CONTENTS

Regions marked in yellow are out of date or incorrect. See the other two documents instead.

A BRIEF PRIMER IN REVERBERATION MAPPING

INTRODUCTION

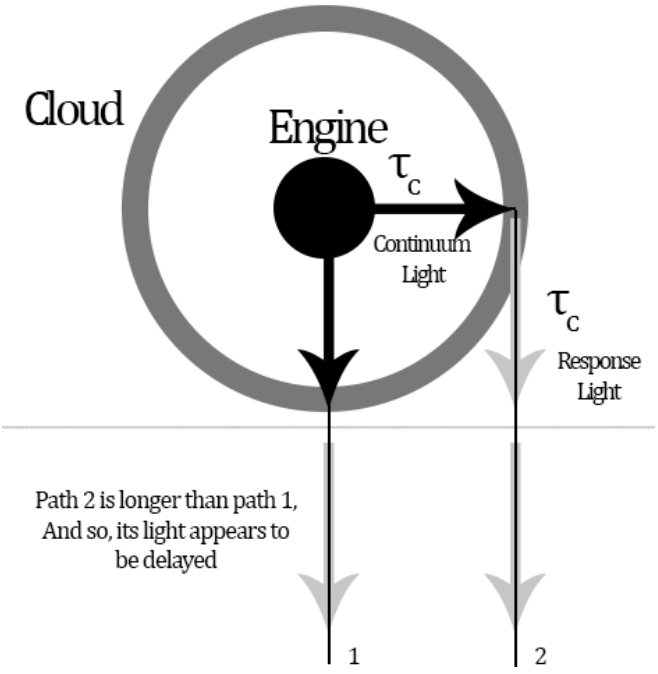
Reverberation mapping is the process of measuring the delays between the different light signals from active galactic nuclei and using that information to recover physically meaningful information about the nuclei's structure. Even the most basic excursion into the field can be difficult without some basic knowledge, and so this guide has been prepared as an extensive rundown of the basics for the completely uninitiated.

THE TASK

"Active Galactic Nuclei" (AGN's) is a catch-all term for the core of galaxies that are actively producing noticeable amounts of light. As far as we're concerned, they have two distinct components:

1. **The Engine:** A central black hole/accretion disk that spits out light at all wavelengths (the "continuum" signal); and

2. **The Dust Cloud:** A hollow gas cloud "shell" of indeterminate shape that surrounds the engine. When the continuum signal excites this gas, spectral emission lines (the photometric signal) are produced.



Simplified AGN Structure

The AGN as a whole is often too small and distant to resolve these parts individually: instead we need to try and discern the structure of the AGN using *only* the continuum and photometric data. This concept is called **reverberation mapping**: using signal responses to try and resolve information about the shape and size of a complex structure. (Peterson & Horne, 2004)

This is of practical interest for a very simple reason: It takes time for the light of the engine to reach the clouds. If we can measure this delay, we measure the AGN's size, and by extension the AGN's mass.

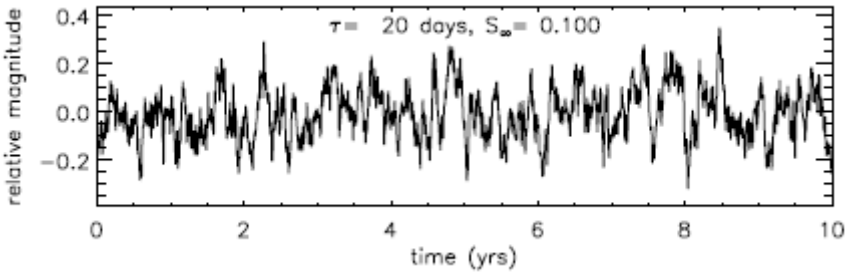
THE HURDLES

Difficulty arises from the fact that light curves tend to have large seasonal gaps in observation, and a lack of consistent light curve shapes means that we have a hard time interpolating the behaviour while we have our back turned

Even in the simple case of estimating the characteristic delay of an AGN, there are three main hurdles that make our life difficult, each of which exacerbate each-other:

1: The Continuum Curve Isn't Smooth or Predictable

AGN signals are inherently random; there's no established 'shape' that we can assume that the light curve will take. This is a big problem for us, because it makes it almost impossible to interpolate what the light curve is doing in between measurements.



A Simulated AGN Light Curve (MacLeod C L, 2010)

Even though the continuum curve isn't easily predictable, it *does* follow a well defined pattern of randomness. Studies have shown that the continuum curves of well-observed AGN's match reasonably well with the "**Damped Random Walk**" (DRW), a stochastic process with bounded uncertainty.

The DRW is similar to the traditional random walk, where the "size" and "direction" of each discrete step are randomized, but with an additional term dragging the signal back towards an equilibrium value:

$$\frac{ds_c}{dt} = -\frac{s_c - \bar{s_c}}{\tau_d} + dW$$

Interpolation is still difficult, but it does tell us two crucial things about the signal:

1. How signal uncertainty increases as we move away from a measurement; and
2. How the signal autocorrelates with itself

Additionally, the DRW is entirely defined by only **three** parameters:

1. Its inherent variance, σ_∞
2. Its damping timescale, τ_d
3. Its average, or "baseline", $\bar{s}$

This is discussed more further on, in the "Damped Random Walk" section.

2: The Response Isn't a Simple Delay

Another issue complicating our life is that there isn't a simple one to one relationship between spikes in the continuum and spikes in the photometric curves. We can see a single spike give a "blurred" response from a finite shell thickness, complications from reflections off the "back" and "sides" off the shell, and any number of other issues arising from the shape and structure of the physical system.

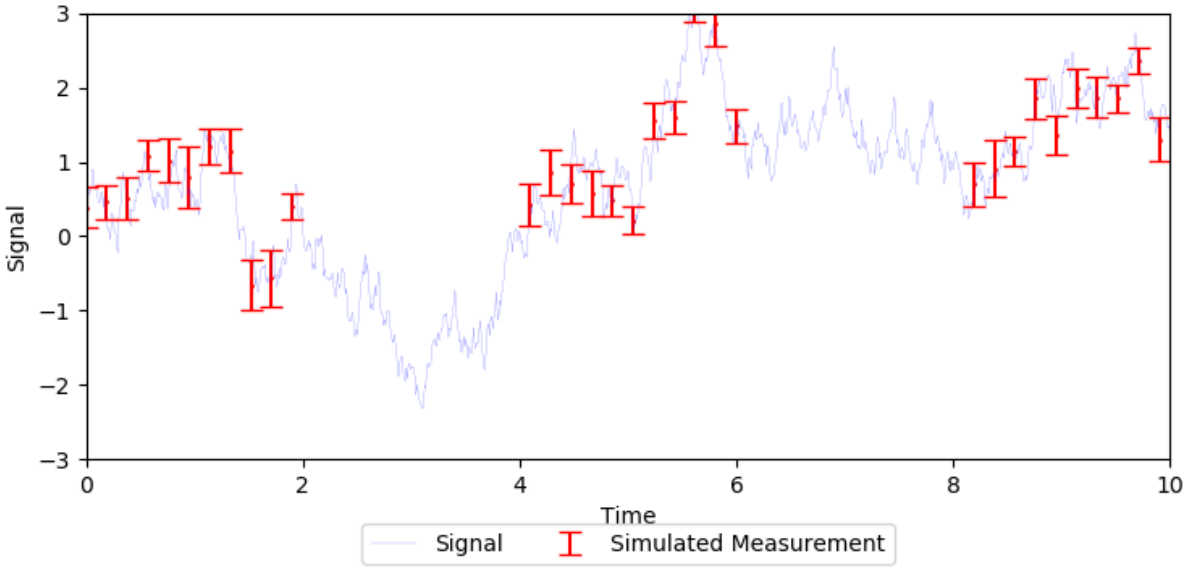
We describe response of the photometric curves $s_{p,i}$ with the **transfer function** ψ, the photometric response that each point on the continuum s_c generates:

$$ds_{p,i}(t) = \psi(t - \Delta t) \cdot df(t)$$
$$s_{p,i}(t) = \int_{\infty}^{\infty} \psi(t - \Delta t) \cdot s_c(t)\, dt$$

Actually measuring the transfer function requires intense observation of both the continuum and photometric curves, and so we instead have to take the messy step of making a reasonable guess as to what ψ might look like, and hope that it doesn't skew our results too badly.

