Linear Systems: Black Boxes and Beyond

These notes combine portions of "Linear Systems Theory" and "Fourier Analysis: Noise and Variability" notes from 2008-2009, and is modified from notes of the same title, 2010-2011. The main difference c/w the 2010-2011 notes is that point processes is moved into the second half, "A more comprehensive approach to noise and variability". The main difference c/w the 2012-2013 notes is a more extensive section on the factor of 1/2π.

The classic "black box" problem and its variants

The general setup is a "black box" with an input, $s(t)$, and an output, $r(t)$, where $r(t) = F[s](t)$.

F is variously called a system, a transformation, a transducer, or, a "functional" (hence, the bracket notation), to emphasize that imply that the value of F at a given time t depends not just on the particular current value $s(t)$, but on the values of s at all times (or all previous times). We have access to $s(t)$ and $r(t)$ but not the insides of the black box, and would like to characterize it.

For example,
apply a voltage, measure a current;
apply a head rotation, measure the eye movement;
apply a volume load to the circulatory system, measure the heart rate response
apply a light stimulus to a photoreceptor, measure the current
apply a pulse train to a synapse, measure the postsynaptic current

We typically assume that we can manipulate $s(t)$ (but sometimes we can only observe it).

We can have many goals: We may be interested simply in a concise way of describing F , or we may be interested in a way of describing F in a manner that suggests what is inside the black box. We may be interested in understanding how a network of such black boxes combines. We may have partial knowledge of the contents of F, and want to use it to infer the rest ("gray box")
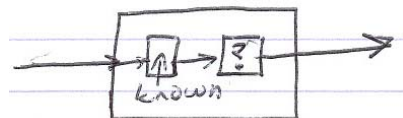
The input can be spatiotemporal rather than just spatial (apply an input to a neural network, measure the firing rate). We can try to reduce to the above case by assuming the inputs don't interact, i.e., $F[s_1,\ldots,s_N](t) = F_1[s_1](t) + \cdots + F_N[s_N](t)$

We may be interested in characterizing the variability of the output of F: in general, to determine $p(s,r)$. We can try to reduce to the above case by assuming that noise is additive, i.e., that $p(s,r) = p_{noise}(r(t) - F[s](t))$. This could be a very bad approximation if, say the output is a spike sequence – but then we might try to analyze in terms of a derived variable, e.g., a "firing rate".

Time-translation invariance

Of course the general class of problems considered above is intractable. So to get off the ground, we make two assumptions: *time-translation invariance* (hard to argue with) and *linearity* (a major simplification). For "systems", we also assume causality, i.e., that the output can only depend on the input from earlier times. Also, need to make "technical" assumptions: finite memory (the output can only depend on the input up to some maximum previous time), and bounded response.

Linearity is the major simplification. It is never exact (saturations, thresholds, can't have negative firing rates), but it is very often useful. Many aspects of the theory do extend to nonlinear transformations, but it is not nearly as concise a theory – it can't be -- and it is not covered here.

Translation-invariance can readily be restated in "functional" notation: If $F[s](t) = r(t)$ and $s' = D_T(s)$ (i.e., $s'(t) = s(t+T)$), then $F[s'](t) = r(t+T)$. Of course this is just the familiar $FD_T = D_T F$.

Linearity in "functional" notation: $F[s_1 + s_2](t) = F[s_1](t) + F[s_2](t)$ and $F[\alpha s](t) = \alpha(F[s](t))$ for any scalar α.

Consequences of linearity

Linearity is simply that F is in $Hom(V,V)$, where V is the vector space of functions of time (the real line). But since we want to work over the field $\mathbb{C}$, how do we interpret functions of time that have complex values? The answer is given by linearity. Any signal $s(t)$ can be decomposed into a real part and an imaginary part: $s(t) = s_{real}(t) + is_{imag}(t)$, where s_{real} and s_{imag} can be determined from s by $s_{real}(t) = \frac{1}{2}(s(t) + \overline{s}(t))$ and $s_{imag}(t) = \frac{1}{2i}(s(t) - \overline{s}(t))$ (Note that

s_{imag} is itself a *real-valued* signal, since it is equal to its complex conjugate). So, if we know how a physical ("real") linear system responds to "real" inputs, then there is only one choice for how it responds to complex-valued inputs, namely:

$$F[s](t) = F[s_{real} + is_{imag}](t) = F[s_{real}](t) + iF[s_{imag}](t).$$

Linearity also provides an intuitive way to characterize how a transducer operates on an arbitrary signal $s(t)$. First, break up time into tiny intervals:
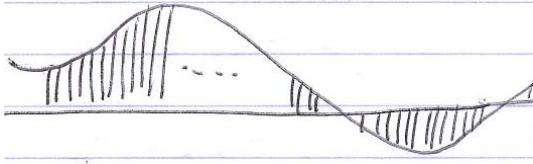
Linearity states that we can consider each interval independently, so we do this:
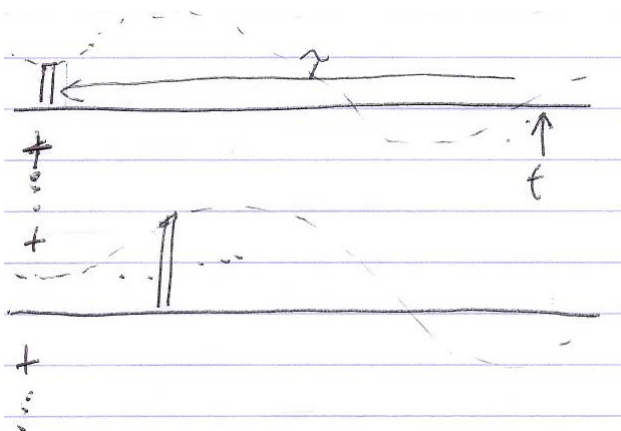
The response r to s at time t contains contributions from inputs at all times in the past, i.e., at all times $t - \tau$, for $\tau \geq 0$.

Let's say that $f(u)$ is the response of the system to a *unit impulse* at time 0. $f(u)$ is known as the "impulse response". Time-translation invariance means that $f(u)$ also determines the response to an impulse presented at time $t - \tau$: it is $f(u - (t - \tau)) = D_{-(t-\tau)}(u)$. That is, at time $t \geq \tau$, the response to an impulse presented at time $t - \tau$ is $f(t - (t - \tau)) = f(\tau)$. Linearity means that we also know the response to an impulse of size α; it is αf. So the total response at time t is a sum over all contributions from times τ earlier:

$$r(t) = F[s](t) = \int_0^\infty s(t - \tau) f(\tau) d\tau,$$ i.e., the response is the convolution of the stimulus with the impulse response. A possibly more intuitive view comes from changing variables to $u = t - \tau$:

$r(t) = F[s](t) = \int_{-\infty}^{t} s(u) f(t-u) du$ -- in both cases, it should be evident that the impulse response f weighs the influence of the stimulus at all previous times, and these combine by linearity.

Note also that the impulse response is, in fact, the response to a formal impulse: if we define $\delta(u)$ as the "function" for which $\int_{0}^{\infty} \delta(u) g(u) du = g(0)$, then

$$F[\delta](t) = \int_{-\infty}^{t} \delta(u) f(t-u) du = f(t).$$

We can always replace the limits of integration with $[-\infty, \infty]$ as long as we keep the requirement that impulse responses are always 0 when their arguments are ≤ 0 (required by causality).

Why not use the time-domain formulation?

It would seem natural to use the above result (the impulse response is the response to a delta-function) to measure it directly, and to use the time domain formulation. Why not?

In addition to the theoretical considerations discussed so far, there are two reasons.

First, measuring the impulse response directly pushes the system into a range in which linearity is likely to be least valid (it is never completely valid), as one needs to realize the delta-function by a short, high-intensity stimulus.

More precisely: if the impulse response (which you don't know yet) is something that does not change much over times on the order of ΔT, then one can approximate

$$f(t) \approx \frac{1}{\Delta T} \left(\int_{t-\Delta T/2}^{t+\Delta T/2} f(u) du \right)$$

This is equivalent to approximating a delta-function by a pulse of width ΔT and height $1/\Delta T$. So, to make the pulse valid as a delta-function, you need to keep the width ΔT small. Making ΔT small (but keeping the response size constant, so that you can measure it in the face of system and instrumentation noise) means increasing the height $1/\Delta T$, since, for sufficiently short times, the response is proportional to the area of the pulse. But increasing the height will eventually recruit nonlinear phenomena that may have no physiologic relevance.

Second, using the impulse response does not lend itself to seeing how systems combine.
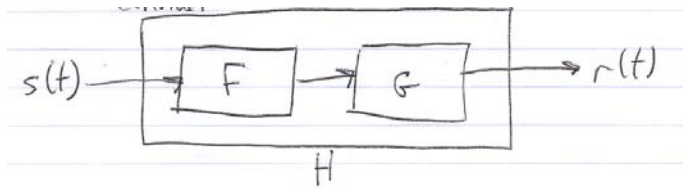
For example,
F=retinal processing, *G*=cortical processing;
F=visual processing, *G*=neural to BOLD transformation
F=the system you are studying, *G*=your measuring device

If you know the impulse responses of the components $f(\tau)$ and $g(\tau)$, then what is the impulse response of the composite system *H*? We can calculate based on impulse responses:

Say $F[s]=q$, $G[q]=r$.
$$r(t)=G[q](t)=\int q(t-\tau)g(\tau)d\tau \text{ and } q(t')=F[s](t')=\int s(t'-\tau')f(\tau')d\tau'.$$
Straightforward substitution,
$$q(t-\tau)=\int s(t-\tau-\tau')f(\tau')d\tau', \text{ so}$$
$$r(t)=\iint s(t-\tau-\tau')f(\tau')g(\tau)d\tau'd\tau. \text{ Change to a new variable for the total lag time,}$$
$$u=\tau+\tau':$$
$$r(t)=\iint s(t-u)f(u-\tau)g(\tau)dud\tau, \text{ and interchange order of integration:}$$
$$r(t)=\int s(t-u)\left(\int f(u-\tau)g(\tau)d\tau\right)du.$$

This writes the response *r* in terms of an impulse, $r(t)=\int s(t-u)h(u)du$, for
$h(u)=\int f(u-\tau)g(\tau)d\tau$. That is, the impulse-response of the composite serial system is the convolution of the component impulse responses, $h=f*g$. It's not even self-evident that $f*g=g*f$ (well, almost self-evident, change variables $\tau'=u-\tau$).

