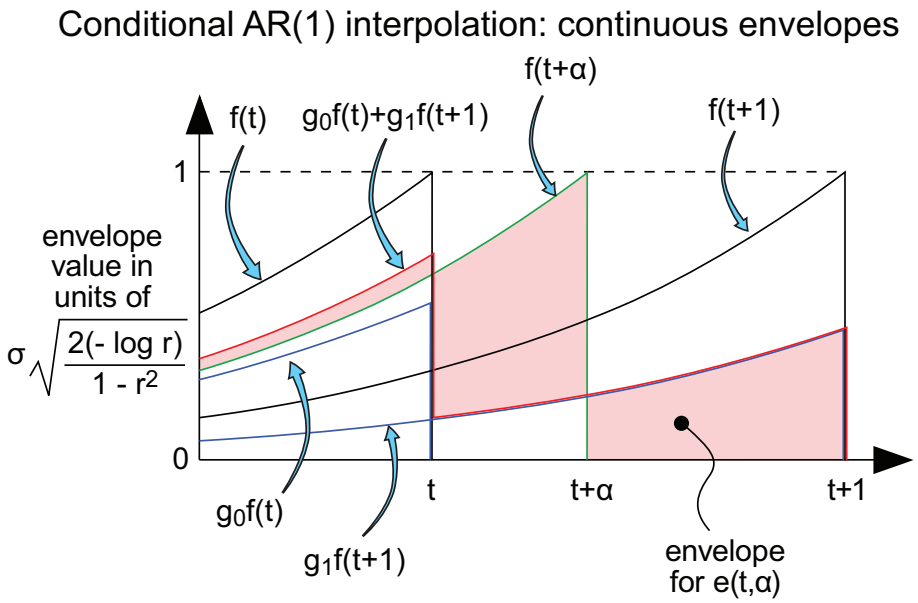

Figure 51: Setting up the conditional AR(1) interpolation problem.

If we minimize $\text{var}[e(t,\alpha)]$ with respect to g_0 and g_1, we get

$$g_0 = \left[\frac{1-r^{2(1-\alpha)}}{1-r^2}\right] r^\alpha, \quad rg_1 = \left[\frac{r^{2(1-\alpha)}-r^2}{1-r^2}\right] r^\alpha \tag{18.14}$$

and the ratio of $\text{var}[e(t,\alpha)]$ to the innovation variance of the original series is:

$$\frac{1}{\sigma^2}\text{var}[e(t,\alpha)] = \frac{1+r^2-r^{2-2\alpha}-r^{2\alpha}}{(1-r^2)^2} \tag{18.15}$$

At this point, this procedure is an educated guess because I have not shown that minimizing $\text{var}[e(t,\alpha)]$ has *anything* to do with autoregressive fields. To test the idea, I generated a conditional AR(1) series with length n, autoregressive root r, and innovation variance 1, and then interpolated $n-1$ points at a known offset α, using the equations above for g_0, g_1, and $\text{var}[e(t,\alpha)]$, and then estimated $\hat{r}$ and $\hat{\sigma}^2$ using just the interpolated points (i.e. no source points). If my procedure has any merit, I should recapture the autoregressive nature of the original series. Figure 52 shows the results that I got.

The results show that my interpolation procedure does indeed capture the autoregressive statistics of the original series. Figure 53 shows the interpolation variance as a function of autoregressive parameter r and offset α.

Now let's compare the AR(1) interpolation result to a simultaneous equivalent. For the original series, we had

$$\begin{aligned} \left|\frac{X(\omega)}{E(\omega)}\right|^2 &= \frac{1}{1+r^2-2r\cos\omega} \\ \rightarrow \left|\frac{X(\omega)}{E(\omega)}\right|^2_{1/2} &= \frac{1-r}{1-r^2}\frac{1}{1+r-2\sqrt{r}\cos\omega} \end{aligned} \tag{18.16}$$

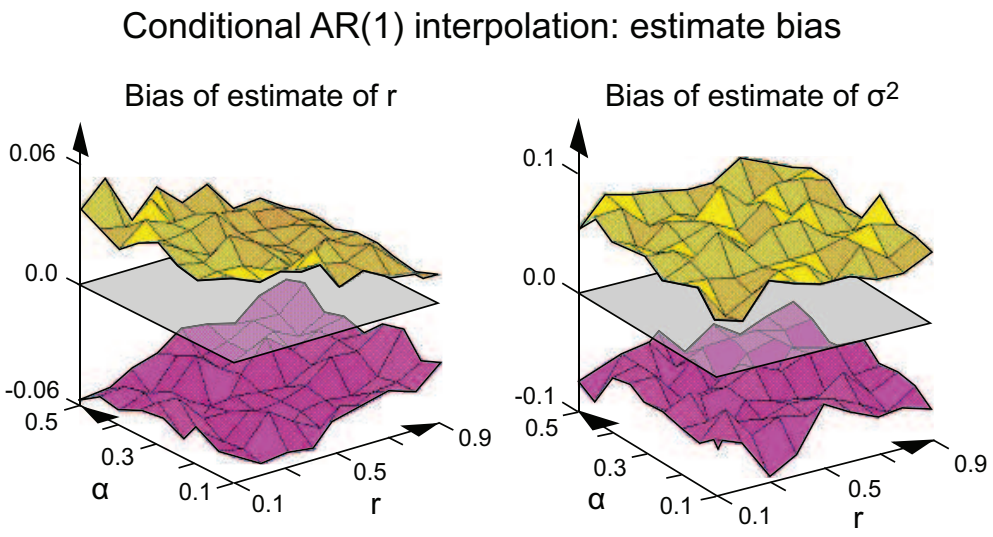

Figure 52: Differences between estimated and theoretical values of autoregressive root and innovation variance for an interpolated AR(1) series. The series length is 2048, and the chart shows the 95% confidence intervals over 25 trials.

Now convert that to simultaneous form:

$$\left[\frac{X(\omega)}{E(\omega)}\right]_{1/2} = \frac{1}{1+r}\frac{1}{\sqrt{1-2\beta\cos\omega}} \qquad (\beta = \frac{\sqrt{r}}{1+r}) \tag{18.17}$$

To second order in β and $\cos\omega$, that is

$$\left[\frac{X(\omega)}{E(\omega)}\right]_{1/2} = \frac{1}{1+r}\frac{1}{(1-\frac{1}{4}\beta^2)-\beta\cos\omega-\frac{1}{4}\beta^2\cos 2\omega\ldots} \tag{18.18}$$

So the equivalent simultaneous form would suggest

$$\begin{aligned}(1-\frac{1}{4}\beta^2)x(t+\frac{1}{2}) &= \frac{1}{2}\beta[x(t)+x(t-1)] \\ &+\frac{1}{8}\beta^2[x(t-\frac{1}{2})+x(t+\frac{3}{2})]+\ldots+\frac{1}{1+r}\sigma z\end{aligned} \tag{18.19}$$

where z is a standard normal $N(0,1)$ variate. Our interpolation procedure, on the other hand, would give

$$x(t+\frac{1}{2}) = \beta[x(t)+x(t+1)]+\frac{1}{1+r}\sigma z \tag{18.20}$$

So, in effect, the interpolation is "squeezing" the effect of all the neglected terms of the simultaneous expansion into its larger coefficient of $[x(t)+x(t+1)]$. Appendix C later on talks about the continuous propagator for a conditional AR(2) system.

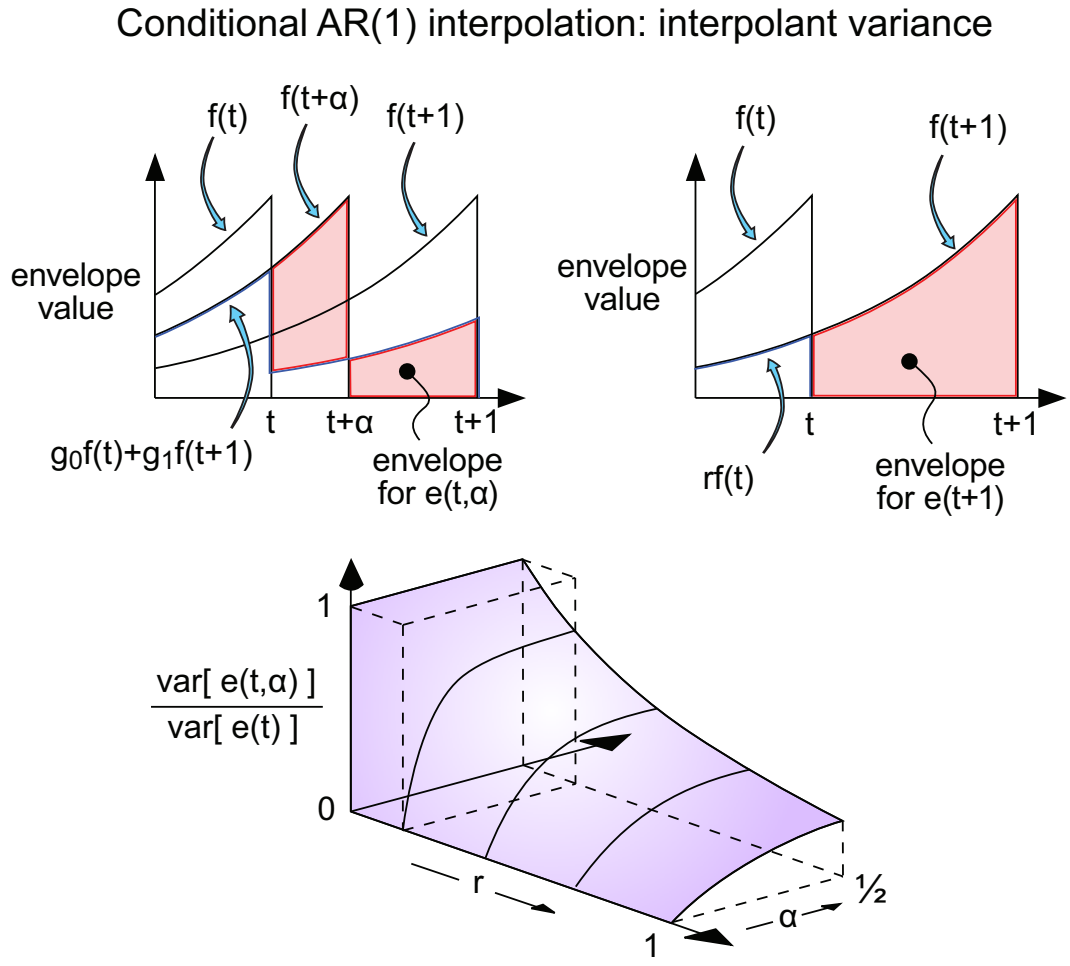

Figure 53: The form of the interpolation variance as a function of autoregressive root and offset for a conditional AR(1) series.

19 Interpolation into a simultaneous AR(1) field

Let's start with the defining equations and autocovariance of a 1D simultaneous AR(1) process with Gaussian innovations:

$$\begin{aligned}(1+r^2)x(t) &= r[x(t-1)+x(t+1)]+(1-r^2)e(t), \quad x(t)=y(t)-\mu, \quad e(t)\simeq N(0,\sigma^2)\\ \frac{1}{\sigma^2}\gamma(s) &= r^{|s|}\left[\frac{1+r^2}{1-r^2}+|s|\right] \qquad [\text{integral } s]\end{aligned} \tag{19.1}$$

If we assume that the series has been plucked from a continuous random field, then the continuous autocovariance would be

$$\frac{1}{\sigma^2}\gamma(u) = r^{|u|}\left[\frac{1+r^2}{1-r^2}+|u|\right] \qquad [\text{real } u] \tag{19.2}$$

Setting $p=-\log r$ and $z=pu$, that is equivalent to

$$\frac{\gamma(z)}{\gamma(0)} = e^{-z}\left[1+\left(\frac{\tanh p}{p}\right)z\right] \tag{19.3}$$

which I have illustrated in Figure 54.

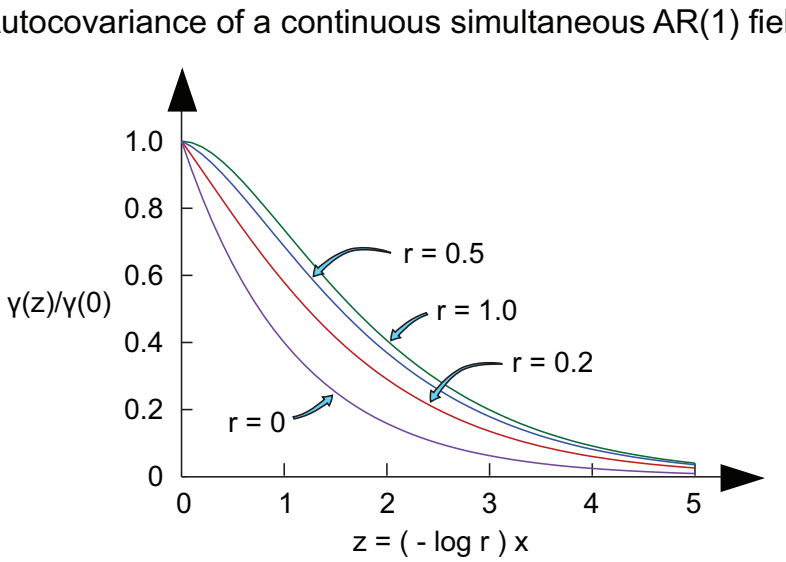

Figure 54: For a simultaneous AR(1) series, the shape of the autocovariance function varies significantly with the autoregressive root.