3: Our Observations Are a Little Patchy

In a perfect world, we'd have arbitrarily precise and continuous measurements of *all* of the signal curves. In practice, our measurements are discrete and often sparse, with significant uncertainties. Worst of all, many distance AGN's will have half-year gaps with no measurement *at all*, simply because we couldn't get a good look at them.



Simulated DRW with Fake Seasonal Measurements

This is the fundamental problem of AGN reverberation mapping: the characteristic delay is of a similar timescale to enormous gaps in the signals. To try and work around this, we need to make the best use of every measurement we have; combining data from every signal we have in a consistent way to maximize the reliability and precision of our output.

INTRODUCING THE DAMPED RANDOM WALK

A major challenge in analysing AGN signals is that they don't follow smooth, predictable light curves like some astronomical objects do; even if we have perfectly accurate measurements at two times, we can't know exactly what the signal was doing between these points.

However, we have found that this random behaviour follows certain probabilistic patterns, specifically those of the "damped random walk" (DRW), a stochastic (random) process (MacLeod C L, 2010) (Zu, Kochanek, Kozlowksi, & Udalski, 2013). By knowing how these random systems evolve, we can at the very least put some reasonable limits on what the light curve *might* have been doing in between measurements.

THE RANDOM WALK

To understand the damped random walk from scratch, it can be helpful to understand its more basic cousin: the regular "random walk". Imagine a person standing at position 0, who flips a coin. If the coin lands heads, they take a step left, if it lands tails, they take a step left. In this discrete case, this looks like:

$$x_{i+1} = x_i + \mathrm{dW}, \qquad dW \in [-1,1\,]$$

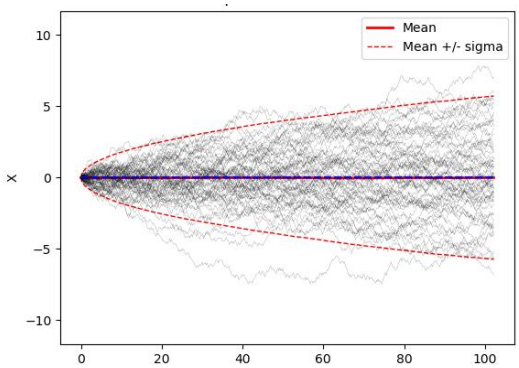
This is an example of an "Ornstein Uhlenbeck (OU) process", which covers any physical process in which the first order derivative has some random element dW added to it. In this more general description, the random walk can be written defined by the stochastic differential equation:

$$\frac{df}{dt} = k\, dW$$

Where dW is some continuously varying stochastic (random) variable with bounded variance:

$$\langle (dW)^2 \rangle = 1, \qquad \langle dW \rangle = 0$$

The 'Walk' part of the name refers to the fact that, over each time increment dt, the signal takes a 'step' up or down, causing the signal to meander diffusively away from its starting point:



Ensemble of Random Walks

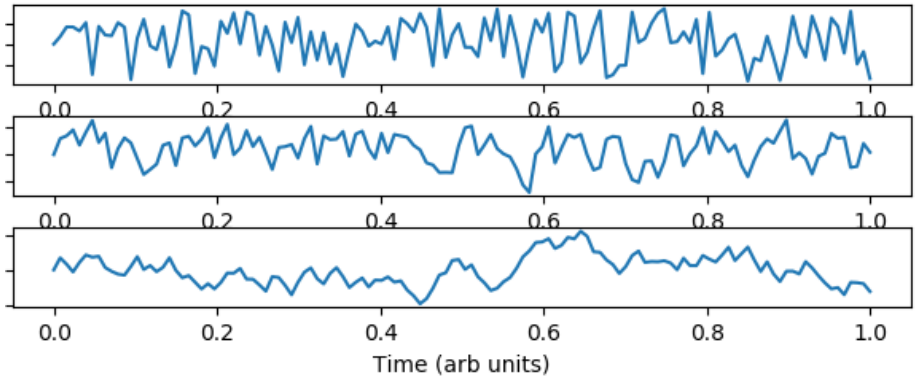
You'll note that, even though each individual "realisation" of the random walk is almost entirely unpredictable, the ensemble behaviour of arbitrarily many *does* follow a smooth distribution. In the random walk's case, the Gaussian spread of the distribution increases to infinity (in a square root fashion, as it happens), and is **unbounded**.

THE DAMPED RANDOM WALK

In well-observed AGN's, we've found that their continuum light curves fit reasonably well with a close cousin of the random walk, the "Damped Random Walk" (DRW):

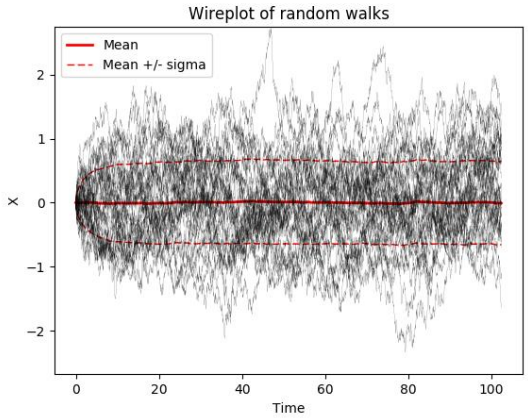
$$\frac{df}{dt} = -\frac{f - \bar{f}}{\tau_d} + AdW$$

The 'damped' part of the name refers to the negative proportional component $-\frac{f}{\tau_d}$. The damping timescale, τ_d, determines the timescale of the DRW as a whole. The figure below shows, from top to bottom, $\tau_d = 1$, 2 and 10.



Simulated DRW's at Increasing Timescales

This has the effect of constraining the signals 'walk', so that, unlike the random walk, its variance plateaus out to a stable maximum:



Ensemble Bheaviour of Many Damped Random Walks

This mean/variation maps nicely to a time-varying Gaussian distribution, which, for the DRW, follows an exponential decay in the mean, and an exponential increase in the variance (square of standard deviation):

$$\mu(t) = \bar{f} + (f_0 - \bar{f})e^{-\frac{\Delta t}{\tau}}, \qquad \sigma(t) = \sigma_\infty \sqrt{1 - e^{-2\frac{\Delta t}{\tau}}}$$

Where f_0 is the starting position of the walk, and σ_∞ is the standard deviation that is plateaus out to.

The description of the standard deviation is sometimes called the **structure function** in reverberation mapping literature. This name can be a bit confusing, because it's also used as the name for a few other unrelated concepts in signal analysis.

Looking at the variance and mean functions, we can see that the DRW at large is entirely described by only three parameters:

1. The average, or "baseline", $\bar{f}$
2. The inherent variance, σ_∞
3. The damping timescale, τ_d

When we have observation measurements for a DRW, we can make decent approximations about the baseline, but the remaining parameters are surprisingly hard to get accurate estimates of from sparse data. Rather than trying to calculate σ_∞ and τ_d from the data, it's common to test many values numerically to find the best fit.