It would be much worse to try to work out the impulse response for a feedback system, as you'd wind up writing an integral equation for the response to a delta-function, *h*, i.e., the system output which, when filtered by *G*, added to the input and then filtered by *F* again, recovers the same output.
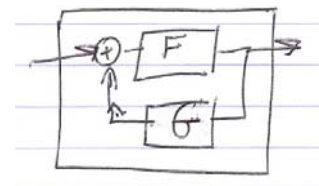
The frequency-domain representation

The group-theoretic formalism guarantees that we will get something much simpler: Translation is the regular representation; the regular representation decomposes into one-dimensional representations, which are parameterized by a frequency ω, in which time translation by an amount T are equivalent to multiplication by an amount $e^{i\omega T}$. Since F commutes with all time-translations, it must act only by multiplication by some number, which we will write as $\hat{f}(\omega)$.

Representing signals

How do we change bases (for signals)? I.e., given a representation of a signal as a function of time, $s(t)$, how do we project it into the subspace in which time-translation by an amount T acts like multiplication by an amount $e^{i\omega T}$? We know what this subspace is: it is the one-dimensional space of scalar multiples of $e^{i\omega t}$. And we know how to project an arbitrary vector v into the space spanned by some u, $P_u(v) = u\dfrac{\langle v,u \rangle}{\langle u,u \rangle}$. So we just apply this with $v = s$ and $u = e^{i\omega t}$.

The problem is that $\langle u,u \rangle$ and $\langle v,u \rangle$ are integrals over an infinite domain, and it is not clear whether they can be defined in a self-consistent way. But working formally, with the "convention" that u has unit length, i.e., that $\langle u,u \rangle = 1$ we find

$P_u(f) = e^{i\omega t} \displaystyle\int_{-\infty}^{\infty} s(t)e^{-i\omega t}dt$. The coefficient of $e^{i\omega t}$ is the coordinate of s in the new system, i.e., "Fourier analysis":

$$\hat{s}(\omega) = \int_{-\infty}^{\infty} s(t)e^{-i\omega t}dt. \tag{1}$$

Similarly, since the $e^{i\omega t}$ is a basis set, we can write $s(t)$ in terms of it, i.e., "Fourier synthesis":

$$s(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{s}(\omega)e^{+i\omega t}d\omega. \tag{2}$$

Note the "$+$" in the exponent – this comes from the algebraic structure, and the need for complex conjugation on the right side of the inner product. However, the $1/2\pi$, which might be a bit of a surprise. Eqs. (1) and (2) are known as Fourier transform pairs. We could make them completely symmetric (except for the sign in the exponent) by moving a factor of $\sqrt{2\pi}$, but then other things would look more complicated.

The infinite line and the factor of $1/2\pi$

The reason the factor of $1/2\pi$ becomes clear when we try to give rigorous interpretations to both $\langle u,u \rangle$ and $\langle v,u \rangle$. If we take $\langle u,u \rangle = 1$, then we are giving a finite (nonzero) value to an integral of a function of magnitude 1 over the entire domain –effectively, taking an average over the

interval $[-\infty, \infty]$. But if we apply this logic to a function v that only has a finite duration in which it is nonzero ("finite support"), we'd want to assign a value of 0 to $\langle v, u \rangle$. This would make the "synthesis" (eq. (2)) nonsensical, as each $\hat{s}(\omega)$ would be zero.. To avoid this kind of collapse, we realize that we have to interpret the coefficient $\hat{s}(\omega)$ the size of the projection, per unit frequency. That is, in a small range of frequencies $\Delta\omega$, the value of $\hat{s}(\omega)$ does not change much, and eq. (2) could then be approximated by a sum. Working out the (nontrivial) details of this lead to the factor of $1/2\pi$.

Had we worked with cyclic time, discretely sampled (in N points per cycle), nothing would be needed besides group theory – the representations of the cyclic group of an N-gon. The group is generated by the element a that corresponds to advancing by one time step ($\Delta T = T/N$). We already saw that the irreducible group representations are maps from a to multiplication by the complex number $e^{2\pi i k/N}$. Since $\dfrac{2\pi i k}{N} = \left(\dfrac{2\pi i}{N/T}\right)\dfrac{k}{T} = \left(\dfrac{2\pi i}{T}k\right)\Delta T$, we can make a correspondence between the infinite case (a representation whose multiplier for a time step of ΔT is $e^{i\omega\Delta T}$) and the finite case (a representation whose multiplier is $e^{2\pi i k/N}$): $i\omega\Delta t = 2\pi i k/N$ corresponds to frequency $\omega = 2\pi k/T$.

This leads to the discrete Fourier transform – "analysis"

$$\hat{s}_k = \frac{1}{N}\sum_{j=0}^{N-1}s(j\Delta T)e^{-2\pi i k j/N} = \frac{1}{N}\sum_{j=0}^{N-1}s(j\Delta T)e^{-2\pi i k(j\Delta T)/T} \ . \tag{3}$$

Since $\Delta T = T/N$ and $t_j = j\Delta T$, this is

$$\hat{s}_k = \frac{1}{T}\sum_{j=0}^{N-1}(\Delta T)s(j\Delta T)e^{-2\pi i k(j\Delta T)/T} = \frac{1}{T}\sum_{j=0}^{N-1}(\Delta T)s(t_j)e^{-2\pi i k t_j/T} \ . \tag{4}$$

In the limit that $N \to \infty$ and $\Delta T \to 0$ (and with $\omega = 2\pi k/T$), this can be interpreted as the integral

$$\hat{s}_k = \frac{1}{T}\int_{-\infty}^{\infty}s(t)e^{-i\omega t}dt \ . \tag{5}$$

The corresponding "synthesis" (eq. (2), with $\omega = 2\pi k/T$) is

$$s(t) = \sum_{k=0}^{N-1}\hat{s}_k e^{+2\pi i k t/T} \ , \tag{6}$$

where the "synthesis" is valid at any time t that is one of the times $t_j = j\Delta T = jT/N$ at which the signal $s(t)$ is sampled. That is,

$$s(t_j) = \sum_{k=0}^{N-1}\hat{s}_k e^{+2\pi i k t_j/T} = \sum_{k=0}^{N-1}\hat{s}_k e^{+2\pi i k(jT/N)/T} = \sum_{k=0}^{N-1}\hat{s}_k e^{+2\pi i k j/N} \ . \tag{7}$$

This also becomes an integral in the limit that $N \to \infty$:

$$s(t_j) = \int_0^N \hat{s}_k e^{+2\pi i k j / N} dk \ . \tag{8}$$

Since $\omega = 2\pi k / T$ we can change variables:

$$s(t_j) = \frac{T}{2\pi} \int_0^{2\pi N/T} \hat{s}_k e^{+i\omega j T/N} d\omega = \frac{T}{2\pi} \int_0^{2\pi N/T} \hat{s}_k e^{+i\omega t_j} d\omega \ , \tag{9}$$

which has the factor of $1/2\pi$ because of the change of variables. The standard Fourier analysis and synthesis equations ((1)and (2)) correspond to equations (5)and (9), with $\hat{s}(\omega) = T\hat{s}_k$ and $\omega = 2\pi k / T$.

Representing transformations: the "transfer function"

How do we represent transformations in the new basis? That is, in the subspace spanned by $e^{i\omega t}$, what is $\hat{f}(\omega)$, the scalar multiplier that corresponds to the action of F?

We use the impulse response to determine how F acts on $s = e^{i\omega t}$:

$$F[s](t) = \int_{-\infty}^{\infty} s(t-\tau) f(\tau) d\tau = \int_{-\infty}^{\infty} e^{i\omega(t-\tau)} f(\tau) d\tau = e^{i\omega t} \int_{-\infty}^{\infty} e^{-i\omega\tau} f(\tau) d\tau = s(t)\hat{f}(\omega) \ .$$ So (relief!) F

really does act like multiplication. The multiplier, which we defined as $\hat{f}(\omega)$, turns out to be

$\hat{f}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} f(\tau) d\tau$, which bears the same relationship to $f(t)$ as $\hat{s}(\omega)$ bears to $s(t)$, as in eq. (1).

    Even though f represents a transformation and s represents a signal (a vector), this is not an accident – because there is a natural mapping between transformations and signals: a transformation is associated with the signal that is its impulse response.

Since F acts separately in each subspace, $r = F[s]$ means $\hat{r}(\omega) = \hat{f}(\omega)\hat{s}(\omega)$.

The quantity $\hat{f}(\omega)$ is known as the "transfer function." It can be measured, directly and practically, by the response of a system to a sinusoidal input $\mathrm{Re}(e^{i\omega t})$. Note that this does not require going to large transient signals in order to have a readily measurable response. Breaking the transfer function down into magnitude and phase $\hat{f}(\omega) = \left|\hat{f}(\omega)\right| e^{i \arg \hat{f}(\omega)}$ is often more useful than breaking it down into a real and imaginary part. Its magnitude $\left|\hat{f}(\omega)\right|$ indicates the amplification of the sinusoid, and its phase $\arg(\hat{f}(\omega))$ indicates whether the sinusoid is advanced (positive phase) or delayed (negative phase) with respect to the input.

There are many other ways to measure the transfer function, including "noise" inputs, multi-sinusoid inputs, and noise inputs, each with their own advantages (and disadvantages).