Taking the continuous Fourier transform yields a spectral density of

$$\frac{1}{2\sigma^2}|G(\omega)|^2 = \frac{1+r^2}{1-r^2}\frac{p}{p^2+\omega^2}+\frac{p^2-\omega^2}{(p^2+\omega^2)^2} \tag{19.4}$$

Here, we will choose a constant phase of zero for our propagator, i.e.

$$\frac{1}{2\sigma^2}[G(\omega)]^2 = \frac{1+r^2}{1-r^2}\frac{p}{p^2+\omega^2}+\frac{p^2-\omega^2}{(p^2+\omega^2)^2} \tag{19.5}$$

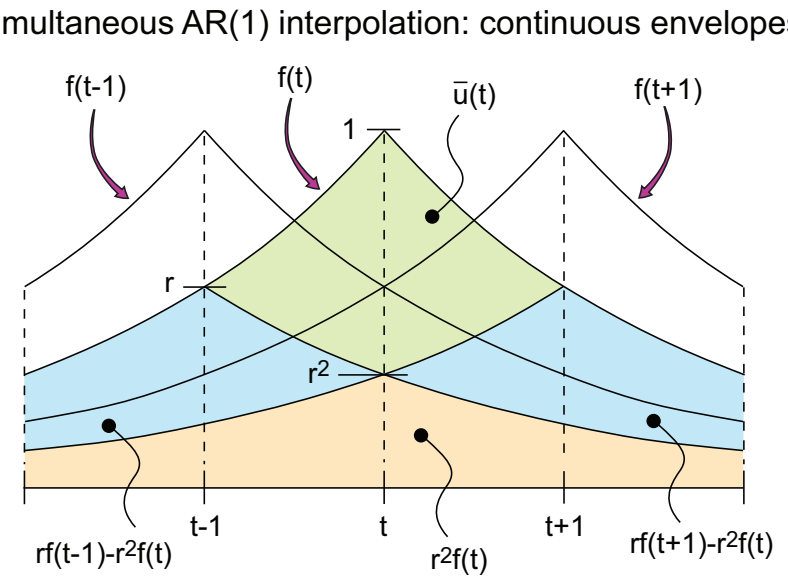

Figure 55: If we take our simulataneous AR(1) propagator to be $g(v) = \sigma r^{|v|}$, then a simple geometric interpretation of the quantities in the defining equation is possible.

For smallish p (i.e. r near 1), we get

$$\frac{1}{\sigma}G(\omega) \approx \frac{2p}{p^2+\omega^2} = \frac{1}{p+j\omega} + \frac{1}{p-j\omega} \tag{19.6}$$

and the corresponding propagator is

$$\frac{1}{\sigma}g(v) = e^{-p|v|} = r^{|v|} \tag{19.7}$$

For the moment, we will take $g(v) = \sigma r^{|v|}$ as an approximate continuous propagator. That then leads to a geometric interpretation of the quantities entering into the discrete series equations

$$\begin{aligned} (1+r^2)x(t) &= r[x(t-1)+x(t+1)] + (1-r^2)e(t) \\ x(t) &= e(t) + \sum_{j=1}^{\infty} r^j[e(t+j)+e(t-j)] \end{aligned} \tag{19.8}$$

Let $f(t)$ be the envelope of $x(t)$, and let $\bar{e}(t)$ be the envelope of $e(t)$. Using the envelope functions as illustrated in Figures 55 and 56, we get

$$\begin{aligned} f(t) &= [rf(t-1) - r^2 f(t)] + [rf(t+1) - r^2 f(t)] + r^2 f(t) + \bar{u}(t) \\ &= \bar{e}(t) + r[\bar{e}(t-1) + \bar{e}(t+1)] + r^2[\bar{e}(t-2) + \bar{e}(t+2)] + \ldots \\ \bar{e}(t) &= \bar{u}(t) + r^2\bar{u}(t) + r^4\bar{u}(t) + \ldots \\ &= \frac{1}{1-r^2}\bar{u}(t) \end{aligned} \tag{19.9}$$

Integrating the envelope equation over a continuous unit Gaussian noise function gives back our discrete defining equation:

$$\begin{aligned} (1+r^2)f(t) &= r[f(t-1)+f(t+1)] + (1-r^2)\bar{e}(t) \\ \rightarrow (1+r^2)x(t) &= r[x(t-1)+x(t+1)] + (1-r^2)e(t) \end{aligned} \tag{19.10}$$

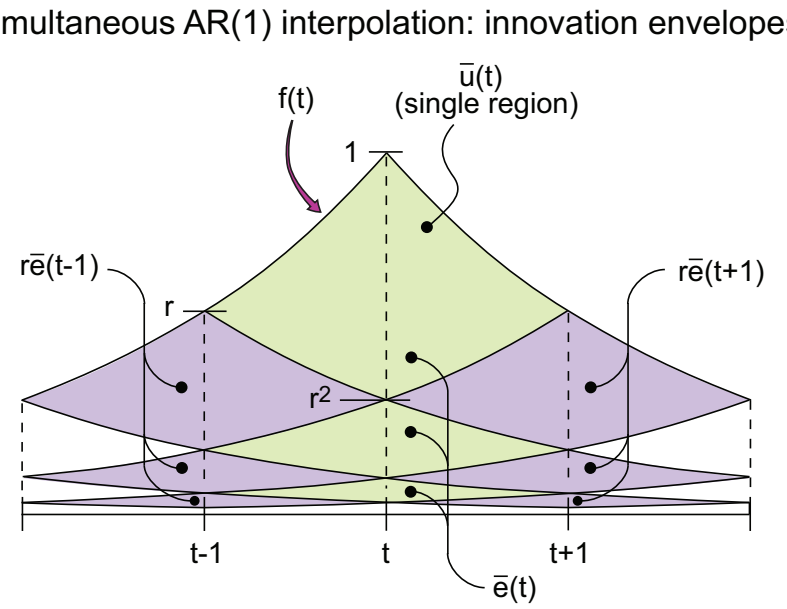

Figure 56: A geometric interpretation of the innovations for a simultaneous AR(1) series.

However, the character of our initial approximation $g(v) = \sigma r^{|v|}$ shows up in the fact that the quantities $u(t) = (1 - r^2)e(t)$ are *not* independent, as illustrated in Figure 57.

They would be independent only if they did not overlap in t (see Equation 13.10 above). Keeping the nature and limitations of our approximation in mind, then, let's set up the interpolation problem as illustrated in Figure 58.

There are now four regions of interest, and we have two possible algorithms. If we require that the linear combination $g_0 f(t) + g_1 f(t+1)$ match up with $f(t+\alpha)$ outside the interval $[t, t+1]$, then we get

$$\begin{aligned} \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \end{bmatrix} &= \begin{bmatrix} r^\alpha \\ r^{1-\alpha} \end{bmatrix} \\ \rightarrow \begin{bmatrix} g_0 \\ g_1 \end{bmatrix} &= \frac{2r}{1-r^2} \begin{bmatrix} \sinh(1-\alpha)(-\log r) \\ \sinh \alpha(-\log r) \end{bmatrix} \end{aligned} \tag{19.11}$$

However, if we require instead that $\text{var}[e(t, \alpha)]$ be a minimum, where

$$\frac{1}{\sigma^2}\text{var}[e(t,\alpha)] = \int_{-\infty}^{+\infty} [g_0 f(t,t') + g_1 f(t+1,t') - f(t+\alpha,t')]^2 \mathrm{d}t' \tag{19.12}$$

and $f(t, t') = r^{|t-t'|}$, then the result is

$$\begin{bmatrix} 1 & r(1-\log r) \\ r(1-\log r) & 1 \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \end{bmatrix} = \begin{bmatrix} r^\alpha(1-\alpha\log r) \\ r^{1-\alpha}[1-(1-\alpha)\log r] \end{bmatrix} \tag{19.13}$$

Note that in the conditional AR(1) case, these two algorithms yield the same result. In the simultaneous AR(1) case, however, the two algorithms yield the same result only in the limiting cases of $\alpha = 0$, $\alpha = 1$, or $r \approx 1$. In other cases, and particularly when r is small but non-zero, the two different approximations look like those illustrated in Figure 59.

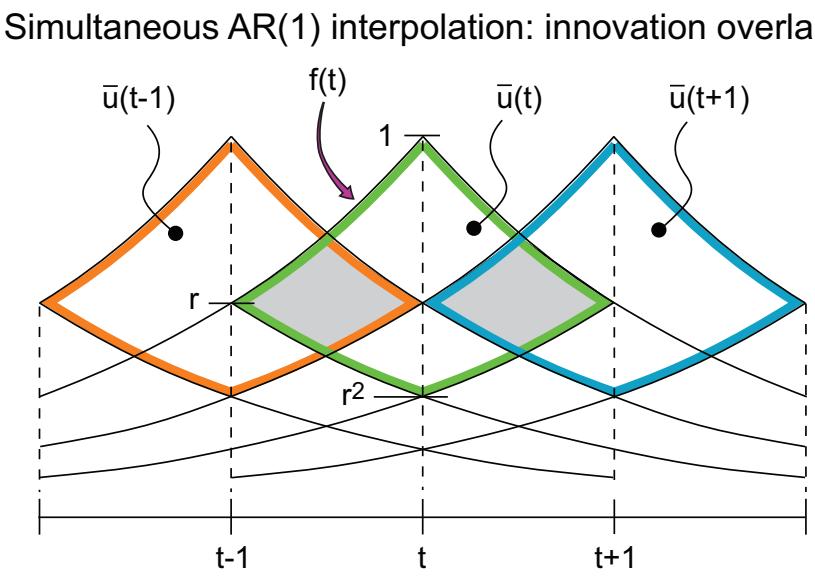

Figure 57: If we take our simulataneous AR(1) propagator to be $g(v) = \sigma r^{|v|}$, then the series innovations are not independent.

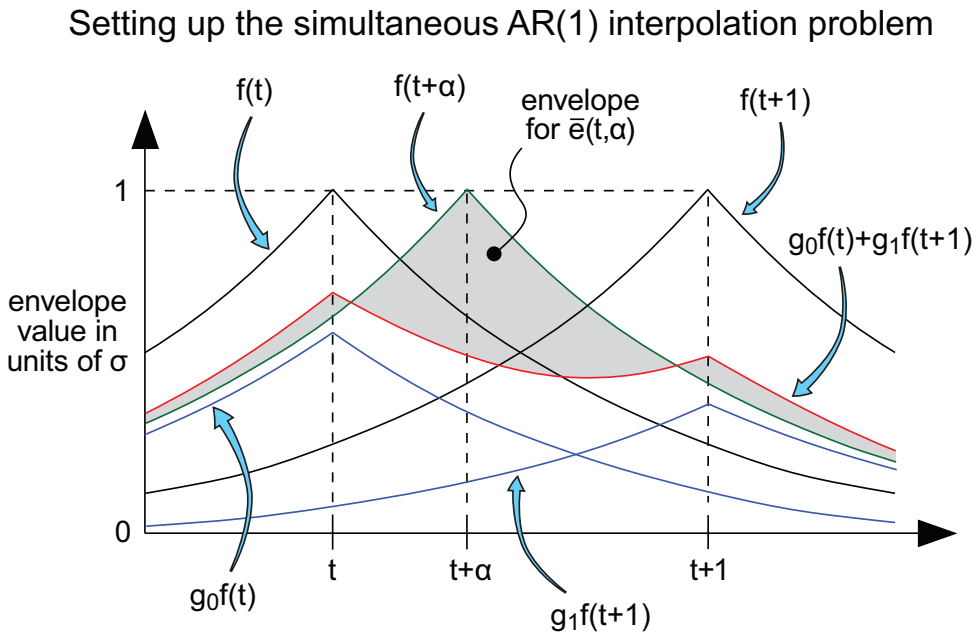

Figure 58: Setting up the interpolation problem for a simultaneous AR(1) series.

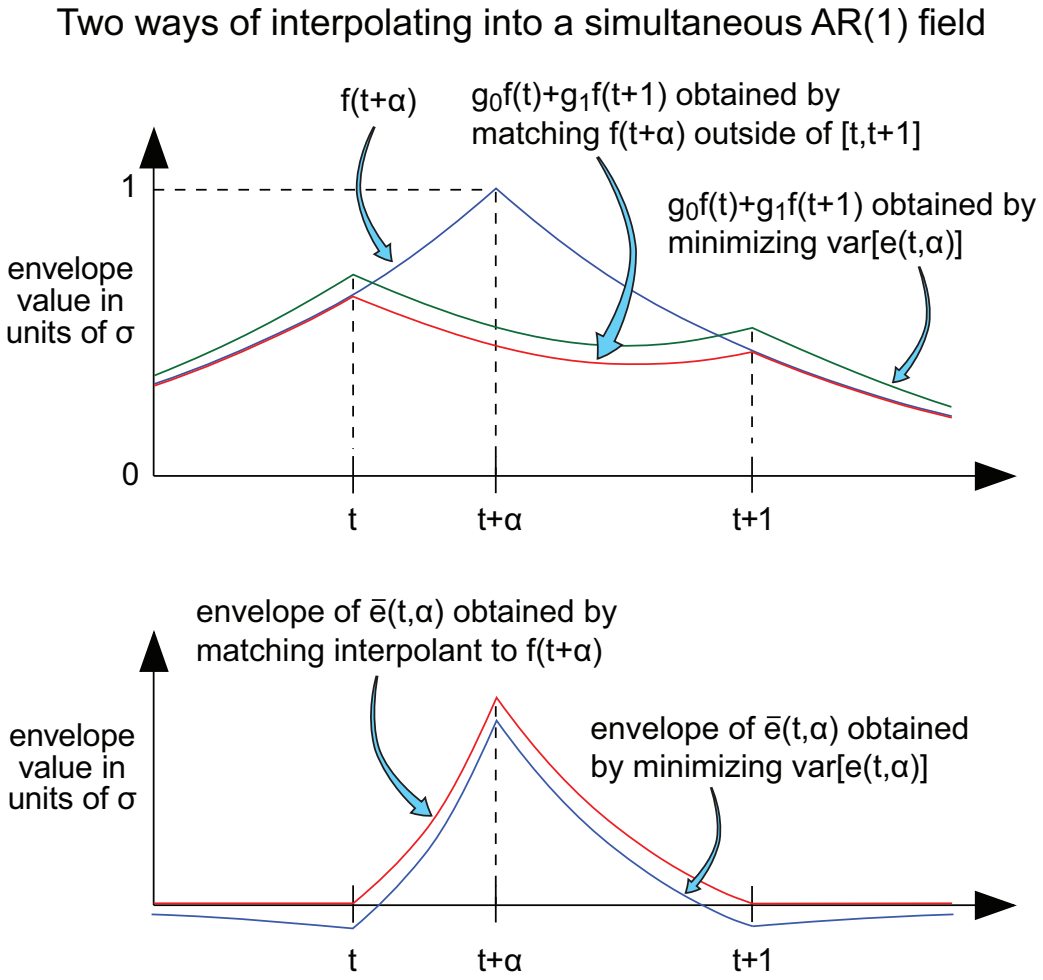

Figure 59: There are two ways we could solve the simultaneous AR(1) interpolation problem, which lead to different innovation variances.