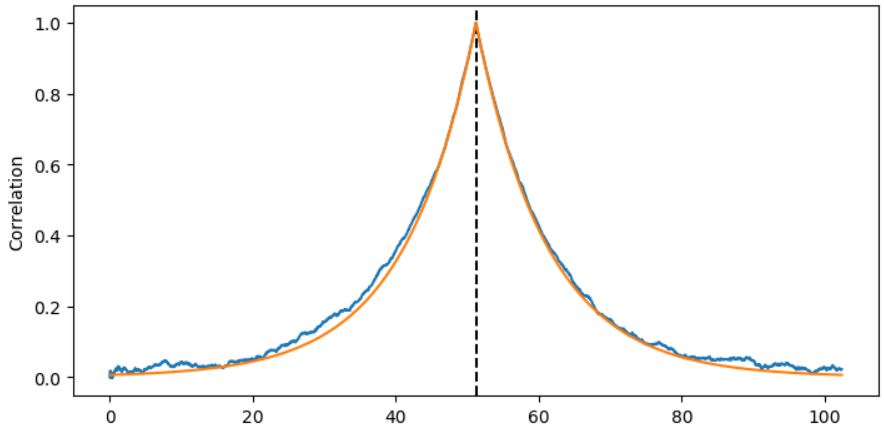
CORRELATION & COVARIANCE

Another useful fact about the DRW is that it has a well-defined **autocorrelation function**:

$$\phi(\Delta t) = \langle s(t)s(t+\Delta t)\rangle = e^{-\frac{|\Delta t|}{\tau_d}}$$

In layman's terms: nearby points in the DRW are likely to be similar, but become exponentially less so at further times. This is incredibly useful for a simple reason: knowing correlations gives us the covariance between any measurements we make:

$$\phi_{ij} = \langle s_i s_j \rangle = \sigma_\infty^2 e^{-\frac{|t_i - t_j|}{\tau_d}}$$



Numerical (Blue) & Theoretical (Orange) Autocorrelation Function of a Simulated DRW

Equipped with this, we can assemble a covariance matrix for our data, and refine our measurements based on the surrounding datapoints.

If you're unfamiliar with correlation and covariance, check out the quick rundown in the accompanying document.

THE EVOLUTION OF SINGLE-POINT UNCERTAINTY

Even though we're often interested in reconstructing continuum curves from many data-points, we can actually arrive at a simple analytical solution for a single point curve. This is also a useful example for illustrating some of the underlying principles at play, and is worth going over for new-comers.

If We Know the Baseline

Recall that the probability distribution for a DRW starting at f_0 at $t = 0$ is:

$$\mu(t) = \bar{f} + \left(f_0 - \bar{f}\right)e^{-\frac{\Delta t}{\tau}}$$
$$\sigma(t) = \sigma_\infty\sqrt{1 - e^{-2\frac{\Delta t}{\tau}}}$$

Notice that this rearranges to:

$$\mu(t) = \bar{f}\left(1 - e^{-\frac{\Delta t}{\tau}}\right) + f_0 e^{-\frac{\Delta t}{\tau}}$$

In the case where there's (Gaussian) error in the initial measurement, we need to account for this variance in $\sigma(t)$. Keeping in mind that uncorrelated variances stack add in quadrature, a measurement uncertainty of σ_E (one standard deviation) gives us:

$$\sigma^2(t) = \sigma_\infty^2\left(1 - e^{-2\frac{\Delta t}{\tau}}\right) + \sigma_E^2 e^{-2\frac{\Delta t}{\tau}}$$

If We Don't Know the Baseline

If we don't know the baseline, we need to add an additional variance:

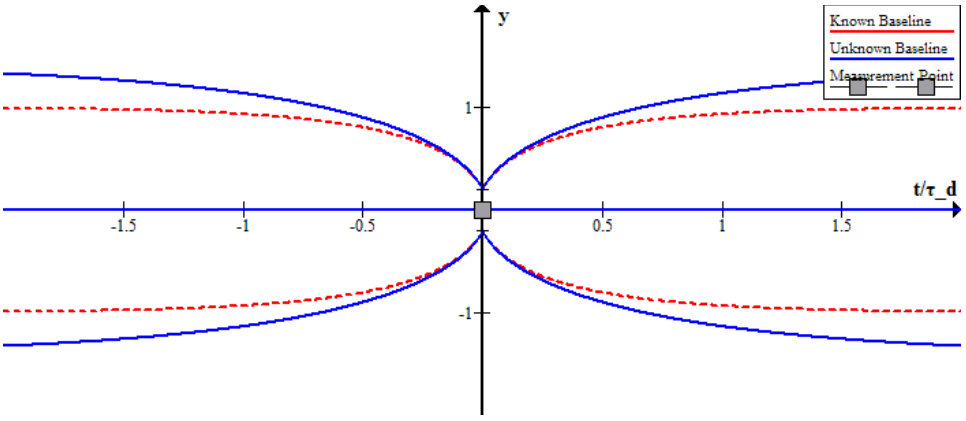
$$\sigma^2(t) = \sigma_\infty^2\left(1 - e^{-2\frac{\Delta t}{\tau}}\right) + \sigma_E^2 e^{-2\frac{\Delta t}{\tau}} + \sigma_{\bar{f}}^2\left(1 - e^{-\frac{\Delta t}{\tau}}\right)^2$$

But, if we only have a single measurement, we can base our uncertainty in the baseline on the measurement error and the DRW's inherent variance:

$$\sigma_{\bar{f}}^2 = \sigma_\infty^2 + \sigma_E^2$$

Which ends up giving us:

$$\sigma_i^2(t) = \sigma_\infty^2\left(2 - 3e^{-\frac{t}{\tau}}\right) + \sigma_E^2\left(1 - 2e^{-\frac{t}{\tau}} - 2e^{-\frac{2t}{\tau}}\right)$$



SIMULATING DRW'S

As long as we're forecasting forward, simulating a DRW is as straight forward as using a discrete Euler scheme with a random element:

$$f_{n+1} = f_n - \frac{f_n - \bar{f}}{\tau_d}\Delta t + \sigma_\infty \sqrt{\frac{\Delta t}{2\tau_d}}\frac{Z}{\sigma_Z}$$

Where Z is some random number, and σ_Z is the variance in the method used to generate it. If Z is generated randomly with symmetrical probability distribution $P(z)$, then:

$$\sigma_z^2 = \langle Z^2 \rangle = \frac{\int_\infty^\infty P(z) z^2 \, dz}{\int_\infty^\infty P(z) \, dz}$$

For example:

Method	$\boldsymbol{\sigma_z}$
Binary	1
Square	$\frac{1}{\sqrt{3}}$
Gaussian	1

e.g., for 'Z' being a Gaussian generated random number $Z \sim norm(0,1)$, and a known baseline of $\bar{f} = 0$:

$$\begin{aligned} f_{n+1} &= f_n - \frac{f_n}{\tau_d}\Delta t + \sqrt{\frac{\Delta t}{2\tau_d}}\, Z \\ &= f_n\left(1 - \frac{\Delta t}{\tau_d}\right) + \sqrt{\frac{\Delta t}{2\tau_d}}\, Z \end{aligned}$$

In cases where we're running a large number of simulations (e.g. brute forcing a continuum curve) it's more efficient to calculate the bracketed terms ahead of time:

$$\begin{aligned} f_{n+1} &= a f_n + bZ \\ a &= \left(1 - \frac{\Delta t}{\tau_d}\right), \qquad b = \sqrt{\frac{\Delta t}{2\tau_d}} \end{aligned}$$

DIFFICULTIES WITH TRANSFER FUNCTIONS

We describe the response of the photometric curves $s_{p,i}$ with the **transfer function** ψ. This is the photometric response that each point on the continuum s_c generates:

$$ds_{p,i}(t) = \psi_i(\Delta t) \cdot s_c(t - \Delta t)\, d\Delta t$$

Such that the entire response is the convolution of the continuum with the transfer function:

$$s_{p,i}(t) = \int_{\infty}^{\infty} \psi_i(\Delta t) \cdot s_c(t - \Delta t)\, d\Delta t$$

This transfer function, along with the auto-correlation function of the continuum light curve, acts as a description of our entire understanding of the physical system.

But what does this transfer function look like? Unfortunately, this doesn't have a simple answer. In a perfect world, the response of the emission line signals from to the continuum would manifest as a simple, constant delay, i.e.:

$$s_p(t) \propto s_c(t - \tau_c), \qquad \psi(\Delta t) = \delta(\tau_c)$$

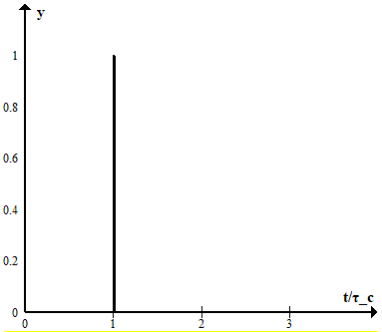
Where τ_c is the 'characteristic delay' of the system: the time taken for the continuum light to traverse the radius of the cavity:

$$\tau_c = \frac{c}{r_{cavity}}$$

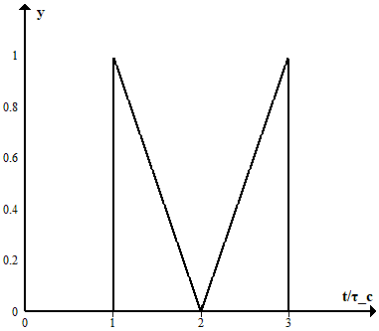
In practice, however, things are rarely so simple. Virtually every element of the AGN's structure can introduce complications to this response.