It is often useful to think of the transfer function $\hat{f}(\omega)$ as a function of a complex variable ω, and where it has zeros and poles – this gives insight into the possible components of a system, and its stability. We'll get a glimpse of this below. There's another idea that we won't go into here in detail: the synthesis (or Fourier inversion) integral

$f(t) = \dfrac{1}{2\pi} \displaystyle\int_{-\infty}^{\infty} e^{i\omega t}\hat{f}(\omega)d\omega$ can be considered a contour integral, by "closing the circle" in the

upper half-plane or the lower half-plane. In order for this integral to be 0 for $t < 0$ (as required for causality), there can't be any zeros in the lower half-plane. The closer that complex zeros get to the lower half plane, the closer the system is to instability.

Transfer functions of composite systems

Transfer functions turn the composition of systems into algebra.

To compose two systems F and G in *series* to make H ($F[s] = q$, $G[q] = r$): Since $\hat{q}(\omega) = \hat{f}(\omega)\hat{s}(\omega)$ and $\hat{r}(\omega) = \hat{g}(\omega)\hat{q}(\omega)$, it follows that $\hat{r}(\omega) = \hat{f}(\omega)\hat{g}(\omega)\hat{s}(\omega)$. The transfer function of the composite system H is therefore $\hat{h}(\omega) = \hat{f}(\omega)\hat{g}(\omega)$. The frequency representation has turned serial composition (convolution) into multiplication. Now it's obvious that F followed by G is the same as G followed by F.
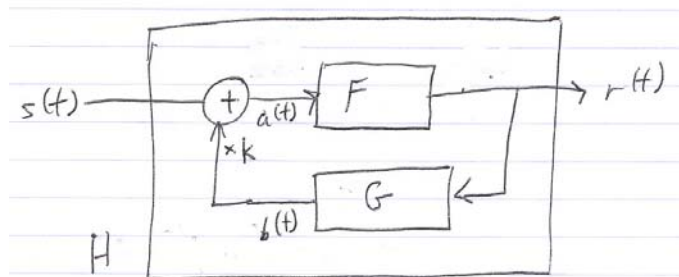
To compose two systems F and G in *parallel* to make H ($F[s] = q_1$, $G[s] = q_2$, $r = q_1 + q_2$): Since $\hat{q}_1(\omega) = \hat{f}(\omega)\hat{s}(\omega)$ and $\hat{q}_2(\omega) = \hat{g}(\omega)\hat{s}(\omega)$, it follows that $\hat{r}(\omega) = \hat{q}_1(\omega) + \hat{q}_2(\omega) = \hat{f}(\omega)\hat{s}(\omega) + \hat{g}(\omega)\hat{s}(\omega) = \left(\hat{f}(\omega) + \hat{g}(\omega)\right)\hat{s}(\omega)$. The transfer function of the composite system H is therefore $\hat{h}(\omega) = \hat{f}(\omega) + \hat{g}(\omega)$. The frequency representation has turned parallel composition into multiplication.

Let's try something more complicated: feedback.

We need to write $\hat{r}(\omega)$ in terms of $\hat{s}(\omega)$, i.e., $\hat{r}(\omega) = \hat{h}(\omega)\hat{s}(\omega)$.

What F does: $\hat{r}(\omega) = \hat{f}(\omega)\hat{a}(\omega)$. What G does: $\hat{b}(\omega) = \hat{g}(\omega)\hat{r}(\omega)$.

At the summation point: $\hat{a}(\omega) = \hat{s}(\omega) + k\hat{b}(\omega)$. Putting it all together:

$\hat{r}(\omega) = \hat{f}(\omega)\left(\hat{s}(\omega) + k\hat{b}(\omega)\right) = \hat{f}(\omega)\left(\hat{s}(\omega) + k\hat{g}(\omega)\hat{r}(\omega)\right)$. Solving for $\hat{r}(\omega)$:

$\hat{r}(\omega)\left(1 - k\hat{f}(\omega)\hat{g}(\omega)\right) = \hat{f}(\omega)\hat{s}(\omega)$, so

$\hat{r}(\omega) = \hat{h}(\omega)\hat{s}(\omega)$ for $\hat{h}(\omega) = \dfrac{\hat{f}(\omega)}{1 - k\hat{f}(\omega)\hat{g}(\omega)}$.

We can always do this with any kind of discrete network, and we will always recover an algebraic combination of the component transfer functions.

Electrical networks

A related, but not identical, situation is that of passive electrical networks. Specifically, apply current, measure voltage, in a network of resistors and capacitors – the ratio is the "impedance," $Z(\omega)$, which characterizes the network, and plays a similar role to that of the transfer function. Since everything is time-invariant, and (for ideal resistors and capacitors) linear, we can still analyze things one sinusoid at a time.

For resistors, by Ohms' Law $V = IR$, it follows that $\hat{V}(\omega) = \hat{I}(\omega)R$, so the impedance function is just the constant R. To find the transfer function for capacitors: $Q = CV$, where Q is charge. Current (I) is the time-derivative of Q. So $I = \dfrac{dQ}{dt} = C\dfrac{dV}{dt}$. Say $I(t) = e^{i\omega t}$ and

$V(t) = Z(\omega)e^{i\omega t}$. So $e^{i\omega t} = C\dfrac{d}{dt}Z(\omega)e^{i\omega t} = i\omega C Z(\omega)e^{i\omega t}$, from which it follows that $Z(\omega) = \dfrac{1}{i\omega C}$.

The distinction between composing electrical components, and composing input-output systems (as above) is that here, the "transfer function" relates voltage and current, not input and output. Current must be conserved in wires and at nodes, and voltages must be independent of path.

The net result is that impedances combine in a different way (familiar from resistor networks). For components in series, the combined impedance is $Z(\omega) = Z_1(\omega) + Z_2(\omega)$. For components in parallel, the combined impedance is given by $1/Z(\omega) = 1/Z_1(\omega) + 1/Z_2(\omega)$, or,

$Z(\omega) = \dfrac{Z_1(\omega)Z_2(\omega)}{Z_1(\omega) + Z_2(\omega)}$.

Combining resistors and capacitors in a discrete network will always yield sums and products of these two kinds of impedances, R and $1/i\omega C$, i.e., always some rational expression in ω.

We can also use this strategy to analyze continuum systems, e.g., the passive (1-D) cable and the passive (2-D) sheet.

Random walks and diffusion

We can also use the frequency-domain approach to analyze random walks. Here, the "signal" is a probability density, $p(x)$, the probability of finding a particle at x. The transformation is how this probability distribution evolves in time step. More precisely, if the probability distribution at time t is a delta-function $\delta(x)$, then let's say that the random walk process results in the particle having a distribution $F_{\Delta T}(x)$ at time ΔT.

$p(x)$, being a probability density, must have $\displaystyle\int_{-\infty}^{\infty} p(x)dx = 1$ and must be non-negative anywhere. But that is not essential to the algebraic setup, since (if linearity holds), then $F_{\Delta T}$ can be formally extended to signals that are not probability densities.

The reason that we are justified in using exactly the same formalism is that the analogs of time-translation-invariance and linearity hold. Time-translation invariance says that the particle's movements do not depend on its position x; that is, $F_{\Delta T}(x)$ gives the probability of a relative displacement x. Linearity says that a mixture of distributions behaves the same way as one would expect from the components: the particles don't interact. That is, $F_{\Delta T}[c_1 p_1 + c_2 p_2] = F_{\Delta T}[c_1 p_1] + F_{\Delta T}[c_2 p_2]$. Of course $c_1 + c_2 = 1$ in order for $c_1 p_1 + c_2 p_2$ to be a properly normalized probability density, but that is not an essential issue here.

We can immediately write down how the probability distribution evolves from one time to the next (and then see what happens in continuous time):

$$p_{t+\Delta T}(x) = \int_{-\infty}^{\infty} p_t(x-u)F_{\Delta T}(u)du.$$

Effectively, $F_{\Delta T}$ is an impulse response. So, in the (spatial) frequency domain,

$\hat{p}_{t+\Delta T}(\omega) = \hat{F}_{\Delta T}(\omega)\hat{p}_t(\omega)$.

Iterating:

$\hat{p}_{t+n\Delta T}(\omega) = \left(\hat{F}_{\Delta T}(\omega)\right)^n \hat{p}_t(\omega)$, or, $\hat{p}_{t+T}(\omega) = \left(\hat{F}_{\Delta T}(\omega)\right)^{T/\Delta T} \hat{p}_t(\omega)$. This tells how the probability distribution evolves over time.

Perhaps the simplest random walk is to allow the particle to move a step of size b to the right (increasing x) or left (decreasing x) with equal probability: $F_{\Delta T}(x) = \dfrac{1}{2}\big(\delta(x-b) + \delta(x+b)\big)$.

This yields

$$\hat{F}_{\Delta T}(\omega) = \int_{-\infty}^{\infty} F_{\Delta T}(x)e^{-i\omega x}dx = \frac{1}{2}\int_{-\infty}^{\infty}\big(\delta(x-b)+\delta(x+b)\big)e^{-i\omega x}dx = \frac{1}{2}\big(e^{-i\omega b}+e^{i\omega b}\big) = \cos(\omega b).$$

To take this to a continuous-time limit: we want $\Delta T \to 0$. Intuitively, the step size b should depend on ΔT, and also go to 0 as ΔT does—but how? We want $\big(\hat{F}_{\Delta T}(\omega)\big)^{T/\Delta T} = \big(\cos(b\omega)\big)^{T/\Delta T}$ to have a stable limit. For small b,

$\cos(b\omega) \approx 1 - \frac{1}{2}b^2\omega^2 \approx e^{-b^2\omega^2/2}$, and $\big(\cos(b\omega)\big)^{T/\Delta T} \approx e^{-b^2\omega^2 T/2\Delta T}$. So, if the limit of $\Delta T \to 0$ is

approached with $b^2/\Delta T = A$, then $\hat{p}_{t+T}(\omega) = e^{-\omega^2 AT/2}\hat{p}_t(\omega)$.