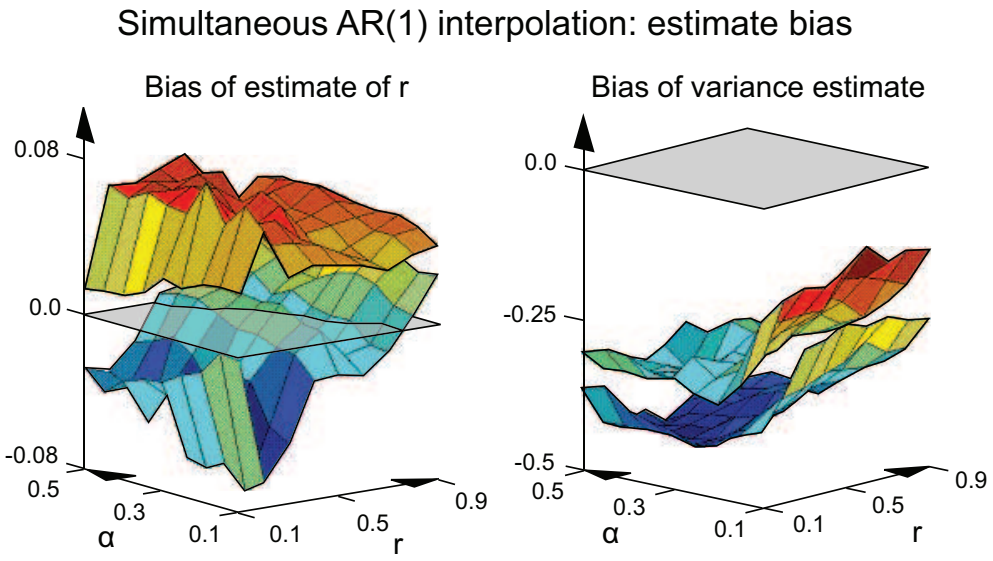

Figure 60: Differences between estimated and theoretical values of autoregressive root and modified innovation variance for an interpolated simultaneous AR(1) series. The series length is 2048, the Hamming window size is 32, and the chart shows the 95% confidence limits over 25 trials.

Now, we don't want to get *too* worked up about these two different approximations. The case in which r is small but non-zero corresponds to a weakly correlated field, which will be noise-like, so slight errors in innovation variance may well be unimportant. A more pressing question is, how closely does this interpolant capture the statistics of the original field? To answer this, I generated a simultaneous series with length n, autoregressive root r, and modified innovation variance $(1-r^2)\sigma^2/(1+r^2) = 1$, and interpolated $n-1$ points at a known offset α using equation (19.11), and then estimated $\hat{r}$ and $(1-\hat{r}^2)\hat{\sigma}^2/(1+\hat{r}^2)$ using *just* the interpolated points. I estimated $\hat{r}$ by fitting a cosine series to the logarithm of the smoothed and averaged spectral density (see Section 11 above).

The results, embodied in figure 60, show that the match-up algorithm overestimates r and seriously underestimates the innovation variance. To see why this happens, we need to develop a more accurate version of the simultaneous AR(1) propagator. Our original equations were

$$\begin{aligned} \frac{1}{\sigma^2}\gamma(u) &= r^{|u|}\left[\frac{1+r^2}{1-r^2} + |u|\right] \qquad [\text{real } u] \\ \frac{1}{2\sigma^2}[G(\omega)]^2 &= \frac{1+r^2}{1-r^2}\frac{p}{p^2+\omega^2} + \frac{p^2-\omega^2}{(p^2+\omega^2)^2} \quad (p = -\log r) \end{aligned} \tag{19.14}$$

As for the actual propagator $g(v)$, it *must* have the form $\sigma r^{|v|}$ at large $|v|$ because that is the only way to get the $\sigma^2|u|r^{|u|}$ term of $\gamma(u)$. At small $|v|$, however, the function must look more like

$$g(v) \approx \sigma\sqrt{\frac{1+r^2}{1-r^2}(-\log r)}\, r^{|v|} \quad \text{small } |v| \tag{19.15}$$

in order to reproduce the $(1+r^2)\sigma^2 r^{|u|}/(1-r^2)$ term of $\gamma(u)$. I calculated the inverse

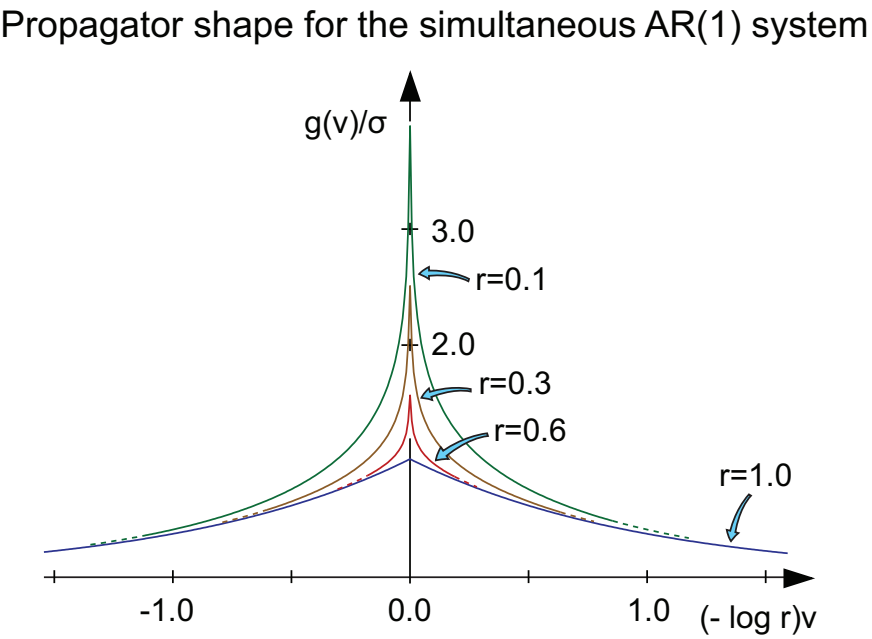

Figure 61: The true propagator shape for a simultaneous AR(1) random field, obtained through numerical integration.

Fourier transform of

$$\frac{1}{2\sigma^2}[G(\omega)]^2 = \frac{1+r^2}{1-r^2}\frac{p}{p^2+\omega^2} + \frac{p^2-\omega^2}{(p^2+\omega^2)^2} \tag{19.16}$$

by numerical means, and got the results encompassed in Figure 61.

So our expectations are indeed fulfilled, and our earlier approximation $g(v) = \sigma r^{|v|}$ is supplemented by a cap, or perturbation, which however always seems to be negligible for $|v| > 1$. So the *true* situation for the simultaneous AR(1) propagator and interpolation problem is something like that shown in Figure 62. Thus, it is indeed possible for the innovations $e(t)$ to be independent, as long as the integrals $\int \bar{e}(t)\bar{e}(t+s)\mathrm{d}t$ are zero, as illustrated in Figure 62.

We can also see why we underestimated the innovation variance in our first stab at a simultaneous AR(1) interpolator. The integral of $\bar{e}^2(t)$ is appreciably larger in the true case than in our first-order approximation. Figure 63 illustrates this fact.

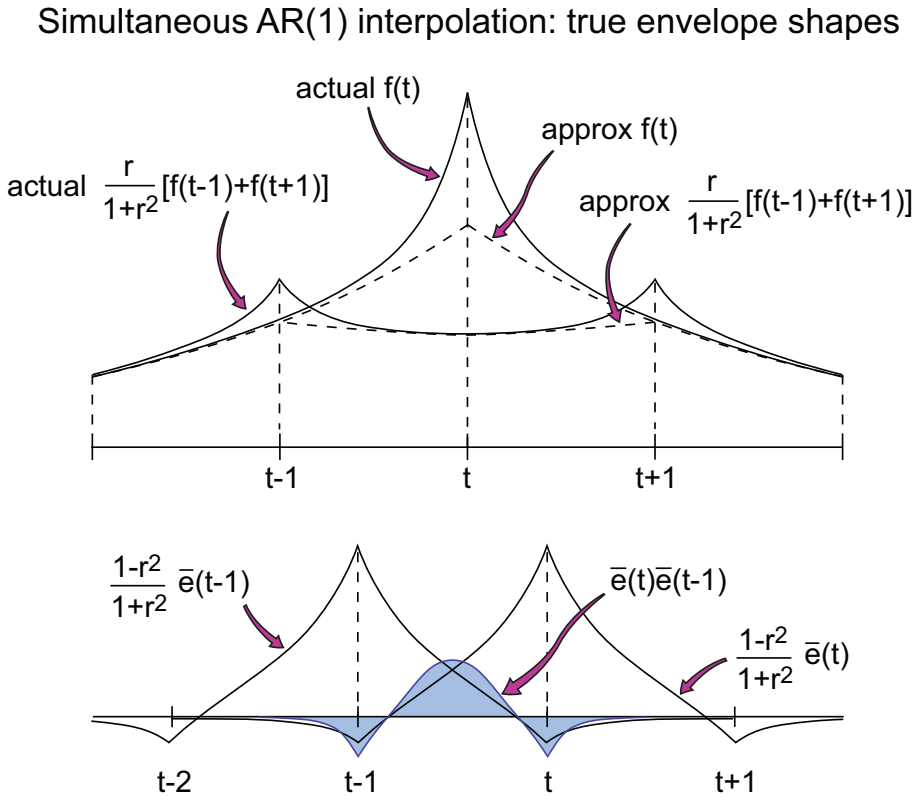

Figure 62: Sketch of the simultaneous AR(1) interpolation problem using the true propagator function. This chart shows how successive innovations can be orthogonal even though their envelopes overlap.

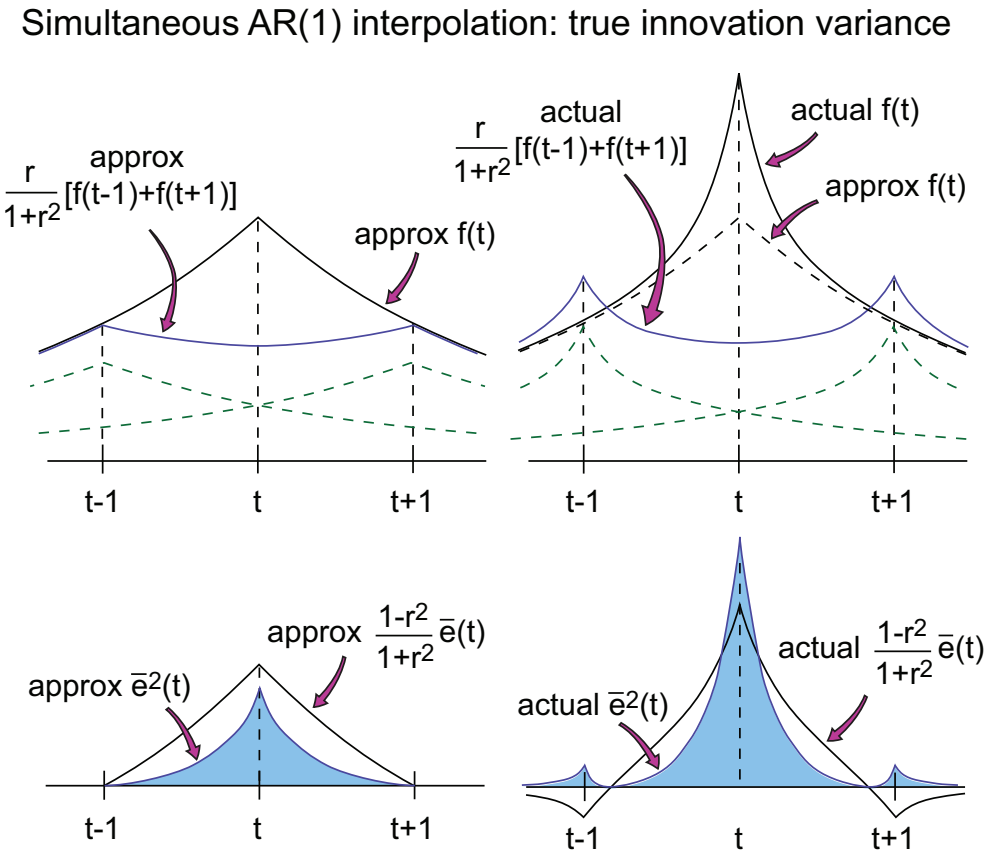

Figure 63: Sketch of the simultaneous AR(1) innovation envelopes using the true propagator function.

20 Wrap-up and acknowledgement

We are at a good stopping point. They say that if you aim for the stars, you may strike the moon. I didn't reach my original goal — arbitrary interpolation into a 2D autoregressive random field — but I found some interesting results along the way.

First of all: my idea concerning the use of a propagator for solving the interpolation problem is sound. The only caveat is that if the model is simultaneous in nature, we have to calculate the propagator function numerically. My approximate one-dimensional version, Equation 19.7, was not good enough to reproduce series statistics accurately.

Next up: standard interpolation methods. These can cause spurious correlations in random fields. Of course, these are calculable, and in Section 14, I show example calculations. In theory then, one can use standard interpolation methods and then reverse the distortion to recover the original discrete autocovariance.

Next up: estimation of parameters for a simultaneous autoregressive field. In this case, the Jacobean factor is not negligible and we cannot use the traditional analysis that looks at the residual sum of squares. Instead, I propose a method that fits a cosine series to the inverse square root of the estimated spectral density. As Figure 21 shows, this method only breaks down when there are double roots close to one in magnitude.

And finally: how *does* an autoregressive model change under rotation and scaling? Here's the prescription:

- Estimate (or calculate) the continuous autocovariance function.
- Sample this autocovariance on a grid with the desired rotation and scaling, averaging if necessary according to the mode of physical sampling.
- Calculate the resulting spectral density function.
- Fit a cosine series to the inverse spectral density to get a conditional model, or to the inverse square root of the spectral density for a simultaneous model.