The Cloud Can Have Different Geometries

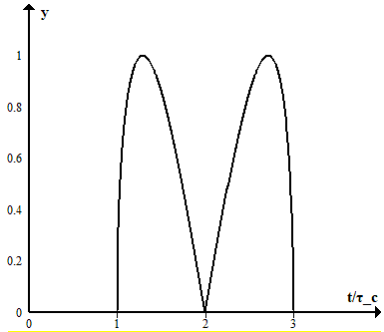
We don't know what shape the cloud might take. Even the simplest cases have massive differences in their corresponding transfer functions, and that's before we take into account other complicating issues. The simplest case we might imagine is that of a thin, axisymmetric cloud in which each angular region gives off a 1-1 response immediately after the engine light reaches it. Even in this fairytail world, there's huge differences based on exactly how that cloud is shaped.



Uniform Disk (Edge-On)



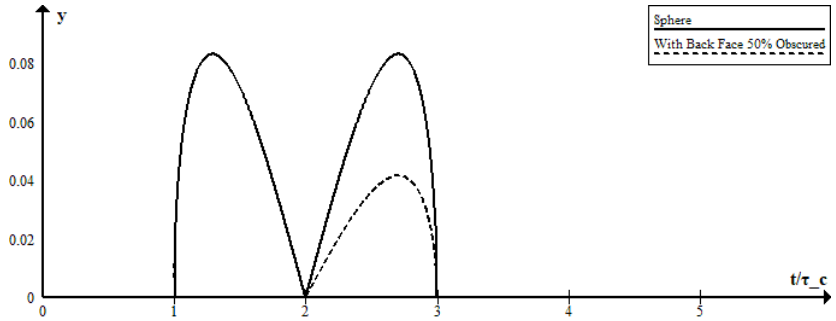
Uniform Disk (Face-On)



Uniform Sphere

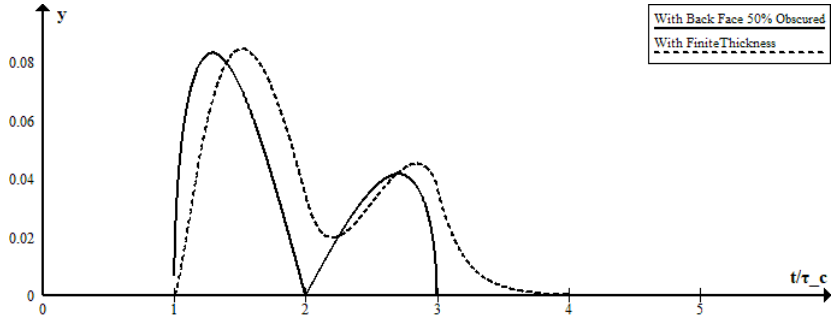
The Shell Might Self-Obscure

Unless the shell perfectly fits on a thin-plane, there's also the possibility that the nearer face will obscure the light from the farther face, limiting or entirely blocking the light from the "back" of the AGN. If the shell is opaque enough, this could even block light from the inside of the shell, making only



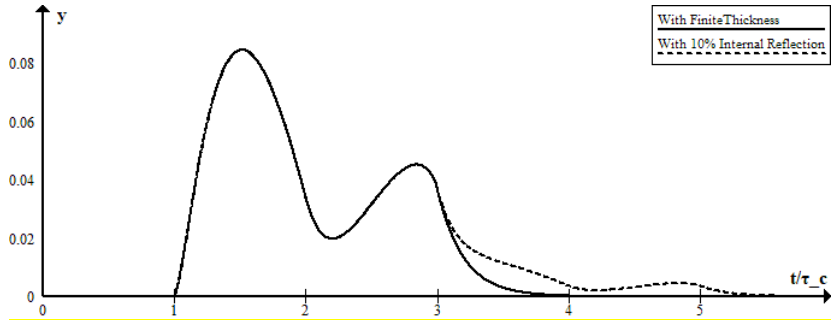
The Shell Might Have Significant Thickness

The above examples assume that each radial line excites one and only one patch of the shell. In practice, each line of sight from the engine outwards will pass through multiple layers, with outer layers giving weaker and more delayed responses:



The Cloud Might Reflect Light Inside the Cavity

Complicating things even further is the possibility that some of the continuum signal could be reflected off the cloud back into its interior, causing even further response:



It's easy to see how quickly a "blind" predictive model of the transfer function gets out of hand: even if we make a bold assumption bout the general behaviour, we still need to introduce any number of extra parameters in doing so. It is technically possible to reconstruct a transfer function through observation of the two signals, but doing so reliably requires us to have reliable measurement data, the opposite of the situation that we're dealing with.

A PRACTICAL SIMPLIFICATION

In practice, we tend to just "guess" a simple shape for the basic transfer functions, and carry on from there. Many simple cases stick with a delta-function (1-1 response with a set delay) but Zu opts for the more advanced option of assuming that each transfer function is a "top-hat"

Tophat Function

The tophat function is pretty straightforward: a set area below a flat line of given width and position:

$$\psi = \begin{cases} \frac{A}{b} & \tau_c \leq \Delta t \leq \tau_c + b \\ 0 & otherwise \end{cases}$$

Note that, as in Zu's papers, the function is sometimes described in terms of its start and end times instead of its start and width:

$$\tau_c = t_1$$
$$b = t_2 - t_1$$

HANDLING CORRELATIONS WITH TRANSFER FUNCTIONS

One advantage of setting our foot down on a rough approximation like this is that we can use it to get correlations *between* different light curves, something that's crucial to making use of as much data as possible.

In general, the correlation between one point on the continuum and one point on a response curve is:

$$\langle s_c(t_i) s_k(t_j) \rangle = \int \psi(t_i - t') \cdot \langle s_c(t') s_c(t_j) \rangle \, dt'$$

While the correlation between two points on the same response curve is:

$$\langle s_a(t_i) s_a(t_j) \rangle = \int \int \psi_a(t_i - t') \psi_a(t_j - t'') \cdot \langle s_c(t') s_c(t'') \rangle \, dt' dt''$$

And the correlation between two points on different responses (for example 'a' and 'b' is):

$$\langle s_a(t_i) s_b(t_j) \rangle = \int \int \psi_a(t_i - t') \psi_b(t_j - t'') \cdot \langle s_c(t') s_c(t'') \rangle \, dt' dt''$$

Continuum-Line Covariance

Given that we now have a fixed description of ψ, and our knowledge of the DRW gives us the continuum-continuum function $\langle s_c(t)s_c(t'')\rangle$. Here, we presents Zu's results for these correlations, assuming that the transfer function is a tophat function of height 'h' and width 'w', i.e.:

$$\psi(\Delta t) = \begin{cases} h & \tau_c \leq \Delta t \leq \tau_c + w \\ 0 & otherwise \end{cases}$$

Firstly, the covariance between the continuum at time 't_i' and a response line s_k at time 't_j', i.e. with a time difference of $t_i - t_j = \Delta t$

$$\langle s_c(t_i)s_k(t_j)\rangle = \tau_d^2\sigma_\infty^2 A \begin{cases} \exp\left(-\frac{\Delta t - \tau_c}{\tau_d}\right) - \exp\left(\frac{\Delta t - (\tau_c + b)}{\tau_d}\right) & \Delta t < \tau_c \\ 2 - \exp\left(\frac{\Delta t - (\tau_c + b)}{\tau_d}\right) - \exp\left(-\frac{\Delta t - \tau_c}{\tau_d}\right) & \tau_c \leq \Delta t \leq \tau_c + w \\ \exp\left(-\frac{\Delta t - (\tau_c + b)}{\tau_d}\right) - \exp\left(-\frac{\Delta t - \tau_c}{\tau_d}\right) & \Delta t > \tau_c + w \end{cases}$$