For $p_0(x) = \delta(x)$, (note that $\hat{\delta}(\omega) = 1$) we can find the distribution at any later time:

$\hat{p}_T(\omega) = \hat{F}_T(\omega) = e^{-\omega^2 AT/2}$. Finally, we can carry out Fourier synthesis:

$$p_T(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{p}_T(\omega)e^{i\omega x}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-\omega^2 AT/2}e^{i\omega x}d\omega = \frac{1}{\sqrt{2\pi AT}}e^{-x^2/2AT}.$$

(The final equality has many steps missing; the origin of the $\sqrt{2\pi}$ is not obvious. This is a standard definite integral, and there are many "tricks" to evaluate it.)

An important comment is that this continuum limit did not depend on the specific discrete-time model, only on the variance per unit time step. Any $\hat{F}_{\Delta T}(\omega)$ which is approximated by $1 - \frac{1}{2}b^2\omega^2$ (for small b) would lead to the same result. (We've almost proved the central limit theorem here.)

Note also that $\hat{F}_T(\omega) = e^{-\omega^2 AT/2}$, which is the "transfer function" of a continuous system, is not a rational expression in ω -- in contrast to all of the examples with discrete compositions of elements.

Note that this strategy works for diffusion in multiple spatial dimensions, rotational diffusion, tumbling (spherical diffusion), etc.

A more comprehensive approach to noise and variability

The goal is to extend the analysis to take into account intrinsic noise and variability.
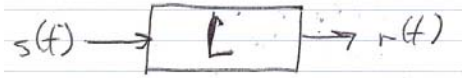
, but repeating the stimulus does not lead to the same response.

We note at the outset that the response to repeated presentations of the same stimulus may differ for two kinds of reasons: one is intrinsic variability, but the second is that the system may have "memory" (or learning). Formally, it is not possible to make the distinction (intrinsic variability vs. memory/learning) rigorously based on laboratory data– since when you present the stimulus a second time, you don't know if the difference in the response was because you had already presented it once, or, if there is intrinsic variability. But conceptually, it is a very important distinction, "memory" vs. "noise."

The formal distinction is made by thinking about an ensemble of parallel experiments, with identical preparations, and using the same stimulus in each – and asking whether the responses are the same. Since we can't do this in the lab, we simply just wait a long time; or, alternatively, ascribe all changes in the responses to "intrinsic variability", rather than noise.

"Noise" may also include measurement noise and noise that is actually generated by the system.

The tools will also deal with other (simpler) situations, such as characterizing the variability without regard to the stimulus


or to characterize the relationship between two response variables

Ensembles of time series

The basic new component of our setup is to consider "ensembles" of time series. Formally, an ensemble Ω of time series is simply an assignment of probabilities $p_\Omega(x)$ to every time series $x(t)$. Our goal is to characterize $p_\Omega(x)$ by sampling it.

An ensemble of time series

Characteris $p(x)$ $x(t)$ $D_\tau(x) = x(t+\tau)$ Ω
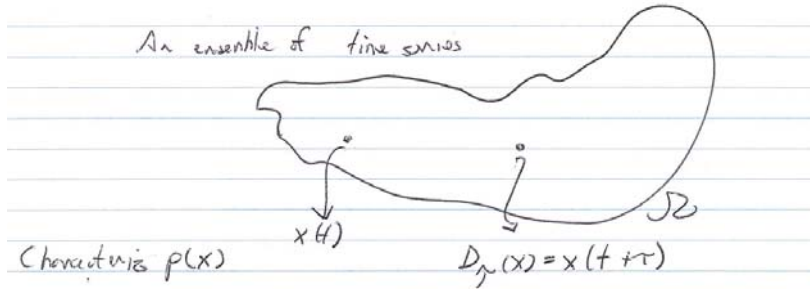
Since Ω represents all the "parallel universe" observations but we can only observe one laboratory at multiple times, we need to formalize the idea that sampling in this fashion (i.e., waiting and sampling again) is equivalent sampling from another parallel universe.

This has two parts. First, the ensemble is translationally invariant (its probabilities do not change with time, i.e., stationarity)

$p_\Omega(D_\tau[x]) = p_\Omega(x)$ for $D_\tau[x](t) = x(t+\tau)$.

Second, we assume that we can replace any kind of average over the ensemble Ω by averages over time:

$\langle \mathcal{F}(x, y,...) \rangle_\Omega = \langle \mathcal{F}(D_\tau[x], D_\tau[y],...) \rangle_\tau$, for any quantity $\mathcal{F}$ .that we wish to average. This condition is known as "ergodicity." Basically, it means that if we wait long enough, the system will exhibit all of its behaviors, with probability equal to the samplings over the parallel universes.

By combining these two assumptions, we can deduce that averages are independent of absolute time:

$$\langle \mathcal{F}(D_T[x], D_T[y],...) \rangle_\Omega = \langle \mathcal{F}(D_{T+\tau}[x], D_{T+\tau}[y],...) \rangle_\tau = \langle \mathcal{F}(D_\tau[x], D_\tau[y],...) \rangle_\tau = \langle \mathcal{F}(x, y,...) \rangle_\Omega,$$

where the middle equality follows because as τ traverses all times, so does $T+\tau$.

Since averages over the ensemble and averages over time are assumed to be identical, we will generally suppress the Ω- or τ-subscript.

We need to be sure that when we define $p(x)$ for a continuous time series x, that this makes sense. In the discrete case, if we had some specific number N of time samples, then $p(x)$ would be the joint distribution of $x(t_1),\dots x(t_N)$, specifically, $p(x)(\Delta x)^N$ is the probability that a sample y drawn from Ω has $x(t_1) \leq y(t_1) < x(t_1) + \Delta x$, etc. So the natural strategy is to interpret $p(x)$ in the same way for continuous time series, but we have to make sure that equations involving the above expressions the equations make sense when N is indefinitely large. (This is typically the case, as there will be the same number of factors Δx on both sides of an equation.)

To make practical use of the ergodicity (i.e., that we can replace ensemble sampling by sampling in time), we also assume that the signals have finite memory – i.e., if we wait long enough, samples are statistically independent. This is really not adding anything new, since we would be applying this to the outputs of systems that we assume have finite memory.

Characterizing an ensemble of time series

Since the ensemble is in effect a multivariate distribution, we can take a cue from simple statistics. The obvious first step is to characterize the mean, $\langle x(\tau) \rangle$. But (because of translation-invariance), this must be independent of τ, i.e., a constant. We could simply have set our measurement scale so that the mean value of the signal to be 0. The mean, therefore, is not that useful. From now on, we will assume that we have set our scales so that the mean signal value is zero.

The next obvious thing to do is to look at the variances (and covariances) of samples, $c_X(\tau_1, \tau_2) = \langle x(\tau_1) x(\tau_2) \rangle$. We can immediately simplify this because of translation-invariance: $c_X(\tau_1 + t, \tau_2 + t) = \langle x(\tau_1 + t) x(\tau_2 + t) \rangle = \langle x(\tau_1) x(\tau_2) \rangle = c_X(\tau_1, \tau_2)$, i.e., $c_X(\tau_1, \tau_2)$ depends only on the difference of its arguments, and we can write: $c_X(\tau_1, \tau_2) = c_X(\tau_1 - \tau_2)$.

Note also that $c_X(\tau_1, \tau_2) = c_X(\tau_2, \tau_1)$, so $c_X(\tau) = c_X(-\tau)$.

$c_X(0)$ is simply the signal variance; $c_X(\tau)$ is known as the autocovariance, and $c_X(\tau)/c_X(0)$ is known as the autocorrelation.

Why not use this straightforward approach to characterize X? The basic problem is that statistical estimates of $c_X(\tau)$ are not independent, and this is progressively more of a problem as we attempt to increase the time resolution (i.e,. decrease the step sizes for τ). For example: say $x(t)$ and $x(t + \Delta t)$ are correlated, then (by time translation invariance), so are $x(t + \Delta t)$ and $x(t + 2\Delta t)$. So now we expect that $x(t)$ and $x(t + 2\Delta t)$ are correlated – but how much? And to what extent is a direct measurement (i.e., a plug-in from the formula for $c_X(\tau)$) provide new information for $c_X(2\Delta \tau)$ that wasn't available from $c_X(\tau)$?

It also is not obvious how the autocorrelation will change when a signal passes through a transducer.

Working in the frequency domain

These problems are solved in the frequency domain, but there is an important difference compared with analyzing deterministic signals. The main thing is that the Fourier representation $\hat{x}(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt$ is not well-defined for infinite signals, which makes it a bit tricky to transform the autocovariance into the frequency domain.
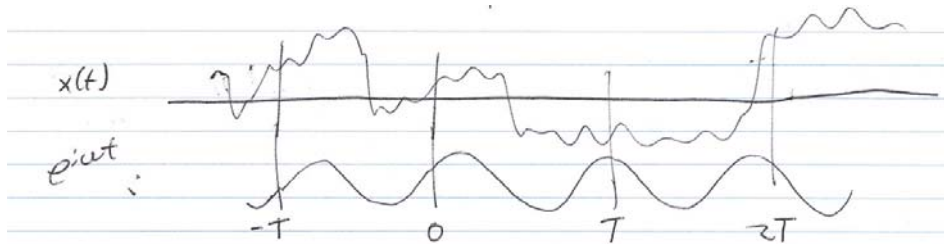
We see this when we break the integral up into periods of length T (not necessarily a repeat period of the frequency ω):

$$\hat{x}(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}dt = \sum_{N=-\infty}^{\infty} \int_{NT}^{N(T+1)} x(t)e^{-i\omega t}dt = \sum_{N=-\infty}^{\infty} F(x,\omega,T,NT) . \qquad (10)$$

Thus, the integral for $\hat{x}(\omega)$ diverges: it is an infinite sum of quantities, each of which is non-negligible.

Here, $F(x,\omega,T,T_0) = \int_{T_0}^{T_0+T} x(t)e^{-i\omega t}dt$ is a "Fourier estimate" of the signal x at the frequency ω,

over an interval T, and beginning at a start time T_0. (We give this quantity a new name, because it is not quite the Fourier transform, being only over a finite interval.)