As for how many terms to include in that cosine series, that can best be decided by using an information criterion such as the Akaike information criterion [48]. My Section 15 illustrates this procedure and its efficacy (see Figures 39 and 40 in particular). Does this suggest any rotational invariants? As I show in Section 17, a Taylor series of the inverse spectral density would have rotational invariants related to its partial derivatives, but these are impractical to compute accurately for the small sample sizes so often met in visual texture classification. A more practical observation is that the locus of poles of the continuous spectral distribution has the same rotation and scaling as the desired transformation, a fact implied by the prescription given earlier.

So what's left to do? Well, the main item left in the agenda is to calculate a general 2D propagator, given a 2D autocovariance. I looked at some simple one-dimensional cases, but we are still a long way from a general 2D case. For me, another interesting question is what happens when a random field is autoregressive in two spatial dimensions and one time dimension, and in particular how it reacts to disturbances distributed over time. We will have to save those questions for the next iteration!

This is also a good point at which to extend, once again, my sincerest thanks to Dave Clausi, for acting both as an instructor and as a thesis supervisor. Dave is a model professor and a great asset to the University of Waterloo.

A Partial fraction expansions

The transfer function of a mean-reduced AR(p) series is

$$\begin{aligned}\frac{X(z)}{E(z)} &= \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_p z^{-p}} \\ &= \frac{z^p}{z^p - a_1 z^{p-1} - a_2 z^{p-2} - \cdots - a_p} \\ &= \frac{z^p}{(z - r_1)(z - r_2)\ldots(z - r_p)} = \frac{z^p}{C(z)} \end{aligned} \tag{A.1}$$

where $\{r_1 \ldots r_p\}$ are the roots of $C(z) = 0$. Near a particular root r_k, the transfer function is dominated by the $(z - r_k)$ term:

$$\begin{aligned}\frac{X(z)}{E(z)} &\approx \frac{z}{z - r_k} \frac{{r_k}^{p-1}}{\prod_{j \neq k}(r_k - r_j)} \quad \text{for } z \approx r_k \\ &\approx \frac{z}{z - r_k} \left[\frac{{r_k}^{p-1}}{C'(r_k)}\right] \quad \text{for } z \approx r_k \\ \Longrightarrow \frac{X(z)}{E(z)} &= \sum_{k=1}^{p} \left[\frac{{r_k}^{p-1}}{C'(r_k)}\right] \frac{1}{1 - r_k z^{-1}} \end{aligned} \tag{A.2}$$

which gives the moving average equivalent directly:

$$x(t) = \sum_{j=0}^{\infty} \sum_{k=1}^{p} \frac{{r_k}^{p-1}}{C'(r_k)} {r_k}^j e(t - j) \tag{A.3}$$

Suppose now that we need the autocovariance of the same series. The discrete spectral density is

$$\left|\frac{X(z)}{E(z)}\right|^2 = \frac{1}{(1 - r_1 z^{-1})\ldots(1 - r_p z^{-1})} \frac{1}{(1 - r_1 z)\ldots(1 - r_p z)} \tag{A.4}$$

where $z = e^{j\omega}$, and ω is the discrete frequency, and hence

$$\left|\frac{X(z)}{E(z)}\right|^2 = \frac{1}{\left[1 + r_1^2 - r_1(z + z^{-1})\right]\ldots\left[1 + r_p^2 - r_p(z + z^{-1})\right]} \tag{A.5}$$

$$\Longrightarrow (1 + r_1^2)\ldots(1 + r_p^2) \left|\frac{X(z)}{E(z)}\right|^2 = \frac{1}{(1 - s_1 u^{-1})\ldots(1 - s_p u^{-1})} = \frac{u^p}{(u - s_1)\ldots(u - s_p)} \tag{A.6}$$

where $u^{-1} = z + z^{-1}$, and $s_1 = r_1/(1 + r_1^2), \ldots s_p = r_p/(1 + r_p^2)$. Define $D(u) = (u - s_1)\ldots(u - s_p)$, and by the results we just developed, we have

$$\begin{aligned}\Longrightarrow (1 + r_1^2)\ldots(1 + r_p^2) \left|\frac{X(z)}{E(z)}\right|^2 &= \sum_{k=1}^{p} \frac{{s_k}^{p-1}}{D'(s_k)} \frac{1}{1 - s_k(z + z^{-1})} \\ &= \sum_{k=1}^{p} \frac{{s_k}^{p-1}}{D'(s_k)} \frac{1 + {r_k}^2}{1 + {r_k}^2 - 2 r_k \cos\omega} \end{aligned} \tag{A.7}$$

Taking the discrete Fourier transform yields

$$(1+r_1^2)\ldots(1+r_p^2)\frac{\gamma(n)}{\sigma^2} = \sum_{k=1} p\frac{{s_k}^{p-1}}{D'(s_k)}\frac{1+{r_k}^2}{1-{r_k}^2}{r_k}^{|n|} \tag{A.8}$$

where σ^2 is the variance of the innovation sequence. For example, suppose we have an AR(2) series

$$\frac{X(z)}{E(z)} = \frac{1}{1-a_1z^{-1}-a_2z^{-2}} = \frac{1}{(1-r_1z^{-1})(1-r_2z^{-1})} \tag{A.9}$$

with innovation variance σ^2. Then

$$\begin{aligned}\left|\frac{X(z)}{E(z)}\right|^2 &= \frac{1}{[1+{r_1}^2-r_1(z+z^{-1})]\,[1+{r_2}^2-r_2(z+z^{-1})]} \\ &= \frac{u^2}{(1+{r_1}^2)(1+{r_2}^2)(u-s_1)(u-s_2)} \\ &= \frac{u^2}{(1+{r_1}^2)(1+{r_2}^2)D(u)}\end{aligned} \tag{A.10}$$

where $u^{-1}=z+z^{-1}, s_1=r_1/(1+{r_1}^2), s_2=r_2/(1+{r_2}^2)$, and $D(u)=u^2-(s_1+s_2)u+s_1s_2$. So $D'(u)=2u-(s_1+s_2)$, and

$$(1+{r_1}^2)(1+{r_2}^2)\left|\frac{X(z)}{E(z)}\right|^2 = \frac{s_1}{s_1-s_2}\frac{1}{1-s_1u^{-1}} + \frac{s_2}{s_2-s_1}\frac{1}{1-s_2u^{-1}} \tag{A.11}$$

and substituting $s_1=r_1/(1+{r_1}^2), s_2=r_2/(1+{r_2}^2)$ gives

$$\left|\frac{X(z)}{E(z)}\right|^2 = \frac{1}{(r_1-r_2)(1-r_1r_2)}\left[\frac{r_1}{1+{r_1}^2-r_1(z+z^{-1})} - \frac{r_2}{1+{r_2}^2-r_2(z+z^{-1})}\right] \tag{A.12}$$

Recalling that the discrete autocovariance resulting from a spectral density component such as $1/[1+r^2-r(z+z^{-1})]$ is $r^{|n|}/(1-r^2)$, we have for this AR(2) process

$$\frac{\gamma(n)}{\sigma^2} = \frac{1}{(r_1-r_2)(1-r_1r_2)}\left[\frac{{r_1}^{|n|+1}}{1-{r_1}^2} - \frac{{r_2}^{|n|+1}}{1-{r_2}^2}\right] \tag{A.13}$$

and the corresponding marginal variance is

$$\begin{aligned}\frac{\gamma(0)}{\sigma^2} &= \frac{1+r_1r_2}{(1-r_1r_2)(1-{r_1}^2)(1-{r_2}^2)} \\ &= \frac{1-a_2}{(1+a_2)(1-a_1-a_2)(1+a_1-a_2)}\end{aligned} \tag{A.14}$$

Requiring $1+a_2>0, 1-a_1-a_2>0$, and $1+a_1-a_2>0$ gives the allowed AR(2) parameter space directly.

Now suppose that we have instead a simultaneous AR(p) series with transfer function

$$\frac{X(z)}{E(z)} = \frac{1}{1-\beta_1(z+z^{-1})-\beta_2(z^2+z^{-2})-\ldots-\beta_p(z^p+z^{-p})} \tag{A.15}$$

To proceed further, we need $z^n + z^{-n}$ in terms of powers of $z + z^{-1}$, or equivalently, $2\cos n\theta$ in terms of powers of $2\cos\theta$. Fortunately, the Chebyshev polynomials of the first kind do that for us [49]:

$$\begin{aligned} T_n(\cos\theta) &= \cos n\theta \\ 2\cos n\theta &= n\sum_{m=0}^{[n/2]})(-1)^m \frac{(n-m-1)!}{m!(n-2m)!}(2\cos\theta)^{n-2m} \end{aligned} \tag{A.16}$$

which leads to the identities

$$\begin{aligned} z^2 + z^{-2} &= (z+z^{-1})^2 - 2 \\ z^3 + z^{-3} &= (z+z^{-1})^3 - 3(z+z^{-1}) \\ z^4 + z^{-4} &= (z+z^{-1})^4 - 4(z+z^{-1})^2 + 2 \quad \text{etc.} \end{aligned} \tag{A.17}$$

Using these polynomials, we get

$$\begin{aligned} 1 - \beta_1(z+z^{-1}) - \quad \ldots \quad & -\beta_p(z^p + z^{-p}) \\ &= c_0 + c_1(z+z^{-1}) + c_2(z+z^{-1})^2 + \ldots + c_p(z+z^{-1})^p \\ &= c_0\left[1 - s_1(z+z^{-1})\right]\left[1 - s_2(z+z^{-1})\right]\ldots\left[1 - s_p(z+z^{-1})\right] \\ &= c_0(1-s_1u^{-1})(1-s_2u^{-1})\ldots(1-s_pu^{-1}) \end{aligned} \tag{A.18}$$

where $\{c_0 \ldots c_p\}$ are functions of $\{\beta_1 \ldots \beta_p\}$, $\{s_1 \ldots s_p\}$ satisfy $D(u) = c_0u^p + c_1u^{p-1} + \ldots + c_p = 0$, and $u^{-1} = z + z^{-1}$. Thus, for example, a simultaneous AR(4) series would have

$$\begin{aligned} \left[\frac{X(z)}{E(z)}\right]^{-1} &= 1 - \beta_1(z+z^{-1}) - \beta_2(z^2+z^{-2}) - \beta_3(z^3+z^{-3}) - \beta_4(z^4+z^{-4}) \\ D(u) &= (1 + 2\beta_2 - 2\beta_4)u^4 + (3\beta_3 - \beta_1)u^3 + (4\beta_4 - \beta_2)u^2 - \beta_3 u - \beta_4 \end{aligned}$$

In the general case, we have

$$\begin{aligned} \frac{X(z)}{E(z)} &= \frac{u^p}{c_0(u-s_1)\ldots(u-s_p)} \\ &= \sum_{k=1}^{p} \frac{{s_k}^{p-1}}{D'(s_k)} \frac{1}{1 - s_k(z+z^{-1})} \\ &= \sum_{k=1}^{p} \frac{{s_k}^{p-1}}{D'(s_k)} \frac{1+{r_k}^2}{1-{r_k}^2} \frac{1-{r_k}^2}{1+{r_k}^2 - r_k(z+z^{-1})} \end{aligned} \tag{A.19}$$

where for each r_k, $r_k/(1+{r_k}^2) = s_k$ and $|r_k| < 1$. The moving average equivalent is then

$$x(t) = \sum_{k=1}^{p} \frac{{s_k}^{p-1}}{D'(s_k)} \frac{1+{r_k}^2}{1-{r_k}^2} \left[e(t) + \sum_{j=1}^{\infty} {r_k}^j [e(t-j) + e(t+j)] \right] \tag{A.20}$$

As for the autocovariance, let

$$\begin{aligned} \frac{X(z)}{E(z)} &= \sum_{k=1}^{p} \frac{{s_k}^{p-1}}{D'(s_k)} \frac{1+{r_k}^2}{1+{r_k}^2 - r_k(z+z^{-1})} \\ &= \sum_{k=1}^{p} \frac{d_k}{1+{r_k}^2 - r_k(z+z^{-1})}, \quad d_k = \frac{{s_k}^{p-1}(1+{r_k}^2)}{D'(s_k)} \end{aligned} \tag{A.21}$$

Then the corresponding discrete spectral density is

$$\left|\frac{X(z)}{E(z)}\right|^2 = \sum_k \left[\frac{d_k}{1+{r_k}^2-r_k(z+z^{-1})}\right]^2 + 2\sum_k\sum_{j\neq k}\frac{d_k d_j}{[1+{r_k}^2-r_k(z+z^{-1})][1+{r_j}^2-r_j(z+z^{-1})]} \quad \text{(A.22)}$$

Performing the inverse discrete Fourier transform, and making use of Equations 6.8 and 21.13, gives the following autocovariance:

$$\frac{\gamma(n)}{\sigma^2} = \sum_k \left(\frac{d_k}{1-{r_k}^2}\right)^2 {r_k}^{|n|}\left[\frac{1+{r_k}^2}{1-{r_k}^2}+|n|\right] 2\sum_k\sum_{j\neq k}\frac{d_j d_k}{(r_j-r_k)(1-r_j r_k)}\left[\frac{{r_j}^{|n|+1}}{1-{r_j}^2}-\frac{{r_k}^{|n|+1}}{1-{r_k}^2}\right] \quad \text{(A.23)}$$

B Maximum likelihood estimates for a simultaneous AR(p)series

I will develop this argument in 1D for simplicity and clarity. Autoregressive series divide into two camps: the conditional (one-sided) and the simultaneous (two-sided). A conditional autoregressive AR(1) model with Gaussian innovations would have

$$[y(t)-\mu] = r[y(t-1)-\mu] + e(t) \text{ i.e. } x(t) = rx(t-1) + e(t) \tag{B.1}$$

where $y(t)$ is the observed series, $x(t) = y(t) - \mu$ is the mean-reduced series, μ is the series mean, and $e(t) \sim N(0, \sigma^2)$ are the innovations. Letting $X(z)$ and $E(z)$ be the z-transforms of $x(t)$ and $e(t)$ respectively, we have

$$\begin{aligned} X(z) &= (1-rz^{-1})^{-1}E(z) \\ &= (1+rz^{-1}+r^2z^{-2}+\ldots)E(z) \qquad (|r|<1) \\ \rightarrow x(t) &= e(t)+re(t-1)+r^2e(t-2)+\ldots \\ \rightarrow \text{var}[x(t)] &= (1+r^2+r^4+\ldots)\sigma^2 \\ &= \sigma^2/(1-r^2) \end{aligned} \tag{B.2}$$