For the sake of simplicity, Zu defines:

$$t_L = \Delta t - (\tau_c + b)$$
$$t_H = \Delta t - \tau_c$$

Such that this may be more compactly written:

$$\langle s_c(t_i)s_k(t_j)\rangle = \tau_d^2\sigma_\infty^2 A \begin{cases} \exp\left(-\frac{t_H}{\tau_d}\right) - \exp\left(\frac{\mathrm{t_L}}{\tau_d}\right) & \Delta t < \tau_c \\ 2 - \exp\left(\frac{\mathrm{t_L}}{\tau_d}\right) - \exp\left(-\frac{t_H}{\tau_d}\right) & \tau_c \leq \Delta t \leq \tau_c + w \\ \exp\left(-\frac{\mathrm{t_L}}{\tau_d}\right) - \exp\left(-\frac{t_H}{\tau_d}\right) & \Delta t > \tau_c + w \end{cases}$$

From: (Zu, Kochanek, & Peterson, An Alternative Approach to Measuring Reverberation Lags in Active Galactic Nuclei, 2011)

Line-Line Covariance

We can similarly arrive at an expression for the covariance *between* two respone lines, $s_1(t)$ and $s_2(t)$ (we've used subscripts 1 and 2 to represent **any** arbitrary pair of response lines here).

> ***Note:*** *There's a good chance that these equations as given by Zu have some minor errors in them. There are some contradictory definitions. They're presented here for completeness, but the reader should be aware that they may have inherited some errors.*

As before, we assume that the transfer function between the continuum and the lines are tophat functions, but set $b_1 \geq b_2$. If this isn't the case, the symmetry of covariance means you can switch the two around so that it is.

Zu first sets up some shorthand variables:

$$\begin{aligned} t_L &= \Delta t - (\tau_{c1} + w_1 - \tau_{c2}), \\ t_{M1} &= \Delta t - (\tau_{c1} + w_1 - \tau_{c2}) \\ t_{M2} &= \Delta t - (\tau_{c1} - \tau_{c2}) \\ t_H &= \Delta t - \big(\tau_{c1} - (\tau_{c2} + w_2)\big) \end{aligned}$$

And gives the covariance as:

$$\langle s_1(t_i) s_2(t_j) \rangle = \tau_d^2 \sigma_\infty^2 A_1 A_2 \cdot \left(\exp\left(-\frac{|t_L|}{\tau_d}\right) + \exp\left(-\frac{|t_H|}{\tau_d}\right) - \exp\left(-\frac{|t_{M1}|}{\tau_d}\right) - \exp\left(-\frac{|t_{M2}|}{\tau_d}\right) + z \right)$$

Where 'z' is a piecewise component:

$$z = \begin{cases} 2\dfrac{t_H}{\tau_d} & t_{M2} \leq 0 < t_H \\ 2w_2 & t_{M2} \leq 0 < t_H \\ -2\dfrac{t_L}{\tau_d} & t_L \leq 0 \leq t_{M1} \\ 0 & t_L > 0 \; or \; t_H < 0 \end{cases}$$

Notice that this also works as the covariance of response curve with itself, we just need to set:

$$A_1 = A_2 = A, \qquad w_1 = w_2 = w, \qquad \tau_{c1} = \tau_{c2} = \tau_c$$

To get:

$$\begin{aligned} t_L &= \Delta t - w \\ t_{M1} &= \Delta t - w \\ t_{M2} &= \Delta t \\ t_H &= \Delta t - w \end{aligned}$$

$$\langle s_k(t_i) s_k(t_j) \rangle = \tau_d^2 \sigma_\infty^2 A^2 \cdot \left(\exp\left(-\frac{|\Delta t|}{\tau_d}\right) - \exp\left(-\frac{|\Delta t - w|}{\tau_d}\right) + z \right)$$

These covariances are crucial to the process of reverberation mapping, as they're used to populate the **covariance matrix**. This matrix (which shows up in the curve recovery section) encodes all of our information and assumptions about the system, and describes how our various measurements relate to one another.

> ***Note:*** *The above result is not given by Zu, but an extension on the (possibly erroneous) prior equations.*
>
> From: (Zu, Kochanek, & Peterson, An Alternative Approach to Measuring Reverberation Lags in Active Galactic Nuclei, 2011)

CORRELATION FUNCTIONS & CURVE RECOVERY FROM DISCRETE DATA

THE DISCRETE CORRELATION FUNCTION

Calculating the correlation function of two datasets can be a bit tricky when they're irregularly or sparsely sampled. We might interpolate and get the correlation by integration or FFT, but:

> *"When, as is usually the case, the fluctuation power spectrum has substantial amplitude at frequencies above the mean sampling rate, interpolation is dangerous"*
>
> **-** (Edelson & Krolik, 1988)

i.e. interpolation can give us a false confidence in misleading results if the timescales of sampling is larger than the delay we're trying to measure. A more conservative approach is the **discrete cross correlation function**, or 'DCF', which makes no assumptions **at all** about the 'shape' of the underlying signals, or of the relationship/transfer between them. In this way, it's probably the most general, though not particularly precise, way of approaching the sparse data correlation problem.

The approach is pretty straightforward:

1. Divide your timeline into 'bins' of arbitrary (but usually regular) size
2. Choose a point on signal x. Run through all the points in signal y. If the time between the two points is inside a particular bin's range, add $(x_i - \bar{x})(y_j - \bar{y})$ to that bin's "score"
3. Repeat for all points in signal 1
4. Average the scores for each bin, giving a measure of covariance
5. Divide by signal variability to get the correlation

Or, in math-talk:

$$S_k = \sum_{ij} (x_i - \bar{x})(y_j - \bar{y}), \qquad \forall t_i - t'_j \in [\tau_k, \tau_{k+1}]$$

$$N_k = \sum_{ij} 1, \qquad \forall t_i - t'_j \in [\tau_k, \tau_{k+1}]$$

$$\phi_k = \frac{S_k}{N_k \sigma_x \sigma_y}$$

With normal discrete calculations of the signal averages and standard deviations.

The advantage of the DCF is its generality: it assumes nothing. However, this is also its biggest shortfall: it fails to make the best use of our understanding of the underlying system that has produced the signals. Useful for an initial reconnaissance, but not suitable for analysis of systems where we have a better grasp of the physics at hand.

CRUDE UNCORRELATED DRW INTERPOLATION

Recall that we had a description of how a DRW's ensemble distribution evolves near a single measurement point.

When we have multiple points, we can crudely stitch these together into a conservative estimate of the curve as a whole. First, we estimate the signal baseline with the inverse variance weighted average:

$$\bar{y} = \left(\sum_i \frac{1}{\sigma_{E,i}^2 + E_i^2} \right)^{-1} \sum_i \frac{x_i}{\sigma_{E,i}^2 + E_i^2}, \qquad \sigma_{\bar{y}}^2 = \left(\sum_i \frac{1}{\sigma_{E,i}^2 + E_i^2} \right)^{-1}$$

And then use these to assemble the distributions about each point. For each measurement, y_i, at time t_i, the probability distribution at time $t = t_i \pm \Delta t$ looks like:

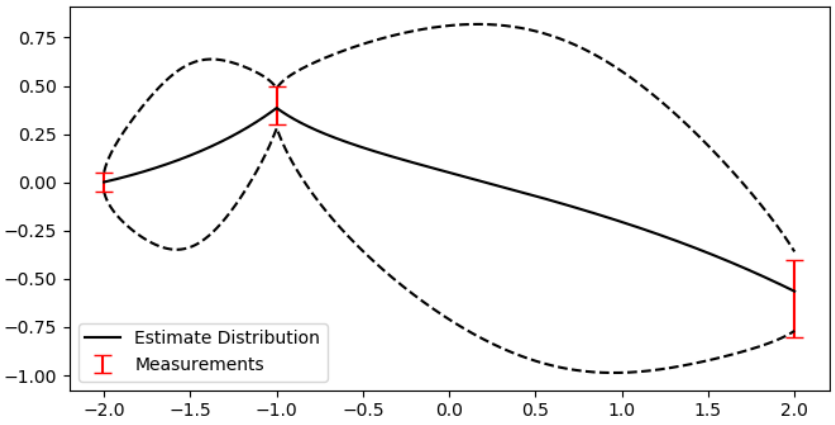
$$\mu_i(t) = \bar{y}\left(1 - e^{-\frac{\Delta t}{\tau}}\right) + y_i e^{-\frac{\Delta t}{\tau}}$$

$$\sigma_i^2(t) = \sigma_\infty^2 \left(1 - e^{-2\frac{\Delta t}{\tau}}\right) + \sigma_{E,i}^2 e^{-2\frac{\Delta t}{\tau}} + \sigma_{\bar{y}}^2 \left(1 - e^{-\frac{\Delta t}{\tau}}\right)^2$$