The distribution of Fourier estimates

We need to understand the distribution of Fourier estimates to make sense of the above infinite integrals.

The distribution of Fourier estimates must be independent of the start time, T_0 – because any average over the ensemble is independent of start time.

Its average, at any time, must be zero. To see this, first use translation-invariance: $\langle F(x,\omega,T,0)\rangle = \langle F(x,\omega,T,T_0)\rangle$ But also, we can change variables:

$$F(x,\omega,T,T_0) = \int_{T_0}^{T_0+T} x(t)e^{-i\omega t}dt = \int_0^T x(u+T_0)e^{-i\omega(u+T_0)}dt$$

$$= e^{-i\omega T_0} \int_0^T x(u+T_0)e^{-i\omega u}dt = e^{-i\omega T_0} \int_0^T x(u)e^{-i\omega u}dt = e^{-i\omega T_0}F(x,\omega,T,0) \qquad (11)$$

Combining this with translation-invariance of the average $\langle F(x,\omega,T,0)\rangle = \langle F(x,\omega,T,T_0)\rangle$

yields $\langle F(x,\omega,T,0)\rangle = \langle F(x,\omega,T,T_0)\rangle = e^{-i\omega T_0}\langle F(x,\omega,T,0)\rangle$.

So this average must be zero, unless $\omega = 0$. At $\omega = 0$, it is just the signal average over the interval from T_0 to T_0+T, and is therefore zero as well.

Since we know that the Fourier estimates have mean zero, what about its second-order statistics? Looking at eq. (10), we will have to think about two cases: the same interval, and non-overlapping intervals. For non-overlapping intervals, the finite-memory condition guarantees that for sufficiently long interval lengths T, the distribution of Fourier estimates is independent, since only a small fraction of the end of the last interval and the beginning of the next interval are close. So we need to focus on the same-interval case, i.e., the behavior of the average of the square of $F(x,\omega,T,0)$.

A simple scaling argument tells us what we need to know: that $\left\langle \left| F(x,\omega,T,0) \right|^2 \right\rangle$ grows in proportion to T. Write $I(T) = \left\langle \left| F(x,\omega,T,T_0) \right|^2 \right\rangle$ (which is independent of T_0). We now consider intervals that are K times longer, and break up this interval of length KT into K intervals of length T.

$$I(KT) = \left\langle \left| F(x,\omega,KT,0) \right|^2 \right\rangle = \left\langle \left| \int_0^{KT} x(t)e^{-i\omega t}dt \right|^2 \right\rangle = \left\langle \left| \sum_{m=0}^{K-1} \int_{mT}^{(m+1)T} x(t)e^{-i\omega t}dt \right|^2 \right\rangle$$

$$= \left\langle \left(\sum_{m=0}^{K-1} \int_{mT}^{(m+1)T} x(t)e^{-i\omega t}dt \right) \left(\sum_{n=0}^{K-1} \overline{\int_{nT}^{(n+1)T} x(t)e^{-i\omega t}dt} \right) \right\rangle = \left\langle \sum_{m,n=0}^{K-1} \int_{mT}^{(m+1)T} x(t)e^{-i\omega t}dt \; \overline{\int_{nT}^{(n+1)T} x(t)e^{-i\omega t}dt} \right\rangle .$$

(Note the complex conjugation in the final term).

Because of the finite-memory condition, each of the cross-terms ($m \neq n$) drop out: They are approximately independent, so that the average of the product is the product of the averages, which we know is zero (from above).

The terms that do not drop out correspond to $m = n$:

$$I(KT) = \left\langle \sum_{m=0}^{K-1} \left| \int_{mT}^{(m+1)T} x(t)e^{-i\omega t}dt \right|^2 \right\rangle = \left\langle \sum_{m=0}^{K-1} \left| F(x,\omega,T,mT) \right|^2 \right\rangle = K \cdot I(T) .$$

Thus, the variance of a Fourier estimate is proportional to the duration of the estimate. Put another way, as time increases, the Fourier estimate is expected to grow indefinitely, but to grow in proportion to $\sqrt{T}$. We can think of the change from $F(x,\omega,(K-1)T,0)$ to $F(x,\omega,KT,0)$ this as taking a new sample out of the distribution of $F(x,\omega,T,0)$ as each amount T of time is added.

Note that the above argument also shows that the distribution of $F(x,\omega,T,0)$ (over the ensemble Ω) is independent of phase (i.e., this distribution is radially symmetric in the complex plane). Eq. (11) shows that rotating this distribution by an amount $e^{-i\omega T_0}$ is the same as translating time by an amount T_0.

The growth rate of $F(x,\omega,T,0)$ will depend on ω. If estimates tend to be uncorrelated, then growth will be slower, but if they are correlated, then the random samples of $F(x,\omega,T,0)$ will tend to line up, and growth will be faster.

The power spectrum

The power spectrum is defined as this growth rate:

$$P_X(\omega) = \lim_{T\to\infty} \frac{1}{T}\left|I_T(\omega)\right|^2 = \lim_{T\to\infty}\frac{1}{T}\left\langle \left|F(x,\omega,T,0)\right|^2\right\rangle = \lim_{T\to\infty}\frac{1}{T}\left\langle \left|\int_0^T x(t)e^{-i\omega t}dt\right|^2\right\rangle. \quad (12)$$

Note the units: P_X has units of time*(units of $X)^2$, or, (units of $X)^2$/(frequency)

The power spectrum thus completely describes the covariances of Fourier estimates at different times.

It also completely describes the covariances of Fourier estimates at different frequencies – but the reason is somewhat surprising. Consider the distribution of a quantity such as $F(x,\omega_1,T,T_0)F(x,\omega_2,T,T_0)$. The mean of this quantity is the covariance of two Fourier estimates, but it turns out that this is zero unless $\omega_1 + \omega_2 = 0$. The reason for this follows along the lines of eq. (11): if we work out the consequence of time-translation, it has to multiply the mean value of this quantity by $e^{-i\omega_1 T - i\omega_2 T}$. And if $\omega_1 + \omega_2 = 0$, i.e., $\omega_2 = -\omega_1$, then $F(x,\omega_2,T,T_0) = \overline{F(x,\omega_1,T,T_0)}$ and $F(x,\omega_1,T,T_0)F(x,\omega_2,T,T_0) = \left|F(x,\omega_1,T,T_0)\right|^2$, which is captured by the power spectrum. So there is nothing to be learned by looking at covariances across frequencies.

Put another way, in the frequency domain, the matrix of variances and covariances is *diagonal*, and the diagonal elements are the power spectrum.

We won't show it here, but the autocorrelation is the Fourier transform of the power spectrum:

$$P_X(\omega) = \int_{-\infty}^{\infty} c_X(\tau)e^{-i\omega\tau}d\tau.$$ See page 12 of 2003 notes on Power Spectra (PSPC01-PSPC14.pdf) if interested.

A "white" noise is a noise whose power spectrum is constant. Equivalently, it is a noise whose autocorrelation is except at $\tau = 0$, i.e., $c(\tau) = P\delta(\tau)$.

Higher moments and cross-moments

A bit of a digression: The power spectrum has natural extensions, based on considering higher moments of the Fourier estimates. For a triple product such as $F(x,\omega_1,T,T_0)F(x,\omega_2,T,T_0)F(x,\omega_3,T,T_0)$, time-translation multiplies it by $e^{-i\omega_1 T - i\omega_2 T - i\omega_3 T}$, which means that the triple product may have a nonzero average for any triplet of frequencies for which $\omega_1 + \omega_2 + \omega_3 = 0$. This is known as the bispectrum:

$$B_X(\omega_1,\omega_2) = \lim_{T\to\infty}\frac{1}{T^{3/2}}\left\langle F(x,\omega_1,T,0)F(x,\omega_2,T,0)\overline{F(x,\omega_1+\omega_2,T,0)}\right\rangle, \qquad (13)$$

which can also be written symmetrically as a function of three arguments for which

$$\omega_1 + \omega_2 + \omega_3 = 0, \; B_X(\omega_1,\omega_2,\omega_3) = \lim_{T\to\infty}\frac{1}{T^{3/2}}\left\langle F(x,\omega_1,T,0)F(x,\omega_2,T,0)F(x,\omega_3,T,0)\right\rangle.$$

Similarly trispectra, etc. can be defined. They inherit the nice properties of the spectrum, concerning how transducers act; and, estimation methods for the spectrum also extend in a straightforward manner.

Estimation of the spectrum is not trivial: the multitaper method

Very much not a digression -- thinking about the definition (12) shows that there is a fundamental problem: given some finite data length *L*, how to divide it up so that we can measure values of the Fourier estimate $F(x,\omega,T,T_0)$ that approach their long-*T* value, yet measure enough of them so that we can get stable estimates of their variance. That is, we could divide the data of length *L* into *M* segments of length $T = L/M$, and obtain *M* samples of Fourier estimates of length *T*, each starting at successive times, $T_0 = mT \; (m = 0,...,M-1)$.

This is essentially a tradeoff between accuracy (freedom from bias), which is enhanced by longer *T*; vs. precision (stability of the estimate) which is enhanced by shorter *T*, enabling more estimates from the same dataset. But how to choose *M*? And, is this the best thing to do?

There is now a standard, but somewhat counterintuitive, approach to this problem – "multitaper spectral estimation" – see Noise and Variability notes NAV16-NAV25 from 2008-2009 for more details; the primary paper is Thomson, D.J. (1982) Spectrum estimation and harmonic analysis. Proc IEEE. 70: p. 1055-1096.

The bottom line is that it still doesn't tell you how to "choose *M*," since you have to decide which you care about – accuracy or precision – but it does tell you the best way to make the estimate – and, surprisingly, it is not simply chopping up the data.