Thus the marginal series variance is $\sigma^2/(1-r^2)$. If we select a group of n successive variates from the series, starting at index $t+1$, then the first will have an apparent marginal probability density of $N(\mu, \sigma^2/(1-r^2))$. Collecting the mean-reduced sample $\{x(t+1), x(t+2), \ldots x(t+n)\}$ and corresponding innovations $\{e(t+1), e(t+2), \ldots e(t+n)\}$ into vectors $\vec{x}$ and $\vec{e}$ respectively, their relation is

$$\begin{bmatrix} \sqrt{1-r^2} & 0 & \cdots & 0 & 0 \\ -r & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -r & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \qquad i.e. \quad L\vec{x} = \vec{e} \tag{B.3}$$

and the corresponding probability density is

$$\begin{aligned} p(\vec{e}) &= \frac{1}{(2\pi)^{n/2}\sigma^n} \exp(-\frac{1}{2}\vec{e}^T\vec{e}) \\ p(\vec{x}) &= \frac{1}{(2\pi)^{n/2}\sigma^n} \sqrt{|\det Q|} \exp(-\frac{1}{2}\vec{x}^T Q\vec{x}) \end{aligned} \tag{B.4}$$

where $Q = L^T L$ is the precision matrix. Forming the maximum likelihood equation for the autoregressive parameter r, we get (see Section 9 above)

$$0 = \frac{\partial}{\partial r}\left[(1-r^2)^{-1/n}\left[(1-r^2){x_1}^2 + \sum_{j=2}^{n}(x_j - rx_{j-1})^2\right]\right] \tag{B.5}$$

So the usual residual sum of squares is modified by *two* effects of order $1/n$: the Jacobean factor $(1-r^2)^{-1/n}$ and the edge term $(1-r^2){x_1}^2$. If $n \gg 1$ and $|r| < 1 - 1/\sqrt{n}$, then we can drop both factors and get the usual estimator

$$\hat{r} = [\sum_{j=1}^{n-1} x_{j+1}x_j]/[\sum_{j=1}^{n-1} {x_j}^2] \tag{B.6}$$

For simultaneous autoregressions, things are not so simple: the Jacobean factor is no longer of order $1/n$. To illustrate this, suppose we have a simultaneous AR(1) model of the form

$$y(t) - \mu = \beta[y(t+1) - \mu] + \beta[y(t-1) - \mu] + e(t) \tag{B.7}$$

where $x(t) = y(t) - \mu$ is the mean-reduced series and $e(t) \sim N(0, \sigma^2)$. Furthermore, we will make the circulant approximation so that for a finite sample of n successive variates, we have

$$\begin{bmatrix} 1 & -\beta & 0 & \cdots & 0 & -\beta \\ -\beta & 1 & -\beta & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\beta & 0 & 0 & \cdots & -\beta & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \qquad i.e. \quad S\vec{x} = \vec{e} \tag{B.8}$$

The corresponding probability density is

$$p(\vec{x}) = \frac{1}{(2\pi)^{n/2}\sigma^n} \sqrt{|\det Q|} \exp(-\frac{1}{2}\vec{x}^T Q \vec{x}) \tag{B.9}$$

where the precision matrix is now $Q = S^2$. As for $\det Q$, first note that if

$$L = \begin{bmatrix} 1 & 0 & \cdots & 0 & -r \\ -r & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -r & 1 \end{bmatrix} \quad (n \text{ rows}) \tag{B.10}$$

then $\det L = 1 - r^n$, and

$$L^T L = \begin{bmatrix} 1+r^2 & -r & 0 & \cdots & 0 & -r \\ -r & 1+r^2 & -r & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -r & 0 & 0 & \cdots & -r & 1+r^2 \end{bmatrix} \quad (n \text{ rows}) \tag{B.11}$$

So setting $\beta = r/(1+r^2)$, we get

$$\begin{aligned} \det L^T L &= (1-r^n)^2 = (1+r^2)^n \det S \\ \rightarrow \sqrt{\det Q} &= \det S = \frac{(1-r^n)^2}{(1+r^2)^n} \end{aligned} \tag{B.12}$$

and putting that in terms of β gives

$$\sqrt{\det Q} = \frac{\left[1 - [(1-\sqrt{1-4\beta^2})/(2\beta)]^n\right]^2}{\left[1 + [(1-\sqrt{1-4\beta^2})/(2\beta)]^2\right]^n} \tag{B.13}$$

Now when we form the maximum likelihood equation for β, the Jacobean factor is

$$(\det Q)^{-1/n} = \frac{\left[1 - [(1-\sqrt{1-4\beta^2})/(2\beta)]^n\right]^{-4/n}}{\left[1 + [(1-\sqrt{1-4\beta^2})/(2\beta)]^2\right]^{-2}} \tag{B.14}$$

which is now of order 1 in β, *not* of order $1/n$, so we can't drop it.

What to do? We'll reformulate the problem. Suppose first that the precision matrix Q corresponding to a simultaneous AR(p) system is circulant with base

$$\vec{q} = [b_0, -b_1, \ldots - b_p, 0, \ldots 0, -b_p, \ldots - b_1] \tag{B.15}$$

That gives us one extra unknown, so we must eventually impose one constraint on $\{b_0, \ldots b_p\}$. The precision matrix is then diagonalized by the discrete Fourier transform:

$$\begin{aligned} [F]_{km} &= \frac{1}{\sqrt{n}} \exp(-2\pi j \frac{km}{n}), \qquad 0 \le k, m < n \\ Q &= F\Lambda F^H, \qquad F^H = F^{-1} \\ \Lambda &= \sqrt{n}\, \mathrm{diag}(F\vec{q}) \end{aligned} \tag{B.16}$$

Our probability density was

$$p(\vec{x}) = \frac{1}{(2\pi)^{n/2}\sigma^n} \sqrt{|\det Q|} \exp(-\frac{1}{2}\vec{x}^T Q \vec{x}) \tag{B.17}$$

Here we have to note that

$$\begin{aligned} \vec{x}^T Q \vec{x} &= \vec{x}^T (F\Lambda F^H)\vec{x} = (F^H\vec{x})^H \Lambda (F^H \vec{x}) \\ &= \sum_{k=0}^{n-1} \lambda(k) |[F^H \vec{x}]_k|^2 = \sum_{k=0}^{n-1} \lambda(k) I_n(k) \end{aligned} \tag{B.18}$$

which is the inner product of the eigenvalue sequence of Q and the n-point periodogram of $\vec{x}$. Here $\lambda(k) = \sqrt{n}[F\vec{q}]_k$. The determinant of the precision matrix Q is the product of its eigenvalues, so if we can hold that product constant while minimizing $\sum \lambda(k)|[F^H\vec{x}]_k|^2$, we're home-free!

So here's the battle plan: the periodogram $\{I_n(k), 0 \le k \le n-1\}$ does not depend on the precision base $\{b_0, -b_1, -b_2, \ldots - b_2, -b_1\}$, so we may scale it arbitrarily. We will scale it so that the average logarithm is zero (i.e. the geometric mean is one):

$$\begin{aligned} \tilde{I}_n(k) &= [\text{scale factor}] I_n(k) = C_{GN} I_n(k) \\ \prod_{k=0}^{n-1} \tilde{I}_n(k) &= 1, \qquad \sum_{k=0}^{n-1} \log \tilde{I}_n(k) = 0 \end{aligned} \tag{B.19}$$

In this case, I say that the sequence $\tilde{I}_n(k)$ is geometrically normalized. We will constrain the eigenvalue sequence of Q, namely $\lambda(k) = \sqrt{n}[F\vec{q}]_k$, to be geometrically normalized also. This provides the extra constraint mentioned earlier. Let m be our Lagrange multiplier; then the minimization target is

$$T = \sum_{k=0}^{n-1} \lambda(k) \tilde{I}_n(k) - m \sum_{k=0}^{n-1} \log \lambda(k) \tag{B.20}$$

If we estimate all possible autoregressive parameters $\{b_0, b_1, \ldots\}$, then we merely accomplish a change of basis between the precision matrix base $\vec{q}$ and its Fourier transform

$\lambda(k) = \sqrt{n}[F\vec{q}]_k$, so we can minimize the target with respect to the eigenvalue sequence $\{\lambda(k)\}$ as opposed to the base $\vec{q}$. Doing so gives the normal equations

$$\begin{aligned} \frac{\partial T}{\partial m} = 0 \quad &\rightarrow \quad \sum_{k=0}^{n-1} \log \lambda(k) = 0 \\ \frac{\partial T}{\partial \lambda(k)} = 0 \quad &\rightarrow \quad \tilde{I}_n(k) - \frac{m}{\lambda(k)} = 0 \end{aligned} \tag{B.21}$$

hence $\tilde{I}_n(k)\lambda(k) = m$. But the product of two geometrically normalized sequences is also geometrically normalized, so

$$\begin{aligned} 0 \quad &= \quad \sum_{k=0}^{n-1} \log[\tilde{I}_n(k)\lambda(k)] \\ \rightarrow 0 \quad &= \quad n \log m, \quad m = 1, \quad \lambda(k) = \frac{1}{\tilde{I}_n(k)} \end{aligned} \tag{B.22}$$

Expressing that in words: the maximum likelihood estimate of the precision matrix base is obtained by fitting a cosine series to the *inverse* of the geometrically normalized sample periodogram. Expressing that as another equation:

$$\sqrt{n}[F\vec{q}]_k = C_{GN}^{-1} |[F^H \vec{x}]_k|^{-2} \tag{B.23}$$

where C_{GN} is the scale factor needed to geometrically normalize the sample periodogram $I_n(k) = |[F^H \vec{x}]_k|^2$.

For a simultaneous autoregressive model, the observation and innovation vectors are related by a symmetric matrix S:

$$S\vec{x} = \vec{e} \qquad \rightarrow \vec{e}^T \vec{e} = \vec{x}^T S^2 \vec{x}, \quad Q = S^2 \tag{B.24}$$

Applying the circulant approximation to both the model matrix S and the precision matrix Q, we find that

$$\sqrt{n}[F\vec{s}]_k = [\tilde{I}_n(k)]^{-1/2} \qquad (\vec{s} = \text{ base of } S) \tag{B.25}$$

which means that the maximum likelihood estimate of the model matrix base is obtained by fitting a cosine series to the *inverse square root* of the geometrically normalized sample periodogram.

C The conditional AR(2) propagator

Let's start with the general form of the conditional AR(2) transfer function, namely

$$\frac{X(z)}{E(z)} = \frac{1}{1 - r_1 z^{-1}} \frac{1}{1 - r_2 z^{-1}} \tag{C.1}$$

where $X(z)$ and $E(z)$ are the discrete z-transforms of the mean-reduced series and innovation series, respectively. The corresponding spectral density and autocovariance are

$$\begin{aligned}
\left|\frac{X(\omega)}{E(\omega)}\right|^2 &= \frac{1}{1 + {r_1}^2 - 2r_1 \cos\omega} \frac{1}{1 + {r_2}^2 - 2r_2 \cos\omega} \\
&= \frac{1}{(r_1 - r_2)(1 - r_1 r_2)} \left[\frac{r_1}{1 + {r_1}^2 - 2r_1 \cos\omega} - \frac{r_2}{1 + {r_2}^2 - 2r_2 \cos\omega}\right] \\
\frac{1}{\sigma^2}\gamma(s) &= \frac{1}{(r_1 - r_2)(1 - r_1 r_2)} \left[\frac{{r_1}^{s+1}}{1 - {r_1}^2} - \frac{{r_2}^{s+1}}{1 - {r_2}^2}\right] = D_1 {r_1}^s + D_2 {r_2}^s \quad [\text{integral } s]
\end{aligned} \tag{C.2}$$

where σ^2 is the innovation variance, and without loss of generality we will assume that $r_1 > r_2$. Let's also assume that the continuous propagator $g(v)$ is real-valued and causal in nature. Then

$$\begin{aligned}
\frac{1}{\sigma^2}\gamma(u) &= \frac{1}{(r_1 - r_2)(1 - r_1 r_2)} \left[\frac{{r_1}^{u+1}}{1 - {r_1}^2} - \frac{{r_2}^{u+1}}{1 - {r_2}^2}\right] = D_1 {r_1}^u + D_2 {r_2}^u \quad [\text{real } u] \\
\gamma(u) &= \int_0^\infty g(v) g(v+u) \mathrm{d}v \\
\frac{1}{\sigma} g(v) &= A_1 {r_1}^v + A_2 {r_2}^v \\
\rightarrow \frac{1}{\sigma^2}\gamma(u) &= \int_0^\infty (A_1 {r_1}^v + A_2 {r_2}^v)(A_1 {r_1}^{v+u} + A_2 {r_2}^{v+u}) \mathrm{d}v
\end{aligned} \tag{C.3}$$

and the corresponding equations for A_1 and A_2 are

$$\begin{aligned}
\frac{{A_1}^2}{2p_1} + \frac{A_1 A_2}{p_1 + p_2} &= D_1 = \frac{r_1}{(r_1 - r_2)(1 - r_1 r_2)(1 - {r_1}^2)} \\
\frac{A_1 A_2}{p_1 + p_2} + \frac{{A_2}^2}{2p_2} &= D_2 = \frac{-r_2}{(r_1 - r_2)(1 - r_1 r_2)(1 - {r_2}^2)} \\
p_1 = -\log r_1, \qquad & p_2 = -\log r_2, \qquad p_1 < p_2
\end{aligned} \tag{C.4}$$

Now set $A_1 = R\cos(\Psi/2)$, $A_2 = R\sin(\Psi/2)$, and $m = 2/R^2$. Then the equation for the $\{\Psi, m\}$ pair is

$$\begin{aligned}
\begin{bmatrix} 1/(2p_1) & 1/(p_1 + p_2) \\ -1/(2p_2) & 1/(p_1 + p_2) \end{bmatrix} \begin{bmatrix} \cos\Psi \\ \sin\Psi \end{bmatrix} &= m \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} - \begin{bmatrix} 1/(2p_1) \\ 1/(2p_2) \end{bmatrix} \\
\rightarrow \begin{bmatrix} \cos\Psi \\ \sin\Psi \end{bmatrix} &= m \begin{bmatrix} 2p_1 p_2 (D_1 - D_2)/(p_1 + p_2) \\ p_1 D_1 + p_2 D_2 \end{bmatrix} + \begin{bmatrix} (p_1 - p_2)/(p_1 + p_2) \\ -1 \end{bmatrix} \\
&= m\vec{a} + \vec{b}
\end{aligned} \tag{C.5}$$

where $\vec{a}$ and $\vec{b}$ depend on r_1 and r_2 only. This equation has a geometric interpretation, which I have illustrated in Figure 64.