We then combine these estimates using normal, uncorrelated Gaussian methods:

$$\sigma^2(t) = \left(\sum_i \frac{1}{\sigma_i^2} \right)^{-1}, \qquad \mu_i(t) = \left(\sum_i \frac{1}{\sigma_i^2} \right)^{-1} \sum_i \frac{x_i}{\sigma_i^2}$$

Provided we're only looking at data from a DRW curve, this gives a quick and easy estimate of the distribution that is conservatively imprecise.



Crude Interpolation for Fake Data

A MORE ROBUST METHOD FROM RYBICKI AND ZU

The DCF is useful in how little it assumes about the signal, but becomes a disadvantage in cases where we *do* have a bit more knowledge. (Rybicki & Press, 1992) presents a method that uses our understanding of the DRW nature of the AGN continuum to not only try and determine the correlation function, but to recover the underlying light curves.

CONSTRUCTING THE CONTINUUM CURVE

Rybicki presents a for using known, or at least assumed, parameters for the DRW to try and recover the signal's behavior in between our measurements, a technique that can be used as a stepping stone to recover the correlation functions between the continuum and response signals *without* knowing these parameters.

First, consider the observed signal, 'y', to be the sum of the true signal, 's', and some random noise signal, 'n':

$$y = s + n + \bar{y}$$

With our best guess for the signal at any point being some linear sum of the measurements (after subtracting away the average):

$$s_* = \sum_i d_i y_i'$$

If the measurements are discrete, the two components of the observation have their own covariance matrices:

$$S, N$$

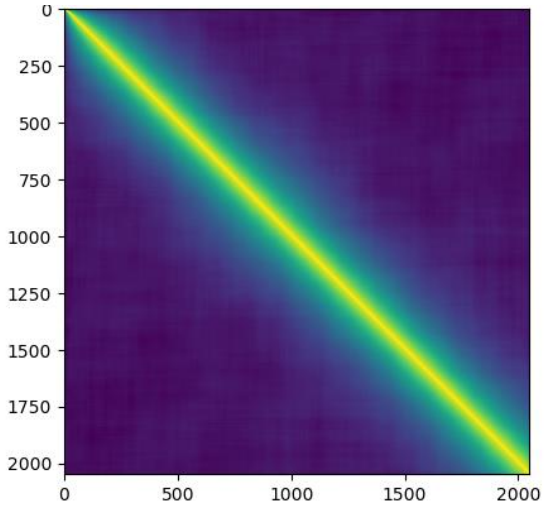
The elements of S are the covariances between our different measurements:

$$S_{ij} = \langle s_i, s_j \rangle$$

If, for example, we're looking **only** at a DRW, the matrix 'S' would have elements:

$$S = \sigma_\infty^2 \cdot \exp\left(-\frac{|t_i - t_j'|}{\tau_d}\right)$$

This produces a characteristic "exponential covariance matrix" that has some conveniently efficient methods of inverting (right).



Very Large Exponential Covariance Matrix

Provided we've used a known transfer function to find the covariances between points that aren't the continuum, we can populate elements that don't correspond to continuum-continuum measurement pairings. In this way, the 'S' matrix encodes all of our knowledge about the underlying physical system.

Meanwhile the (uncorrelated) measurement errors would give a diagonal matrix with:

$$N_{ii} = E_i$$

Under these conditions, our best estimate for the baseline is the normal result from Gaussian statistics:

$$\bar{y} = \frac{L^T C^{-1} y}{L^T C^{-1} L}$$

We need to subtract this away from the data:

$$y' = y - L\bar{y}$$

The argument then goes that, in estimating the signal at some time 's(t)', we should weight each value $y' = y - \bar{y}$ by:

- The inverse of variance of that measurement; and
- How strongly the thing we're estimating correlates with the measurement

If $\vec{S}_*(t)$ is a vector of covariance between the point we're estimating and the signal at the measurements:

$$\vec{S}_* = \sigma_\infty^2 \begin{pmatrix} \exp\left(-\frac{|t - t_1|}{\tau_d}\right) \\ \exp\left(-\frac{|t - t_2|}{\tau_d}\right) \\ \vdots \end{pmatrix}$$

We can get that kind of weighting with a best estimate:

$$s(t) = \vec{S}_*(t) \cdot C^{-1} y'$$

Alternately written:

$$s(t) = \vec{S}_*(t) \cdot C^{-1}(y - \bar{y}) + \bar{y}$$

As for the uncertainty in this estimate, it's found by finding the component of the variance that is orthogonal (can't be accounted for) with the data:

$$\Delta s_*^2 = \langle s_*^2 \rangle - \vec{S}_* \cdot C^{-1} \vec{S}_*$$

Which, because $\langle s_*^2 \rangle = \sigma_\infty^2$, means:

$$\Delta s_* = \sigma_\infty^2 - \vec{S}_* \cdot C^{-1} \vec{S}_*$$

From: (William, Rybicki, & Hewitt, 1992)

EXTENDING TO MULTIPLE POINTS

To summarise Rybicki's method for a single point:

$$s(t) = \vec{S}_*(t) \cdot C^{-1}(y - \bar{y}) + \bar{y}$$

$$\Delta s_*^2(t) = \langle s_*^2 \rangle - \vec{S}_* \cdot C^{-1} \vec{S}_*$$

Where:

$$\bar{y} = \frac{L^T C^{-1} y}{L^T C^{-1} L}$$

You may notice that all of these are linear operations, and we can pretty easily generalize this to estimate multiple points at once.

- Instead of $\overrightarrow{S_*}$ being a vector, instead make it a matrix, S_{cd}, the signal covariance of the curve-data. We do this by placing multiple covariance "vectors" side by side as columns.
- Similarly, $\langle s_*^2 \rangle$ needs to be replaced with a matrix-like object, S_{cc}, constructed the same as the normal 'S' matrix from earlier, but using the curve-times instead of the data-times
- For clarity, we'll rename the old covariance matrices with subscripts 'dd' to indicate data-data covariance matrices.

So, if we want to estimate an entire curve at once, we'd use:

$$\vec{s}_c = S_{cd}^T C_{dd}^{-1} (y - \bar{y}) 3$$
$$\Delta s_*^2 = S_{cc} - S_{cd}^T \, C_{dd}^{-1} \, S_{cd}$$

Where:

$$C_{dd} = S_{dd} + N_{dd}$$

Notice that this returns the mean and uncertainty of the stochastic component of the signal, 's', not the signal itself, 'y'. To recover that, you will need to add the baseline back on**:**

$$y_c = s_c + \bar{y}$$

USING RYBICKI'S METHOD TO REFINE DATA

A meaningful special case is when the signal points we're trying to predict **are** the measurements, i.e. when we're trying to refine the measurements. In this case, we have:

$$S_{dd} = S_{df} = S_{ff} = S$$

Giving:

$$\vec{s}_c = S C^{-1} (y - \bar{y}) + \bar{y}$$
$$\Delta s_*^2 = S - S C^{-1} S$$

Where we've used the fact that, in this instance, $S^T = S$.

AN EXTENSION TO NON-CONSTANT BASELINES

One limitation of the method described is that it assumes that the noise and DRW of the underlying signal are varying about a fixed average, $\bar{y}$, and doesn't allow for any more complicated behaviour of the baseline. A more general extension that accounts for this is presented by Rybicki, and adopted by Zu in their Javelin program, and will be summarized here.