The overview is to work out how the length of the time window biases the Fourier estimate. $F(x,\omega,T,T_0)$ is essentially the Fourier transform (from $-\infty$ to ∞) of the signal *x* multiplied by a function that is 1 on the interval $[T_0,T_0+T]$, and zero elsewhere. This can be regarded as a

general "window" function (a.k.a. "taper"), $W(t)$. We then ask for the "best" collection of M window functions to use, rather than assume that these should just be pedestal functions on the M intervals $[mT,(m+1)T]$. Of course "best" has to be given a specific meaning – and we mean the following: that the estimated spectrum $P_X(\omega)$ is the one that corresponds to a Gaussian ensemble, whose spectrum is constant over frequency intervals $\dfrac{2\pi M}{L}$, that is most likely to have yielded the observed signal x. That is, you are effectively estimating a piecewise-constant spectrum, and the pieces are of length $\dfrac{2\pi M}{L}$.

Note that as M increases, you lower the resolution of the spectrum – but increase its precision (i.e., you decrease the error bars of what you measure, but the piece-wise estimate is now on progressively larger pieces.)

These taper functions turn out to be (a) nonzero over the entire interval from 0 to $L = MT$, (b) orthogonal, and (c) surprisingly, sometimes negative. They are alternatively known as "Slepian functions", "prolate spheriodal functions", and "discrete prolate spheroidal sequences", the latter to emphasize that one is working on a discrete-time mesh. Matlab's "dpss" produces these functions.

The net result is that eq. (12) is replaced by a "multitaper" estimate:

$$P_X(\omega) = \frac{1}{M}\sum_{m=1}^{M}\left|\int_0^L W_m(t,L)x(t)e^{-i\omega t}dt\right|^2,$$

and one should try a few values of M to be sure that findings don't depend on a particular choice of the accuracy-vs-precision tradeoff. Matlab's "pmtm" carries out the above estimate.

This procedure has two other advantages. First, error bars are readily obtained, since each of the M estimates are approximately independent – even from the same data. This is because the M taper functions are orthogonal. Second, spectral estimates at different frequencies are approximately independent too. This ultimately is a result of the orthogonality of the sinusoids, and the maximum-likelihood framework for the estimation. Here, the fine print in "approximately" is that the spectra are constant in the estimation window $\dfrac{2\pi M}{L}$.

Typically, M is chosen to be a relatively low odd number, e.g., 3 to 15; an odd number is chosen so that the center of the frequency window is always a frequeny of the form $\dfrac{2\pi k}{L}$, and the edges are frequencies $\dfrac{2\pi}{L}(k\pm\dfrac{M-1}{2})$.

The real and imaginary part of each integral in the multitaper estimate is approximately Gaussian-distributed (central limit theorem!), and, since the integrals are approximately

independent, the power spectrum, which is the sum of their magnitudes-squared (M integrals, each with a real and imaginary part) is distributed like χ^2 with $2M$ degrees of freedom.

An analogous procedure can (and should) be followed for other spectral quantities.

A full-featured package is available at www.chronux.org; it calculates error bars and also includes extensions to multiple signals, point processes, higher-order spectra, and related signal-processing steps.

Using the power spectrum

The power spectrum solves the problem of describing the way that the noise structure of signal is changed by a linear transformation, as well as facilitating other ways of analyzing noise or using it to understand a system.

Effect of a filter on noise
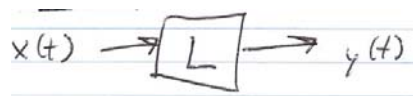
Since $y = L[x]$, then a Fourier estimate of y, for sufficiently long times, is related to Fourier estimate of x by $F(y,\omega,T,0) = \hat{L}(\omega)F(x,\omega,T,0)$, from which it follows that

$$P_Y(\omega) = \left|\hat{F}(\omega)\right|^2 P_X(\omega).$$

Power spectra of simple processes

In the above scenario, the input x need not be a continuous signal, it also might be a sequence of events (i.e., a point process). For example, x may represent the fusion of vesicles at spikes at a synapse, and L might represent the postsynaptic response to each fusion event (a "miniature end-plate potential"). Or x might represent a sequence of action potentials, and L might represent the overall synaptic response. So, calculating the power spectrum of a point process can be a crucial ingredient for modeling the variability of a continuous process.

So we take a detour to discuss point processes, and then their power spectrum. Intiutively, a point process is simply an event sequence; more formally, it is an ensemble from which example even sequences are drawn. We focus on a specific kind of point process – a "renewal process", because it is a nice compromise between simple and flexible (it includes Poisson processes and also clock-line ones), and we can do many things analytically.

A renewal process, by definition, is a point process in which the probability of the next event depends only on the probability since the previous event. This probability is the "renewal density", $p(t)$. That is, given a spike at time 0, the probability that the next spike occurs

between t and $t + \Delta t$ is $p(t)\Delta t$. $p(t)$ is a probability density and therefore is unit-normalized, $\int_0^\infty p(t)dt = 1$. We are still assuming time-translation invariance, so $p(t)\Delta t$ also can be interpreted as the probability that, given a spike at any time T, that the next spike is between $T + t + \Delta t$ and $T + t + \Delta t$.

The renewal density determines the charactistics of the process. For example, if $p(t)$ has a well-defined peak, the process is clock-like; if $p(t) = \lambda e^{-\lambda t}$, the process is Poisson. That may not be obvious: the reason for that behavior is that in a Poisson process, by definition, there is a fixed hazard of a spike per unit time, λ -- so in a Poisson process, $dp/dt = -\lambda p(t)$ -- which has $p(t) = \lambda e^{-\lambda t}$ as the unique solution, given unit normalization.

If $p_1(t) = p(t)$ is the distribution of the time of the first spike time (after a spike at time 0), we can calculate the probability distribution of the *second* spike (after a spike at time): $p_2(t)$:

$$p_2(t) = \int_0^t p(\tau)p(t-\tau)dt \text{, where } \tau \text{ is the time of the intervening spike.}$$

So we're essentially doing forward diffusion in discrete steps, in which each "step" corresponds to the transformation from the distribution of spikes, to the distribution of next spikes – a random walk that only moves forward. Iterating, we can find the probability distribution of the Nth spike:

$$p_N(t) = \int_0^t p_{N-1}(\tau)p(t-\tau)dt.$$

These are convolutions: $p_N = p_{N-1} * p$, so we can write $\hat{p}_N(\omega) = \hat{p}_{N-1}(\omega)\hat{p}(\omega)$, or, iterating,

$$\hat{p}_N(\omega) = \left(\hat{p}(\omega)\right)^N.$$

It's now easy to determine the probability $q(t)$ that there is *any* spike at time t:

$$q(t) = \sum_{N=1}^\infty p_N(t) \text{, so } \hat{q}(\omega) = \sum_{N=1}^\infty \hat{p}_N(\omega) = \sum_{N=1}^\infty \left(\hat{p}(\omega)\right)^N = \frac{\hat{p}(\omega)}{1 - \hat{p}(\omega)}.$$

For a Poisson process of rate λ, $p(t) = \lambda e^{-\lambda t}$ and

$$\hat{p}(\omega) = \int_0^\infty e^{-i\omega t}p(t)dt = \int_0^\infty \lambda e^{-i\omega t}e^{-\lambda t}dt = \frac{\lambda}{-i\omega - \lambda}e^{-(i\omega + \lambda)t}\Big|_0^\infty = \frac{1}{1 + i\omega/\lambda}. \text{ So in this case,}$$

$$\hat{q}(\omega) = \frac{\hat{p}(\omega)}{1 - \hat{p}(\omega)} = \frac{\lambda}{i\omega}. \text{ This is the Fourier transform of } q(t) = \begin{cases} \lambda, t > 0 \\ 0, t < 0 \end{cases}.$$

We can use the Fourier transform-formula for the probability distribution of the Nth spike,

$p_N(t)$, to calculate the power spectrum of the point process. The strategy is to first calculate the cross-correlation, and then use the fact that the power spectrum is the Fourier transform of the cross-correlation. The cross-correlation will be a sum of terms that correspond to pairs of events separated by 0, 1, 2, … spikes, so the Fourier transform of $p_N(t)$ will be crucial, as above, $\hat{p}_N(\omega) = \left(\hat{p}(\omega)\right)^N$. (The full analysis, with appropriate care about limits) is on pages 15-19 of PSPC15-PSPC27.pdf, 2003-2004 notes.)

To calculate the cross-correlation, we assume that there IS an event at time 0 (since, if there is no event, the time series at time 0 has a value of 0, so there is no contribution to the cross-correlation). We break the cross-correlation into three pieces: the correlation of an event at time 0 with the future, the correlation of an event at time 0 with the past, and its correlation with itself.

The Fourier transform of the distribution of events in the future, given an event at time 0, is

$$\sum_{N=1}^{\infty} \hat{p}(\omega)^N = \frac{\hat{p}(\omega)}{1 - \hat{p}(\omega)}.$$

For events in the past, it is the same expression but for the time-inverse of $p(t)$, i.e.,

$$\sum_{N=1}^{\infty} \overline{\hat{p}(\omega)}^N = \frac{\overline{\hat{p}(\omega)}}{1 - \overline{\hat{p}(\omega)}}, \text{ since}$$

$$\overline{\hat{p}(\omega)} = \int_0^{\infty} p(t) e^{+i\omega t} dt = \int_0^{\infty} p(-t) e^{-i\omega t} dt.$$

The contribution of the correlation of the event at time 0 with itself is 1.

Since the probability that there is an event at time 0 is the overall event rate, we will have to multiply these quantities (the probability that there is an event at a time t, given an event at time 0) by the overall rate, λ, to obtain the Fourier transform of cross-correlation – which is the power spectrum.

Thus, the power spectrum is the sum of these contributions, multiplied by the rate λ:

$$P_X(\omega) = \lambda \left(\frac{\hat{p}(\omega)}{1 - \hat{p}(\omega)} + \frac{\overline{\hat{p}(\omega)}}{1 - \overline{\hat{p}(\omega)}} + 1 \right) = \lambda \frac{1 - \left| \hat{p}(\omega) \right|^2}{\left| 1 - \hat{p}(\omega) \right|^2}.$$

Note that in general, as $\omega \to \infty$, $\hat{p}(\omega) \to 0$, so $P_X(\omega) \to \lambda$. I.e. at sufficiently high frequencies, all temporal structure is lost.