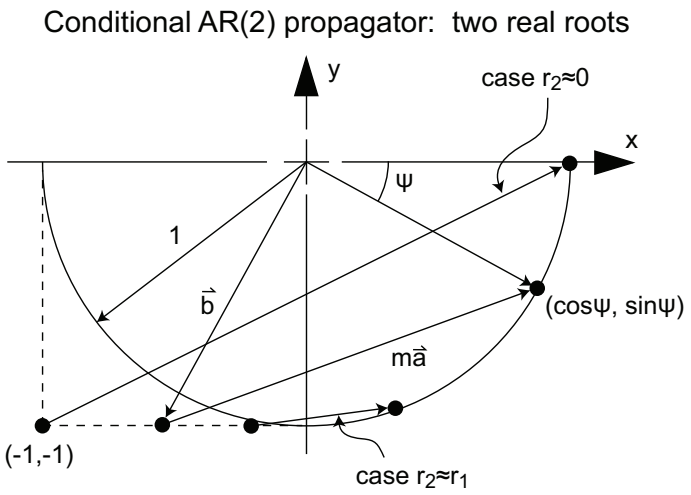

Figure 64: Geometric interpretation of the quantities involved in forming a conditional AR(2) propagator.

The algorithm for calculating the AR(2) propagator is then:

$$
\begin{array}{lll}
(1) & - & \text{compute } \vec{a}, \vec{b} \text{ from } r_1, r_2 \\
(2) & - & m_+ = \text{larger root of } |m\vec{a} + \vec{b}| = 1 \\
(3) & - & \text{solve } \left[\begin{array}{c} \cos\Psi \\ \sin\Psi \end{array} \right] = m_+\vec{a} + \vec{b} \text{ for } \Psi \\
(4) & - & g(v) = \sigma\sqrt{\dfrac{2}{m_+}}({r_1}^v \cos\dfrac{\Psi}{2} + {r_2}^v \sin\dfrac{\Psi}{2})
\end{array}
$$

Now let's consider the case in which the two autoregressive roots form a conjugate pair, i.e. $r_1 = re^{+j\theta}$, $r_2 = re^{-j\theta}$, and $0 \le \theta \le \pi$. Then the autocovariance is

$$
\frac{1}{\sigma^2}\gamma(u) = \frac{1}{1-r^2}\frac{r^u}{1-2r^2\cos 2\theta + r^4}\left[\frac{\sin(u+1)\theta - r^2\sin(u-1)\theta}{\sin\theta}\right] \tag{C.6}
$$

Now define the pair $\{R, \Psi_0\}$ so that

$$
\begin{aligned}
R\cos\Psi_0 &= (1+r^2)\sin\theta, \quad R\sin\Psi_0 = (1-r^2)\cos\theta \\
\rightarrow R^2 &= 1 - 2r^2\cos 2\theta + r^4 = (1 + 2r\cos\theta + r^2)(1 - 2r\cos\theta + r^2) \\
\rightarrow \frac{1}{\sigma^2}\gamma(u) &= \frac{1}{1-r^2}\frac{r^u}{R\sin\theta}\cos(u\theta - \Psi_0)
\end{aligned} \tag{C.7}
$$

Now, suppose that the continuous propagator $g(v)$ is real-valued and causal in nature:

$$
\begin{aligned}
\frac{1}{\sigma}g(v) &= r^v(A\cos v\theta + B\sin v\theta) \\
\rightarrow \frac{1}{\sigma^2}\gamma(u) &= \int_0^\infty r^v(A\cos v\theta + B\sin v\theta) \bullet \\
&\qquad r^{v+u}[A\cos(v+u)\theta + B\sin(v+u)\theta]\mathrm{d}v
\end{aligned} \tag{C.8}
$$

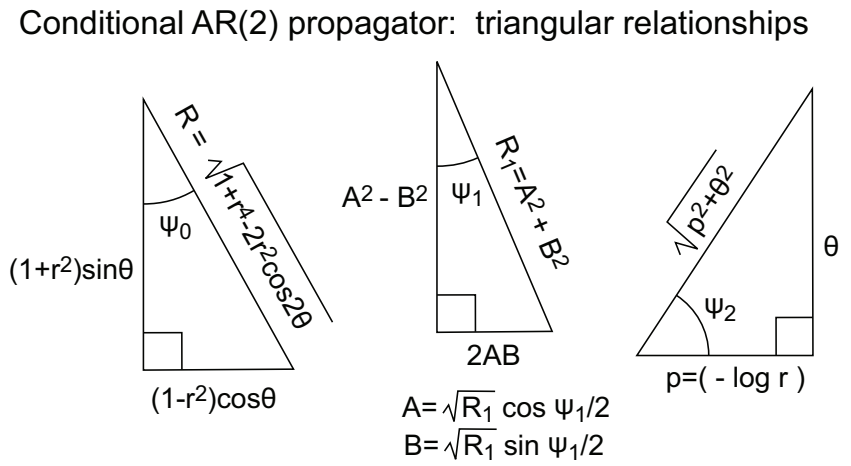

Figure 65: Definition and relationships of quantities entering into the calculation of the conditional AR(2) propagator for the case of conjugate roots.

We will transform the pair $\{A, B\}$ into the pair $\{R_1, \Psi_1\}$ and define an additional angle Ψ_2 as illustrated in Figure 65. With these definitions, we have

$$\frac{1}{1-r^2}\frac{1}{R\sin\theta}\begin{bmatrix}\cos\Psi_0\\ \sin\Psi_0\end{bmatrix} = \frac{R_1}{4p}\begin{bmatrix}1+\cos\Psi_2\cos(\Psi_1-\Psi_2)\\ \cos\Psi_2\sin(\Psi_1-\Psi_2)\end{bmatrix} \tag{C.9}$$

where $p = -\log r$. This also has a geometric interpretation, which I have illustrated in Figure 66.

From Figure 66, we can calculate that

$$\begin{aligned} 0 &= {R_0}^2 - 2R_0\cos\Psi_0 + \sin^2\Psi_2 \\ \rightarrow R_0 &= \cos\Psi_0 + \sqrt{\cos^2\Psi_0 - \sin^2\Psi_2} \end{aligned} \tag{C.10}$$

The algorithm for constructing $g(v)$ is then:

(1) – solve $R\cos\Psi_0 = (1+r^2)\sin\theta$ and $R\sin\Psi_0 = (1-r^2)\cos\theta$ for R and Ψ_0

(2) – solve $\tan\Psi_2 = \dfrac{\theta}{(-\log r)}$ for Ψ_2

(3) – solve $\tan\Psi_0 = \dfrac{\cos\Psi_2\sin(\Psi_1-\Psi_2)}{1+\cos\Psi_2\cos(\Psi_1-\Psi_2)}$ for Ψ_1

(4) – $R_0 = \cos\Psi_0 + \sqrt{\cos^2\Psi_0 - \sin^2\Psi_2}$

(5) – $R_1 = \dfrac{4(-\log r)}{R_0}\dfrac{1}{1-r^2}\dfrac{1}{R\sin\theta}$

(6) – $\dfrac{1}{\sigma}g(v) = \sqrt{R_1}r^v\cos(v\theta - \dfrac{\Psi_1}{2})$

Now that we have a continuous propagator, it's time to put it to use. In particular, we will use this function to interpolate into a conditional AR(2) random field. The version that interests me here is the last one, in which the autoregressive roots are complex conjugates. The propagator takes the form

$$g(v) = [\text{const}]r^v\cos(v\theta-\phi) = [\text{const}]e^{-pv}\cos(v\theta-\phi) \tag{C.11}$$

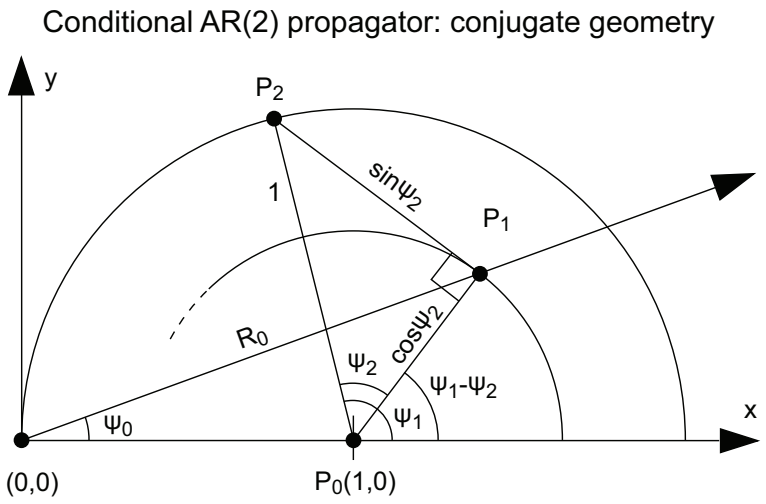

Figure 66: Geometric interpretation of the quantities involved in forming a conditional AR(2) propagator in the case of conjugate roots.

where $p = -\log r$, $\phi = \Psi_1/2$, the autoregressive roots are $r\exp(j\theta)$ and $r\exp(-j\theta)$, and Ψ_1 is defined as in the previous discussion. Let's say that we know $x(t)$ and $x(t+1)$, and wish to create a field value $x(t+\alpha)$ in between them which preserves the process statistics. Figure 67 illustrates the interpolation problem.

Suppose now that we wish to choose the linear combination of the envelopes $r^t\cos(t\theta - \phi)$ and $r^{t+1}\cos(t\theta + \theta - \phi)$ that best matches the envelope $r^{t+\alpha}\cos(t\theta + \alpha\theta - \phi)$ for $t < 0$. Then we would need to have the following equation identically satisfied:

$$g_0 r^t \cos(t\theta - \phi) + g_1 r^{t+1}\cos(t\theta + \theta - \phi) = r^{t+\alpha}\cos(t\theta + \alpha\theta - \phi), \quad t < 0 \tag{C.12}$$

Equating coefficients of $\cos(t\theta - \phi)$ and $\sin(t\theta - \phi)$ gives

$$\begin{aligned} g_0 + g_1 r\cos\theta &= r^\alpha \cos(\alpha\theta) \\ -g_1 r\sin\theta &= -r^\alpha \sin(\alpha\theta) \end{aligned} \tag{C.13}$$

and the solution is

$$\begin{bmatrix} g_0 \\ r g_1 \end{bmatrix} = \frac{r^\alpha}{\sin\theta}\begin{bmatrix} \sin(1-\alpha)\theta \\ \sin\alpha\theta \end{bmatrix} \tag{C.14}$$

Armed with these equations, we can now give a geometrical interpretation of the series variance and interpolation variance in terms of envelope functions, which I have illustrated in Figure 68.

An important case is that in which $p = -\log r$ and θ are both near zero, i.e. the random field is highly correlated. In this case we get

$$\begin{aligned} \frac{1}{\sigma}g(v) &= \frac{1}{\theta}r^{-v}\sin(-v\theta) \quad (v <= 0) \\ \frac{1}{\sigma^2}\gamma(u) &= \frac{1}{4p\theta}\frac{r^{|u|}}{p^2+\theta^2}(\theta\cos|u|\theta + p\sin|u|\theta) \end{aligned} \tag{C.15}$$

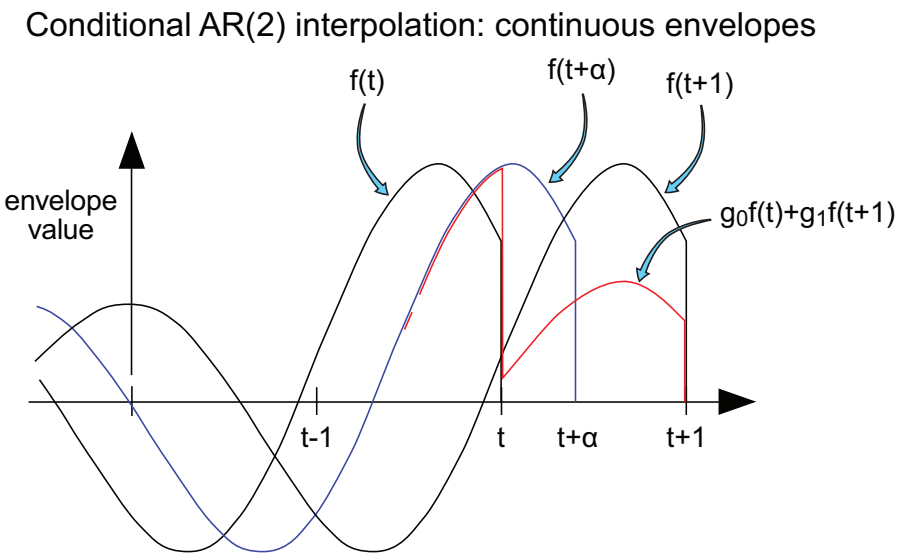

Figure 67: Setting up the interpolation problem for a conditional AR(2) random field.