First, we describe the observed signal as being the superposition of the natural variations, the noise, and a linear combination of some other set of known (or guessed) basis functions:

$$y(t) = s(t) + n(t) + \sum q_i f_i(t)$$

Now we'll introduce a matrix, L, which encodes these basis functions. For example, suppose we assume the baseline to behave quadratically in time, and we're analysing only one curve, i.e.:

$$\sum q_i f_i(t) = q_1 + q_1 t + q_2 t^2$$

The L matrix would look like:

$$L^T = \begin{bmatrix} 1 & 1 & 1 & 1 & \dots \\ t_1 & t_2 & t_3 & t_4 & \dots \\ t_1^2 & t_2^2 & t_3^2 & t_4^2 & \dots \end{bmatrix}$$

If we've got data from more than one curve, we staple together matrices like above, but and leave the remaining elements blank. For example, suppose we have two curves, where we're assuming linear baseline behaviour for the first and constant for the second. In this case, our L matrix looks like:

$$L^T = \begin{bmatrix} 1 & 1 & \dots & 0 & 0 \\ t_1 & t_2 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 1 \end{bmatrix}$$

First, as before, we combine the DRW and noise covariances into a single matrix:

$$C = S + N$$

Then use this to estimate the linear coefficients of the baseline:

$$\hat{q} = (L^T C^{-1} L)^{-1} L^T C^{-1} y$$

Alternately written as:

$$\hat{q} = C_q L^T C^{-1} y$$

Where we've define the covariance matrix of these coefficients, C_q:

$$C_q = (L^T C^{-1} L)^{-1}$$

$$\langle \Delta \hat{q}^2 \rangle = C_q$$

You'll notice that, if we have 'L' just be a vector of 1's, this is the same average estimate from the simple method.

Meanwhile, the estimated light curve is given by the same method as before, just subtracting away the moving baseline instead of a fixed average:

$$\hat{s} = SC^{-1}(y - L\hat{q})$$

With variance:

$$\langle \Delta\hat{s}^2 \rangle = \vec{S}_* - \vec{S}_* \cdot C_\perp S$$

Where $C_\perp$ is the portion of the covariance matrix that is orthogonal (i.e. can't be accounted for) by the moving baseline.

$$C_\perp^{-1} = C^{-1} - C^{-1} L C_q L^T C^{-1}$$

As presented, this method is used to refine the measurements at the times we observe them, however we can pretty easily extend this to generate any arbitrary time series in the same way as before:

$$\vec{s}_c = S_{cd}^T C_{dd}^{-1}(y - L\hat{q})$$
$$\Delta s_*^2 = S_{cc} - S_{cd}^T \, C_{dd}^{-1} \, S_{cd}$$

Where:

$$C_{dd} = S_{dd} + N_{dd}$$

The uncertainties in these estimates come from the **diagonal elements** of the matrix Δs_*^2.

Notice that, as with the other methods, this returns the mean and uncertainty of the stochastic component of the signal, 's', not the signal itself, 'y'. To recover that, you will need to **add the baseline evolution back in:**

$$y_c = s_c + \sum q_i f_i(t)$$

This is only really useful as a diagnostic to compare the predictions directly to the observation.

> From: (Zu, Kochanek, & Peterson, An Alternative Approach to Measuring Reverberation Lags in Active Galactic Nuclei, 2011)

USING RYBICKI & ZU'S METHODS TO RECOVER CORRELATION FUNCTIONS

Looking at the continuum reconstruction method, we can see that we encode all of our model parameters in the 'S' matrix. The continuum-continuum elements require us to known σ_∞ and τ_d, and all of the other elements require that we know something about the transfer function.

We can use some principles of Gaussian statistics to say that the probability of a particular signal 's' being true is proportional to:

$$P(\boldsymbol{s}) \propto |S|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{s}^T S^{-1}\boldsymbol{s}\right)$$

And, similarly, the probability of a particular noise line, **n,** being true is:

$$P(\boldsymbol{s}) \propto |N|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{n}^T N^{-1}\boldsymbol{n}\right)$$

Where S and N are the **covariance matrices** of **s** and **n**. Thus, the probability of a particular signal realization is:

$$P(y|\,\boldsymbol{s},\boldsymbol{q},\boldsymbol{n}) \propto |SN|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}[\boldsymbol{s}^T S^{-1}\boldsymbol{s} + \boldsymbol{n}^T N^{-1}\boldsymbol{n}]\right)$$

But we don't care about a "particular" realisation, we care about how well the ensemble of possible realisation implied by a particular set of parameters is, i.e. ,marginalizing over the variances we calculated in the previous section. Doing so gives us the probability that a particular set of parameters corresponds to our measurements:

$$P(y|\boldsymbol{p}) \propto |SN|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left[\Delta s^T(\mathrm{S}^{-1} + N^{-1})\Delta s + \Delta q^t C_q^{-1}\Delta C_q + y^T C_\perp^{-1} y\right]\right)$$

So, we have a single number telling us how good of a job a given choice of parameters, '**p**', does at describing our measurements, 'y'. From here, we marginalize over the linear parameters (i.e. q over the interval Δq) to arrive at:

$$P(y|\boldsymbol{p}) \propto |S+N|^{-\frac{1}{2}} |L^T C^{-1} L|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} y^T C_\perp^{-1} y\right)$$

Alternately written as:

$$P(y|\boldsymbol{p}) \propto \mathcal{L} = \frac{\exp\left(-\frac{1}{2} y^T C_\perp^{-1} y\right)}{\sqrt{|C| \times |L^T C^{-1} L|}}$$

This gives us an engine for turning a set of model parameters, i.e. the damping timescale, natural DRW variance and the transfer function width and characteristic delay. From here, trying to get the best fits for these parameters becomes a numerical optimization problem, which Zu approaches with a mix of Nelder-Mead and grid-interpolation.

If we trial a sufficient number of datapoints, we then also have the option of marginalizing over the parameters we're not concerned about to build up a decent estimate of the correlation function as a whole.

> From: (Zu, Kochanek, & Peterson, An Alternative Approach to Measuring Reverberation Lags in Active Galactic Nuclei, 2011)

VALDIATION OF CURVE GENERATION METHODS

Before moving forward to estimating correlation functions and estimating delays, it's worthwhile taking a moment to see how well these continuum methods actually work at recovering light curves from data.

As a part of the attached python program is a monte-carlo module that generates many DRW's of known parameters to "brute force" the distribution of possible light curves for a given set of data. This process is unreasonably expensive to use in general, but does give us a reliable curve to validate our faster methods against. This section will outline how the Monte Carlo data is generated, and use this data to quickly compare the curve estimate models.

GENERATING MONTE-CARLO CURVES

The Monte Carlo curves that we're comparing against were generated only for DRW's of a constant baseline, using the following process.:

Beginning with some set of (faked) signal measurements with known times and uncertainties, and set DRW parameters τ and σ_∞:

1. Generate a single DRW realization
2. Offset that curve by some baseline (see further down)
3. Evaluate the χ^2 score of the realisation against the data
4. Generate a random number 'r' on (0,1)
5. If $r < \exp\left(-\frac{1}{2}\chi^2\right)$, add keep that realization to the ensemble, else discard
6. Repeat steps 1-5 until a sufficiently large ensemble has been built up
7. Calculate the evolving mean and variance of the ensemble and save

The χ^2 score is generated by:

$$\chi^2 = \sum_i \left(\frac{y_i - f(t_i)}{E_i}\right)^2$$

Where y_i is the ith signal measurement, at time t_i and with (1 standard deviation) uncertainty E_i, and $f(t_i)$ is the DRW value at that time.