Note also that for a Poisson process, for which $p(t) = \lambda e^{-\lambda t}$ and $\hat{p}(\omega) = \dfrac{1}{1 + i\omega / \lambda}$, $P_X(\omega) = \lambda$, exactly.

There's a classic application in neuroscience (Stevens): The postsynaptic noise P_Y can be understood, quantitatively, as the net effect of a synaptic impulse response F and the variability of the presynaptic spike train, P_X. By assuming that the presynaptic activity was Poisson, and measuring the postsynaptic noise, Stevens was able to infer the shape of the synaptic impulse response – which was then directly verified experimentally.

Power spectrum of a channel noise

A simple extension of this approach allows us to calculate the power spectrum of simple channel noise, i.e., random switching processes:

,
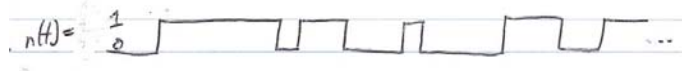
in which we assume that switching between states happens randomly at a rate λ.
The basic idea is that this is the integral of a noise of alternating positive and negative events.

.
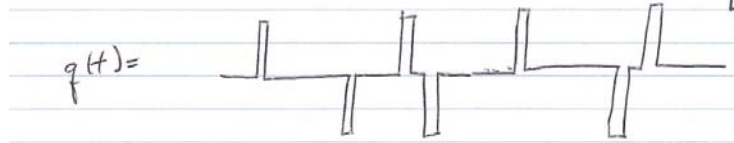
Event times are governed by a renewal process, but they alternate in sign.
Since the channel noise $n(t)$ is the integral of the alternating point process $q(t)$, it follows that each Fourier estimate of $n(t)$ is $1/i\omega$ times the corresponding Fourier estimate of $q(t)$, and that

the power spectra are related by $P_N(\omega) = \dfrac{1}{\omega^2} P_Q(\omega)$.

We can calculate $P_Q(\omega)$ as above, but with alternating signs for each contribution, e.g.,

$$-\hat{q}(\omega) + \hat{q}(\omega)^2 - \hat{q}(\omega)^3 + \hat{q}(\omega)^4 \ldots = \frac{-\hat{q}(\omega)}{1 + \hat{q}(\omega)}.$$

This leads to

$$P_Q(\omega) = \lambda\left(\frac{-\tilde{q}(\omega)}{1 + \tilde{q}(\omega)} + \frac{-\overline{\tilde{q}(\omega)}}{1 + \overline{\tilde{q}(\omega)}} + 1\right) = \lambda\left(\frac{1}{1 + \tilde{q}(\omega)} + \frac{1}{1 + \overline{\tilde{q}(\omega)}} - 1\right) = \lambda\left(\frac{1 - |\tilde{q}(\omega)|^2}{|1 + \tilde{q}(\omega)|^2}\right),$$

and to

$$P_N(\omega) = \frac{\lambda}{\omega^2}\left(\frac{1 - |\tilde{q}(\omega)|^2}{|1 + \tilde{q}(\omega)|^2}\right).$$

For the Poisson case for openings and closings,

$$\hat{q}(\omega) = \frac{1}{1 + i\omega/\lambda}, \text{ and this works out to } P_Q(\omega) = \lambda\frac{\omega^2}{4\lambda^2 + \omega^2} \text{ and } P_N(\omega) = \lambda\frac{1}{4\lambda^2 + \omega^2}.$$

The analysis can be readily extended to more complex channel dynamics (e.g., different opening and closing dynamics, hidden states, etc.); see NAV26-NAV33 of 2008-2009 notes on this.

Combination of independent noise sources

How does the power spectrum of Z relate to that of X and Y?

Fourier estimates simply add up $F(z,\omega,T,T_0) = F(x,\omega,T,T_0) + F(y,\omega,T,T_0)$.

$$|F(z,\omega,T,0)|^2 = \left|F(z,\omega,T,0)\overline{F(z,\omega,T,0)}\right|$$

So we can calculate
$$= \left(F(x,\omega,T,0) + F(y,\omega,T,0)\right)\left(\overline{F(x,\omega,T,0)} + \overline{F(y,\omega,T,0)}\right)$$ But the
$$= F(x,\omega,T,0)\overline{F(x,\omega,T,0)} + F(x,\omega,T,0)\overline{F(y,\omega,T,0)}$$
$$+ F(y,\omega,T,0)\overline{F(x,\omega,T,0)} + F(y,\omega,T,0)\overline{F(y,\omega,T,0)}$$

averages of the cross-terms are zero, since x and y are independent, and individual Fourier estimates have an average of zero.

$$\left\langle|F(z,\omega,T,0)|^2\right\rangle = \left\langle F(x,\omega,T,0)\overline{F(x,\omega,T,0)}\right\rangle + \left\langle F(y,\omega,T,0)\overline{F(y,\omega,T,0)}\right\rangle$$
$$= \left\langle|F(x,\omega,T,0)|^2\right\rangle + \left\langle|F(y,\omega,T,0)|^2\right\rangle$$
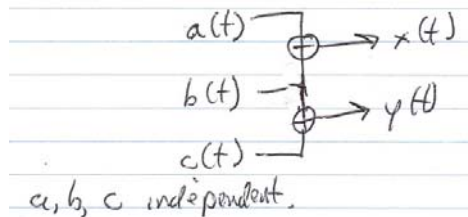
So it follows that $P_Z(\omega) = P_X(\omega) + P_Y(\omega)$, i.e., power spectra of independent noise sources add.

Notice that independence is crucial – if $X = Y$ (and $Z = X + Y = 2X$), then $F(z,\omega,T,T_0) = 2F(x,\omega,T,T_0)$ and $P_Z(\omega) = 4P_X(\omega)$.

Common source

Here, X and Y have their own sources, and also a source in common. This should lead to some kind of a relationship between X and Y, that reflects the properties of Z. How is it manifest?

Since $x = a + b$ and $y = b + c$, we can write the power spectra for X and Y in terms of the power spectra for A, B, and C: $P_X(\omega) = P_A(\omega) + P_B(\omega)$ and $P_Y(\omega) = P_B(\omega) + P_C(\omega)$. This gives no hint that X and Y share a common input.

But if we go back to the Fourier estimates

$$F(x,\omega,T,T_0) = F(a,\omega,T,T_0) + F(b,\omega,T,T_0) \text{ and } F(y,\omega,T,T_0) = F(b,\omega,T,T_0) + F(c,\omega,T,T_0),$$

we see that we can isolate the b-term, since

$$F(x,\omega,T,T_0)\overline{F(y,\omega,T,T_0)} = \left(F(a,\omega,T,T_0) + F(b,\omega,T,T_0)\right)\left(\overline{F(b,\omega,T,T_0)} + \overline{F(c,\omega,T,T_0)}\right).$$

Independence of a, b, and c means that when we calculate the average, the cross-terms don't contribute:

$$\left\langle F(x,\omega,T,T_0)\overline{F(y,\omega,T,T_0)} \right\rangle = \left\langle F(b,\omega,T,T_0)\overline{F(b,\omega,T,T_0)} \right\rangle \text{ -- but this term is not present in the}$$

spectra of X or Y.

The cross-spectrum

The above example motivates the definition of the cross-spectrum, which is a crucial quantity in multichannel analysis:

$$P_{X,Y}(\omega) = \lim_{T\to\infty} \frac{1}{T}\left\langle F(x,\omega,T,0)\overline{F(y,\omega,T,0)} \right\rangle.$$

In the particular case above, $P_{X,Y} = P_B(\omega)$ -- the cross-spectrum identifies and characterizes the common source. So covariance between X and Y, as measured by the cross-spectrum or the coherence, indicates the presence of a common source of variability.

But the cross-spectrum need not be a purely real quantity; the above was a special case in which there is no delay between the influence of the common source on the two observables x and y. If there is a delay (or in general, a difference in dynamics), $P_{X,Y}$ will have a phase shift that reflects this.

As a simple example, say Y receives input from X (which is noisy, and characterized by $P_X(\omega)$) via a linear filter L, and also has an independent noise input B that adds to the signal from X where the noise input is characterized by $P_B(\omega)$.
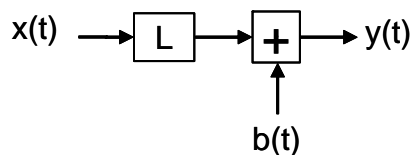
In terms of Fourier estimates, $F(y,\omega,T,T_0) = \hat{L}(\omega)F(x,\omega,T,T_0) + F(b,\omega,T,T_0)$. By the same techniques above (and, making crucial use of the independence of x and b),

$P_Y(\omega) = \left|\hat{L}(\omega)\right|^2 P_X(\omega) + P_B(\omega)$ and $P_{X,Y}(\omega) = \hat{L}(\omega)P_X(\omega)$. Thus, the power spectrum of Y reflects both the variability in the input from X and its own private source, but the cross-spectrum reflects how X and Y are linked, and does not see the common noise source.

Just as the Fourier transform of the spectrum is the autocorrelation, the Fourier transform of the cross-spectrum is the cross-correlation:

$$P_{X,Y}(\omega) = \int_{-\infty}^{\infty} c_{X,Y}(\tau)e^{-i\omega\tau}d\tau \text{ where } c_{X,Y}(\tau) = \langle x(t)y(t-\tau)\rangle.$$

There's an arbitrariness in the definition: whether to put the minus sign on the first or the second term, or, equivalently, what phase indicates that y leads x, or vice-versa. To establish a consistent convention: if y leads x by some amount u, then $y(t-u) = x(t)$, and $c_{X,Y}(\tau)$ will be proportional to $\delta(\tau-u)$, and

$$F(y,\omega,T,T_0) = \int_{T_0}^{T_0+T} y(t)e^{-i\omega t}dt \approx \int_{T_0}^{T_0+T} x(t+u)e^{-i\omega t}dt = e^{i\omega u}\int_{T_0+u}^{T_0+T+u} x(t')e^{-i\omega t'}dt' \approx e^{i\omega u}F(x,\omega,T,T_0)$$

so that $P_{X,Y}(\omega) = e^{-i\omega u}P_X(\omega)$. Note that $P_{X,Y}(\omega) = \overline{P_{Y,X}(\omega)}$.