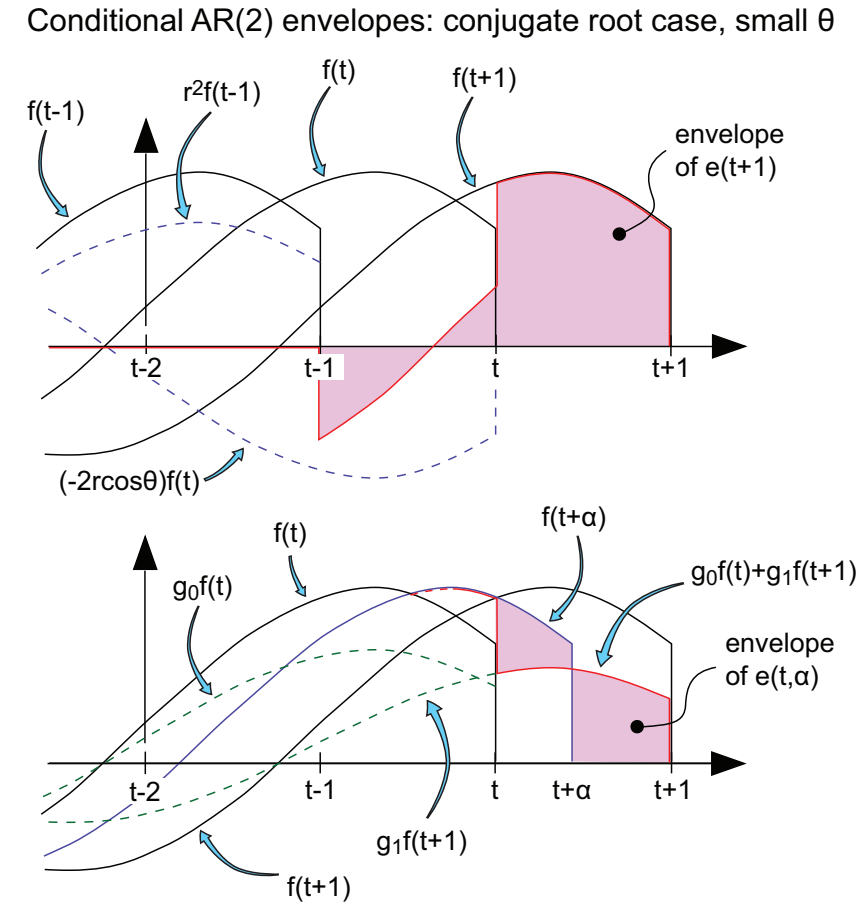

Figure 68: A sketch of the innovation variance for the conditional AR(2) interpolation problem when $\theta < \pi/2$.

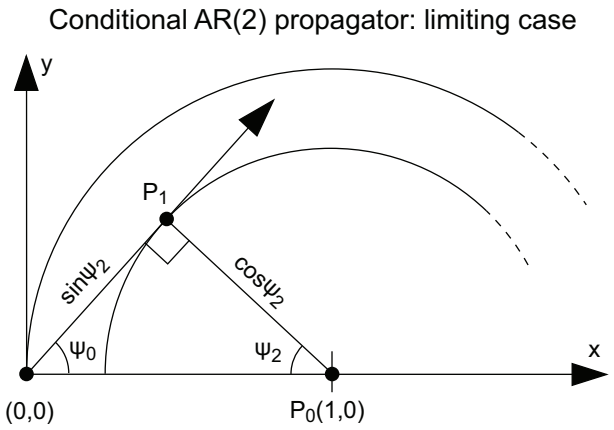

Figure 69: In the case of conjugate autoregressive roots, the conditional AR(2) propagator may not exist. Here I have sketched the limiting case.

and the second-order expansions for g_0 and g_1 are

$$\begin{aligned} g_0 &= (1-\alpha)\left[1 - \alpha p + \frac{1}{2}\alpha^2 p^2 + \frac{1}{6}\theta^2[1-(1-\alpha)^2]\right] \\ g_1 &= \alpha\left[1 + (1-\alpha)p + \frac{1}{2}(1-\alpha)^2 p^2 + \frac{1}{6}\theta^2(1-\alpha^2)\right] \end{aligned} \tag{C.16}$$

Before leaving the conditional AR(2) propagator, we must consider whether it always exists. In particular, when the roots form a conjugate pair, there is indeed a limiting case, which Figure 69 illustrates.

Here we have $\Psi_0 + \Psi_2 = \pi/2$, $\tan\Psi_0 \tan\Psi_2 = 1$, and consequently

$$\begin{aligned} &\frac{\theta}{p} = \frac{(1+r^2)\sin\theta}{(1-r^2)\cos\theta} = \frac{1+r^2}{1-r^2}\tan\theta \qquad (p = -\log r) \\ &\rightarrow \frac{1}{(-\log r)}\frac{1-r^2}{1+r^2} = \frac{1}{p}\tanh p = \frac{1}{\theta}\tan\theta \end{aligned} \tag{C.17}$$

For a solution to exist, we need

$$\Psi_0 < \frac{\pi}{2} - \Psi_2 \qquad \rightarrow \frac{1}{p}\tanh p < \frac{1}{\theta}\tan\theta \tag{C.18}$$

So we're okay for $\theta \leq \pi/2$, but larger values of θ are problematic. Suppose that we attempt to solve the interpolation problem in the manner described above, in other words by postulating an interpolant of the form

$$x(t+\alpha) = g_0 x(t) + g_1 x(t+1) + e(t,\alpha) \tag{C.19}$$

where $x(t)$ and $x(t+1)$ are successive reduced observations from the conditional AR(2) series under study, and $e(t,\alpha)$ is a quantity whose variance we would like to minimize. The solution obtained above was

$$\begin{bmatrix} g_0 \\ r g_1 \end{bmatrix} = \frac{r^\alpha}{\sin\theta}\begin{bmatrix} \sin(1-\alpha)\theta \\ \sin\alpha\theta \end{bmatrix} \tag{C.20}$$

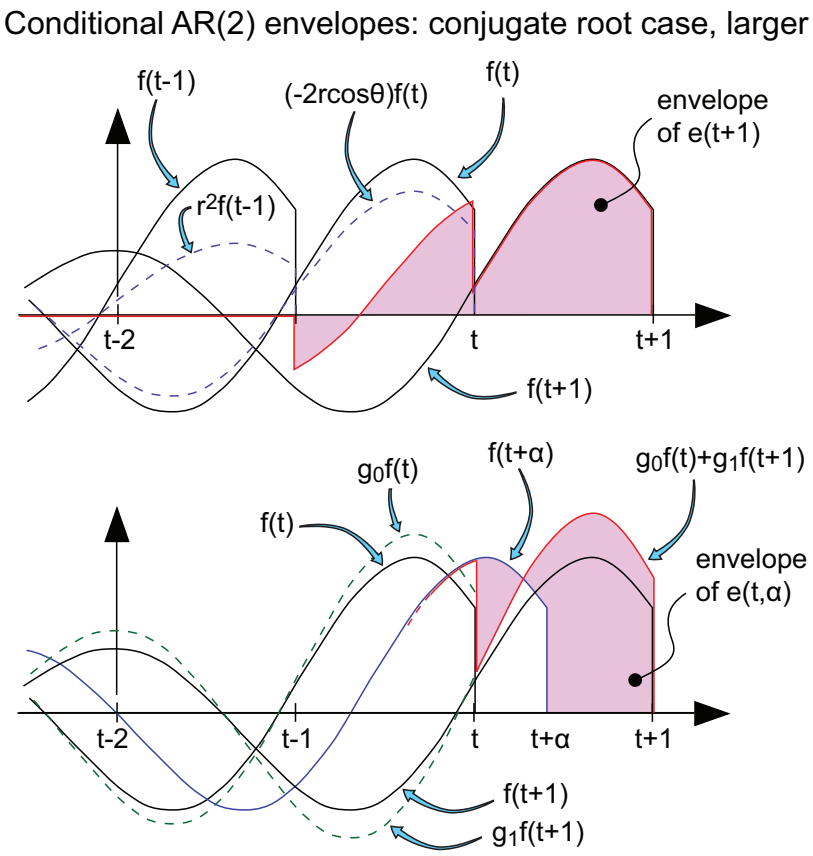

Figure 70: For the case of conjugate roots, the conditional AR(2) interpolant may have an innovation variance greater than that of the original series if $\theta > \pi/2$.

which is innocuous enough for small values of θ, as Figure 68 shows. However, the approximation definitely gets problematic for $\pi/2 < \theta < \pi$. Figure 70 is a chart showing one such case, where the variance of the interpolation appears to be *greater* than that of the series itself.

In fact, if we plot out the continuous autocovariance as specified by Equation 23.6, as I have done in Figure 71, we see that the function becomes non-physical when θ approaches π. The values at integral u are fine, but those in between are suspect. The bottom line is simple: we can use this method of interpolating the series with confidence only when $\theta < \pi/2$.

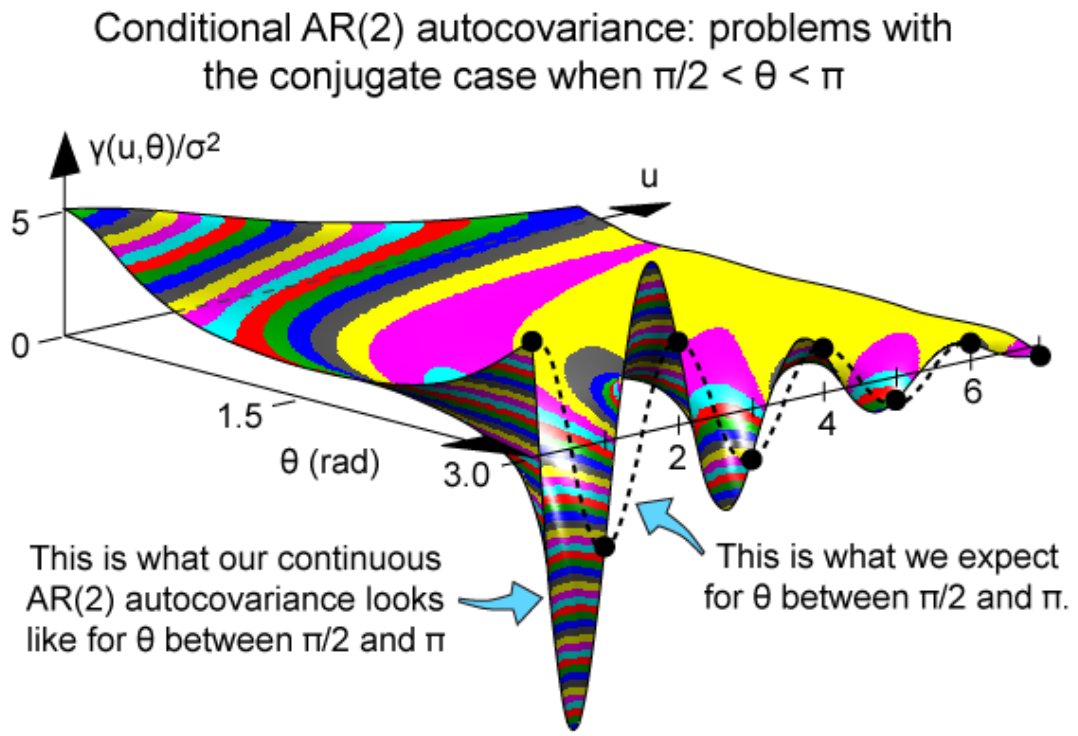

Figure 71: The continuous autocovariance that we assumed for a conditional AR(2) series in Equation 23.6 leads to non-physical result when θ is close to π.

D The skew-separable Fourier transform

The aim of this section is to demonstrate that the 2D Fourier transform retains its form under a shear transformation of the coordinate axes, a fact that I use in Section 16. First, let's define the quantities of interest. I imagine a linear transform that takes the usual (x, y) coordinates into coordinates (x_1, x_2), where the new axes are at angles θ_1 and θ_2 with respect to the original x axis. Figure 72 illustrates this state of affairs.

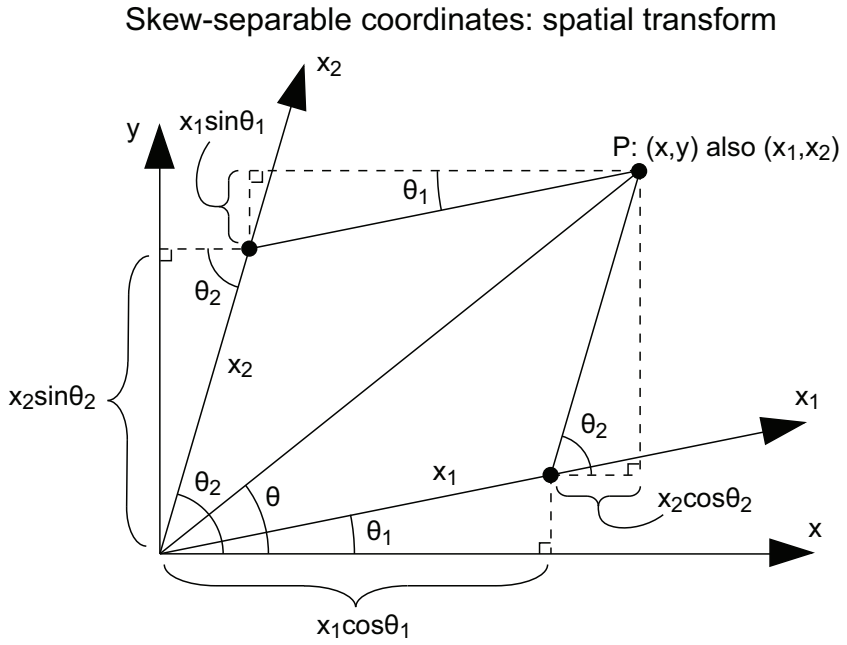

Figure 72: A geometric description of skewed coordinates in two dimensions.