As for estimating the baseline, three methods were used:

1. Setting $\bar{y} = 0$, representing cases where the baseline was already known beforehand
2. Generating a purely random value for $\bar{y}$ from a sufficiently broad range
3. Using the value that minimizes the χ^2 for that particular realization, i.e. using:

$$\bar{y} = \frac{\sum_i \frac{y_i}{E_i^2}}{\sum_i \frac{1}{E_i^2}}$$

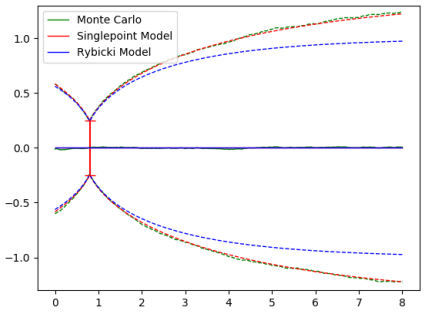
As it happens, the last two of these generate virtually identical results, with the second being much faster in a computational sense due to each realization maximizing its chance of being used in the final ensemble.

VALIDATING ESTIMATE MODELS

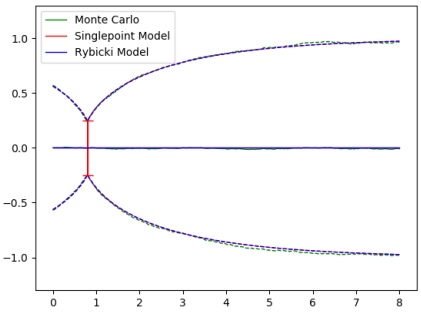
Now that we have (computational expensive) numerical data, we have some result to compare our data to. Note that the following validation extends **only** to DRW's of constant baselines (exponential covariance matrices)

For a Single Datpoint

Before even considering more complicated cases, it's worth validating against the simplest case: a single datapoint, the behaviour for which we have analytical models for (the 'singlepoint' model). When we test the Rybicki method against these, we can see a perfect agreement between all three models when the baseline is fixed (right), i.e. all the DRW's in the monte carlo ensemble are generated with $\bar{y} = 0$, but some disagreement arises when this isn't the case (left).



DRW Curve Recovery Methods for Unknown Baseline

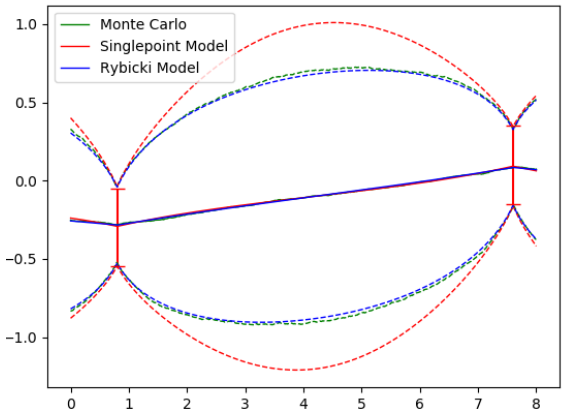


DRW Curve Recovery Methods for Known Baseline

The Rybicki method over estimates the confidence in the baseline, thereby under-estimating the variance in the overall distribution.

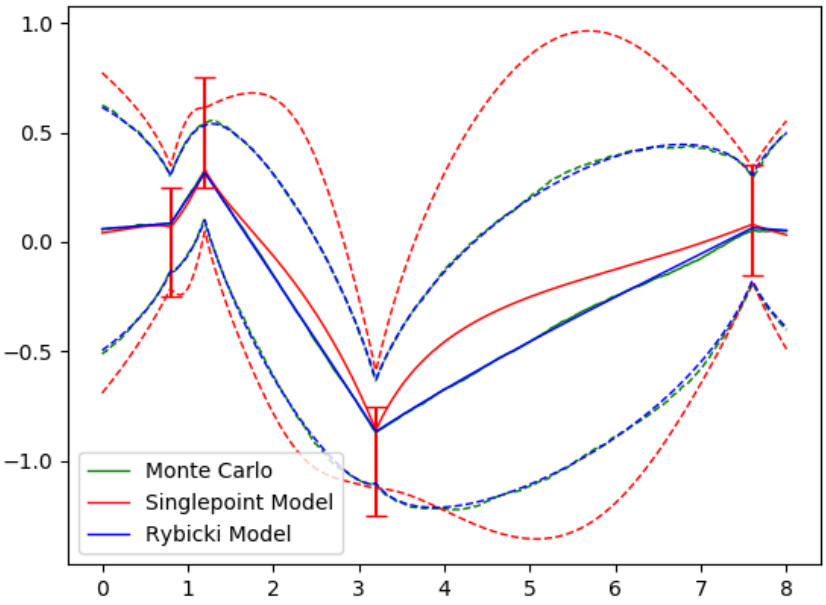
Fortunately, this issue disappears once we have enough data to properly constrain the measurement: with even two datapoints, the rybicki method comes into close agreement with the monte carlo curves.

This example also illustrates the use of the singlepoint method as a quick and cheap estimate of the outside behaviour of the curve, sacrificing resolution for ease of calculation.



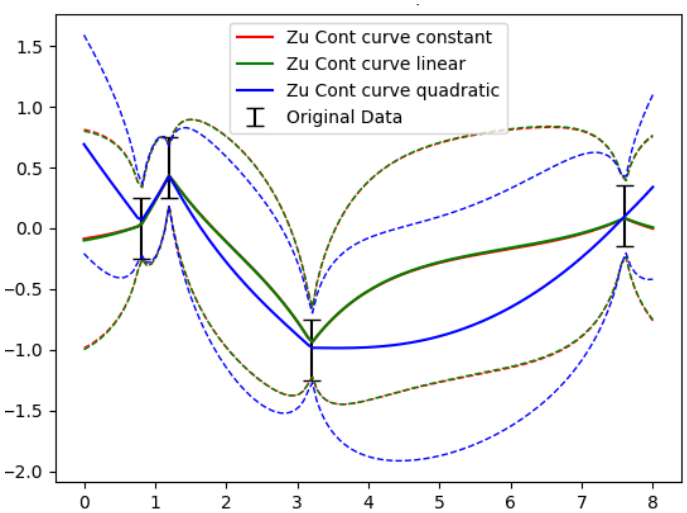
DRW Curve Recovery Methods For Two Datapoints

The more data we have available, the better the Rybicki and Zu methods perform. As shown below, the robust method matches almost perfectly to the monte-carlo results for multiple datapoints, drastically outperforming any cruder alternatives.



DRW Curve Recovery Methods For Multiple Datapoints

Something that has to be considered in the Rybicki/Zu methods is the affect of different baseline basis functions: changing the elements of the L matrix can drastically alter the final curve distribution. Particular care should be taken not to overfit the data with an overly complex model, as this may obscure the random variations that we're trying to isolate.



DRW Curve Recovery with Multiple Baseline Behaviour Types

REFERENCES

Edelson, R. A., & Krolik, J. H. (1988). *The Discrete Correlation Function: A New Method For Analyzing Unvevenly Sampled Variability Data.* Center for Astronomy and Space Astrophysics, University of Colorado.

Ivezic, Z., & Macleod, C. (2014). *Optical variability of Quasars: A Damped Random Walk.* Seattle, WA: Department of Astronomy, University of Washington.

MacLeod C L, E. A. (2010). *Modelling the Time Variability of the SDSS Stripe 82 Quasars as a Damped Random Walk.* Seattle, WA: Department of Astronomy, University of Washington.

Peterson, B. M., & Horne, K. (2004). *Reverberation Mapping of Active Galactic.* Columbus, OH: Department of Astronomy, The Ohio State University.

Rybicki, G. B., & Press, W. H. (1992). *Interpolation, Realization and Reconstruction of Noisy, Irregularly Sampled Data.* Cambridge, MA: Harvard-Smithsonian Center for Astrophysics.

William, P. H., Rybicki, G. B., & Hewitt, J. H. (1992). *The Time Delay of Gravitational Lens 0957+561. I. Methodology And Analysis of Optical Photometric Data.* Cambridge, Ma: Harvard-Smithsonian Center for Astrophysics.

Zu, Y., Kochanek, C. S., & Peterson, B. M. (2011). *An Alternative Approach to Measuring Reverberation Lags in Active Galactic Nuclei.* Columbus, OH: Ohio State University.

Zu, Y., Kochanek, C. S., Kozlowksi, S., & Udalski, A. (2013). *Is Quasar Optical Variability a Damped Random Walk?* Columbus, OH: Department of Astronomy, The Ohio State University.