Related quantities: coherency and coherence

The coherency is the cross-spectrum normalized by the spectrum of the component signals:
$C_{X,Y}(\omega) = \dfrac{P_{X,Y}(\omega)}{\sqrt{P_X(\omega)P_Y(\omega)}}$ (see Homework for why this is useful). Coherency is a complex number whose absolute value must be ≤ 1. The coherence is defined as the magnitude of the coherency, $\left|C_{X,Y}(\omega)\right|$.

Sometimes the term "coherence" is used to refer to the coherency, and sometimes it is used to refer to the square of the quantity defined above (similar to the confusion between "R" and "R-squared" in goodness-of-fit and correlation measures).

Estimation and statistical issues

Multitaper spectral measures of cross-spectra, coherence, and coherency have the same nice statistical properties as the spectrum: each taper provides an approximately independent measure (making it possible to get error bars), and estimates at different frequencies are approximately independent.

The coherence is essentially a correlation coefficient, and its distribution (under the null hypothesis that X and Y are independent) is identical to the distribution of the Pearson correlation coefficient $\rho = \dfrac{<x_i y_i>}{\sqrt{<x_i^2><y_i^2>}}$ of $2M$ pairs of normally-distributed quantities. This is the Fisher distribution; $\tanh^{-1}(\rho)$ is normally distributed.

Nonstationary processes

Can we dispense with the notion of translation-invariance? Strictly speaking, unless we have access to the "parallel universes" (or, multiple instances of exactly the same experiment, including its history), we cannot.

However, if we are willing to assume that the system changes gradually (i.e., on a slow timescale, in comparison with fluctuations of the signal itself), then we can give meaning to the various spectral quantities local in time. For example, we could attempt to make the best estimate of the power spectrum $P_X(\omega)$ from the data in some window $[\tau, \tau + Q]$, and assign this value to $P_X(\omega, \tau)$. This quantity is known as the "spectrogram."

The multitaper method

The above idea puts a premium on using the data in $[\tau, \tau + Q]$ efficiently. The multitaper estimate is

$$P_X(\omega, \tau) = \frac{1}{M} \sum_{m=1}^{M} \left| \int_0^Q W_m(t,Q) x(t+\tau) e^{-i\omega t} dt \right|^2, \text{ where } W_m(t,Q) \text{ is the } n\text{th Slepian function on the}$$

interval $[0,Q]$.

$$\int_{-\infty}^{\infty} P_X(\omega, \tau) d\tau = P_X(\omega) \text{ and } \int_{-\infty}^{\infty} P_X(\omega, \tau) d\omega = \left\langle x(\tau) \right\rangle^2, \text{ so the spectrogram can be thought of as}$$

parsing out variability into frequency-dependent components, time-dependent components, and their interactions.

One can think of a spectrogram as being a descriptor of how the original signal projects onto each of the basis functions shown below. For each non-overlapping window of length Q, there is a separate subset of basis functions that are nonzero just in that window (Slepian functions).
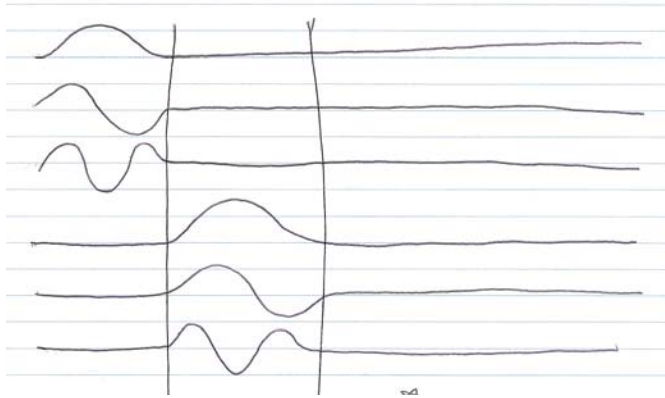
Similarly, cross-spectra can be generalized to cross-spectrograms, coherences to coherograms, etc. All are readily calculable, with the correct normalizations, units, error bars, etc., in Chronux.

Generally, these quantities are calculated with the start time τ stepping through the data, and a fixed Q. Larger Q enables greater frequency resolution ($2\pi M / Q$) and/or higher precision (M at the same frequency resolution), but lower time resolution ($1/Q$)– a tradeoff that cannot be avoided. Results are typically displayed as a "spectrogram",
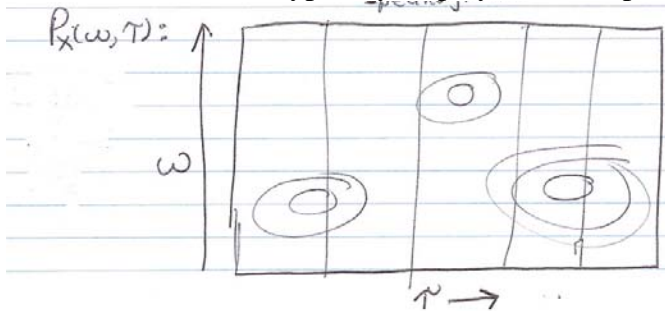
A nice aspect of this is that non-overlapping estimates (frequencies separated by ($2\pi M / Q$, or start times separated by Q) are approximately statistically independent, since tapers that start at separate times are orthogonal, and tapers at the same time are also orthogonal.

Wavelets

"Wavelets" are another approach to time-frequency analysis. Briefly, one first chooses a "mother wavelet" $M(t)$
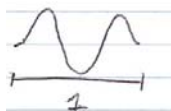

and then convolves it with the signal at all start positions, and with all scales:

$$M_X(s,\tau) = \int_{-\infty}^{\infty} x(t+\tau)M(t/s)dt = \int_{-\infty}^{\infty} x(t)M(\frac{t-\tau}{s})dt \, .$$

Then, $\langle |M_X(s,\tau)|^2 \rangle$ (or, in the absence of parallel universes, $|M_X(s,\tau)|^2$), can be considered analogous to $P_X(1/s,\tau)$. Cross-"spectral" objects can also be defined when multiple time series are available, such as $\langle M_X(s,\tau)M_Y(s,\tau)\rangle$.

Typically, one restricts to scales s that differ by powers of 2. This, essentially, is projecting the signal onto functions that look like:
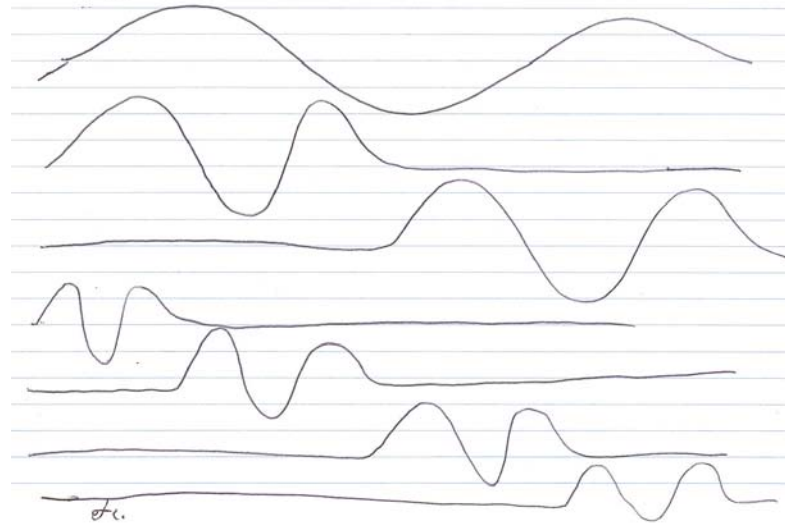

The "Morlet" wavelet is standard: $M(t) = c_\sigma \pi^{-1/4} e^{-t^2/2}\left(e^{i\sigma t} - e^{-\sigma^2/2}\right)$, where $\sigma = 5$ is a typical choice that yields a few oscillations per envelope.
But there are many others, which can be chosen based on considerations such as the basic aspect ratio of the time-frequency tile, speed of computation, whether their envelopes are infinite in both directions, one direction, or neither, etc.

Note that in contrast to the basis functions used in multitaper time-frequency analysis, the functions associated with longer periods have greater durations. Thus, for any given wavelet analysis, time resolution and frequency resolution are inversely proportional *within the analysis*; in the multitaper case, they are uniform.

Thus, wavelet analysis is likely to provide a more efficient description if events that have short timespans are composed of higher frequencies, while the multitaper approach is likely to provide a more efficient description if the frequency composition and duration of events are independent.

This can be summarized in terms of how the two approaches parcellate the time-frequency plane:
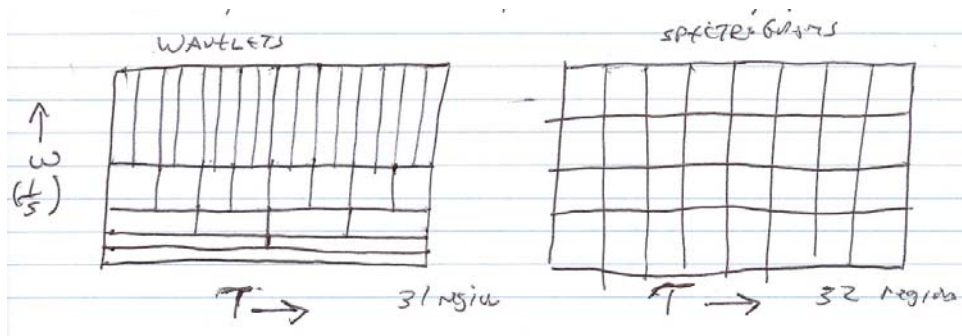
Wavelet functions are typically not orthogonal (but, with the proper choices, can be made to form a basis) – so it is harder to do statistics on them.