The corresponding coordinate transformations are

$$\begin{aligned}
\begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} \cos\theta_1 & \cos\theta_2 \\ \sin\theta_1 & \sin\theta_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \frac{1}{\sin(\theta_2 - \theta_1)} \begin{bmatrix} \sin\theta_2 & -\cos\theta_2 \\ -\sin\theta_1 & \cos\theta_1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\
&= \frac{\sqrt{x^2+y^2}}{\sin(\theta_2 - \theta_1)} \begin{bmatrix} \sin(\theta_2 - \theta) \\ \sin(\theta - \theta_1) \end{bmatrix}
\end{aligned} \tag{D.1}$$

Suppose now that our target function is separable in the sheared coordinate system, i.e.

$$f(x, y) = p_1(x_1)p_2(x_2) \tag{D.2}$$

Then the standard 2D Fourier transform is:

$$F(\omega_x, \omega_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_1(x_1)p_2(x_2)\exp[-j(x\omega_x + y\omega_y)]\mathrm{d}x\mathrm{d}y \tag{D.3}$$

and if we set

$$\begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ \cos\theta_2 & \sin\theta_2 \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} \tag{D.4}$$

then we get

$$\begin{aligned} F(\omega_x, \omega_y) &= \sin(\theta_2 - \theta_1) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_1(x_1) p_2(x_2) \exp[-j(x_1\omega_1 + x_2\omega_2)] \mathrm{d}x_1 \mathrm{d}x_2 \\ &= \sin(\theta_2 - \theta_1) P_1(\omega_1) P_2(\omega_2) \end{aligned} \tag{D.5}$$

where $P_1(\omega_1)$ and $P_2(\omega_2)$ are the standard 1D Fourier transforms of $p_1(x_1)$ and $p_2(x_2)$ respectively. Taking the inverse Fourier transform reproduces our starting point, namely that $f(x, y) = p_1(x_1)p_2(x_2)$. The bottom line: a separable spectral density implies a separable autocovariance, even if the coordinate axes are skewed. Why is that possible? In short, because the shear transforms that we developed preserve both dot products and phase space volume elements. To see this more clearly, consider the basic 2D Fourier transform set:

$$\begin{aligned} f(\vec{x}) &= \int f(\vec{s}) \delta_2(\vec{x} - \vec{s}) \mathrm{d}^2\vec{s} \\ &= \frac{1}{4\pi^2} \int \int f(\vec{s}) \exp[j\vec{\omega}^T(\vec{x} - \vec{s})] \mathrm{d}^2\vec{\omega} \mathrm{d}^2\vec{s} \\ &= \frac{1}{4\pi^2} \int \left[\int f(\vec{s}) \exp(-j\vec{\omega}^T\vec{s}) \mathrm{d}^2\vec{s} \right] \exp(+j\vec{\omega}^T\vec{x}) \mathrm{d}^2\vec{\omega} \end{aligned} \tag{D.6}$$

Any transform that preserves dot products $\vec{\omega}^T\vec{x}$ and phase space elements $\mathrm{d}^2\vec{\omega}\mathrm{d}^2\vec{s}$ will leave the Fourier transform untouched. In particular, if we have $\bar{x} = L\vec{x}$, $\bar{\omega} = G\vec{\omega}$, and $G^T L = I_2$ where I_2 is the 2 by 2 identity matrix, then we have

$$f(\bar{x}) = \frac{1}{4\pi^2} \int \left[\int f(\bar{s}) \exp(-j\bar{\omega}^T\bar{s}) \mathrm{d}^2\bar{s} \right] \exp(+j\bar{\omega}^T\bar{x}) \mathrm{d}^2\bar{\omega} \tag{D.7}$$

In the case of the shear transform mentioned earlier, we had

$$G^T L = \begin{bmatrix} \cos\theta_1 & \cos\theta_2 \\ \sin\theta_1 & \sin\theta_2 \end{bmatrix} \frac{1}{\sin(\theta_2 - \theta_1)} \begin{bmatrix} \sin\theta_2 & -\cos\theta_2 \\ -\sin\theta_1 & \cos\theta_1 \end{bmatrix} = I_2 \tag{D.8}$$

and so the 2D Fourier transform is valid in the skewed coordinate system. One thing I should point out is that under these transformations, the (x_1, x_2) and (ω_1, ω_2) coordinate systems do not necessarily overlap each other. Indeed, the frequency transform is a shear followed by a contraction. Figure 73 summarizes this state of affairs.

$$\begin{aligned} \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} &= \begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ \cos\theta_2 & \sin\theta_2 \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} \\ \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} &= \frac{1}{\sin(\theta_2 - \theta_1)} \begin{bmatrix} \sin\theta_2 & -\sin\theta_1 \\ \cos\theta_2 & \cos\theta_1 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} \\ &= \frac{1}{\sin(\theta_2 - \theta_1)} \begin{bmatrix} \cos(\theta_2 - \frac{\pi}{2}) & \cos(\theta_1 + \frac{\pi}{2}) \\ \sin(\theta_2 - \frac{\pi}{2}) & \sin(\theta_1 + \frac{\pi}{2}) \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} \end{aligned} \tag{D.9}$$

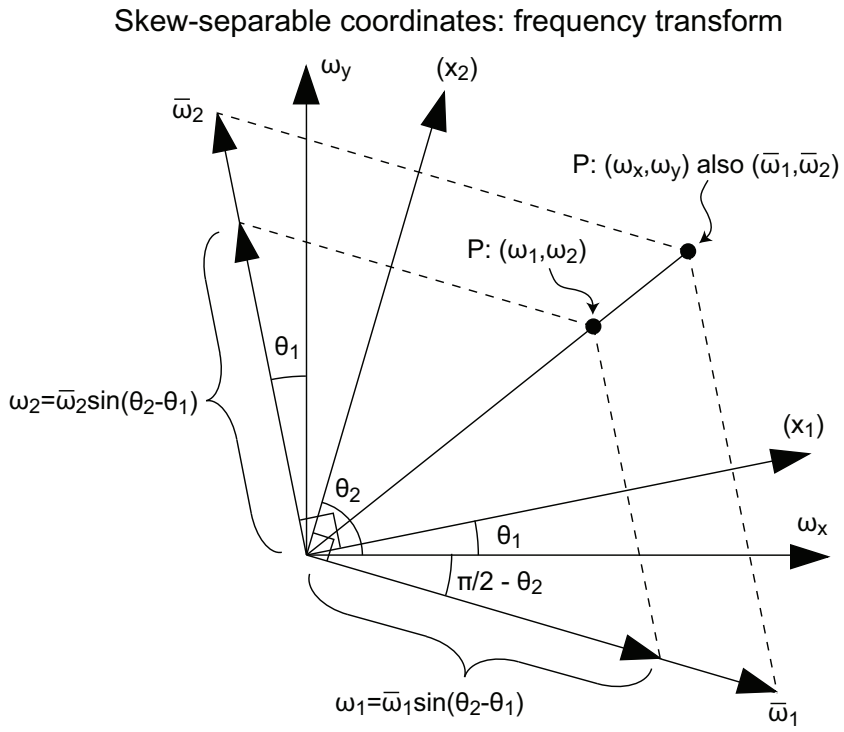

Figure 73: A geometric description of the skewed frequency coordinates that correspond to the skewed spatial coordinates described in Equation 24.1.

References

[1] Wold, H. *A Study in the Analysis of Stationary Time Series*. Uppsala, Sweden: Almquist & Wicksell, 1954. Quoted in Hipel & McLeod (reference 16), p. 82.

[2] B. Julesz, E. N. Gilbert, & J. D. Victor, "Visual Discrimination of Textures with Identical Third-Order Statistics," *Biological Cybernetics*, vol. 31, 1978, pp. 137-140

[3] This and several other histological images are available at http://anatomy.iupui.edu/courses/histo_D502

[4] Neill Campbell,Colin Dalton, David Gibson, David Oziem, and Barry Thomas, "Practical generation of video textures using the auto-regressive process," *Image and Vision Computing*, vol. 22, pp. 819-827.

[5] Franco Wolfe & Andrew Fitzgibbon, "Shift-invariant Dyamic Texture Recognition," *Computer Vision - ECCV 2006*, Part II, pp.549-562

[6] J. P. Lewis, "Generalized Stochastic Subdivision," *ACM Transactions on Graphics*, vol. 6, no. 3, July 1987, pp. 167-190

[7] H. Xie, H. Sun, Y. Ju, & Z. Feng, "Study on generation of rock fracture surfaces by using fractal interpolation," *International Journal of Solids and Structures*, Vol. 38, No. 32-33, August 2001, Pages 5765-5787

[8] H. Erzgraber, H. Touchette, D. K. Arrowsmith, & W. Just, "Time series analysis of the Norwegian electricity spot market prices," http://manmadenet.eu/publications/

[9] P. Whittle, "On Stationary Processes in the Plane," *Biometrika*, vol. 41, 1954, pp. 434-449

[10] Mihran Tuceryan & Anil K. Jain, "Texture Analysis," in *The Handbook of Pattern Recognition and Computer Vision*, 2nd edition, World Scientific Publishing Co, 1998, pp. 207-248

[11] J. M. Francos, A. Z. Meiri, & Boaz Porat, "A Unified texture Model Based on a 2D Wold-like Decomposition," *IEEE Transactions on Signal Processing*, vol. 41, no. 8, August 1993, pp. 2665-2678

[12] James A. Cadzow, D. M. Wilkes, & R. A. Peters , "Image Texture Synthesis-by-Analysis Using Moving-Average Models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 4, October 1993, pp. 1110-1122

[13] Rosalind W. Picard, "Structured Patterns from Random Fields," *Twenty-sixth Asilomar Conference on Signals, Systems, and Computers 1992*, Oct 1992, pp 1011-1015

[14] Maria Petrou & Pedro Garcia Sevilla, *Image Processing Dealing with Texture*, Chichester:John Wiley and Sons, 2006, p. 195

[15] I. M. Elfadel & Rosalind W. Picard, "Gibbs Random Fields, Co-occurrences, and Texture Modelling," *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, January 1994, pp. 24-37

[16] Keith W. Hipel & Ian A. McLeod, *Time Series Modelling of Water Resources and Environmental Systems*, Amsterdam:Elsevier, 1994 p. 390

[17] Jacek Ilow & Henry Leung, "Self-similar Texture Modelling Using FARIMA Processes with Applications to Satellite Images," *IEEE Transactions on Image Processing*, vol. 10, no. 5, May 2001, pp. 792-797

[18] Jianguo Zhang & Tieniu Tan, "Brief review of invariant texture analysis methods," *Pattern recognition*, vol. 35, 2002, pp. 735-747

[19] R. L. Kashyap & A. Khotanzad, "A Model-Based Method for Rotation Invariant Texture Classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 4, July 1986, pp. 472-481

[20] Kie B. Eom, "Generalized circular autoregressive models for isotropic and anisotropic Gaussian textures," *Journal of the Optical Society of America*, vol. 18, no. 8, August 2001, pp. 1822-1831

[21] Jianchang Mao & Anil K. Jain, "Texture Classification and Segmentation Using Multiresolution Simultaneous Autoregressive Models," *Pattern Recognition*, vol. 25, no. 2, 1992, pp. 173-188

[22] Franci Lahajnar & Stanislav Kovacic, "Rotation-invariant texture classification," *Pattern Recognition Letters*, vol. 24, 2003, pp. 1151-1161

[23] Huawa Deng & David Clausi, "Gaussian MRF rotation-invariant features for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, July 2004, pp. 951-955

[24] Hipel & McLeod (reference 16), p. 573

[25] F. S. Cohen, Zhigang Fan, & Maqbool A. Patel, "Classification of Rotated and Scaled Textured Images Using Gaussian Markov Random Field Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 2, February 1991, pp. 192-202

[26] Eric Lengyel, *Mathematics for 3D Game Programming and Computer Graphics*, 2nd edition, Hingham, Mass:Charles River Media, 2004, p. 80

[27] R. M. Corless, M. W. Giesbrecht, M. van Hoeij, I. S. Kotsireast, S. M. Watt, "Towards Factoring Bivariate Approximate Polynomials," *Proceedings of the 2001 International Symposium on Symbolic and Algebraic Computation*, 2001, pp. 85-92

[28] Havard Rue and Leonhard Held, *Gaussian Markov Random Fields: Theory and Applications*, Boca Raton, Florida: Chapman & Hall/CRC, 2005, p. 79

[29] Sridhar Lakshmanan & Haluk Derin, "Valid Parameter Space for 2D Gaussian Markov Random Fields," *IEEE Transactions on Information Theory*, vol 39, no. 2, March 1993, pp. 703-709

[30] Hipel & McLeod (reference 16), p. 218

[31] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, Mineola, NY:Dover, 1966. Available in electronic form at: http://sipi.usc.edu/database/database.cgi?volume=textures

[32] Rue and Held (reference 28), p. 71

[33] Rue and Held (reference 28), pp. 58-64

[34] Helmut Luhtkepohl, *Introduction to Multiple Time Series Analysis*, 2nd edition. New York:Springer-Verlag, 1993.

[35] A. V. Oppenheim & R. W. Schafer, *Digital Signal Processing*, Englewood Cliffs, NJ:Prentice-Hall, 1975, p. 541

[36] Oppenheim & Schafer (reference 35), p. 547

[37] Oppenheim & Schafer (reference 35), p. 553

[38] I. S. Gradshteyn & I. M. Ryzhik, *Table of Integrals, Series, and Products*, New York:Academic Press, 1980, p. 41

[39] M. Abramowitz & I. A. Stegun, *Handbook of Mathematical Functions*, New York:Dover, 1972, p. 260

[40] Historical data courtesy of MSN, available at: http://moneycentral.msn.com/investor/charts/chartdl.aspx

[41] Robert F. Engle (editor), *ARCH Selected Readings*, Oxford:Oxford University Press, 1995, p. 314

[42] Daniel Revuz & Marc Yor, *Continuous Martingales and Brownian Motion*, 3rd edition, New York:Springer-Verlag, 1999.

[43] The rotated Brodatz images are available at: http://sipi.usc.edu/database/database.cgi?volume=rotate

[44] Oppenheim & Schafer (reference 35), p. 539

[45] Oppenheim & Schafer (reference 35), p. 29

[46] Robert J. Adler, *The Geometry of Random Fields*, Toronto:John Wiley and Sons, 1981, p. 259

[47] Abramowitz & Stegun (reference 39), p. 375

[48] Y. Sakamoto, M. Ishiguro, & G. Kitigawa, *Akaike Information Criterion Statistics*, Hingham, Mass:Kluwer Academic Publishers, 1986, p. 188

[49] Abramowitz & Stegun (reference 39), p. 775