Problem 2. Write up your Project Proposal with the following sections. The result should be 2-3 pages long.

2a. *Goal*: What is your primary project goal? What you would like to learn?

2b. *System*: Describe how the dynamical system is nonlinear and time-dependent.

What's the state space?

What's the dynamic?

Why is the system behavior interesting?

2c. *Dynamical properties*: What dynamical properties are you going to investigate?

2d. *Intrinsic computation properties*: What information processing properties are you going to investigate?

2e. *Methods*: What methods will you use? Why are they appropriate?

2f. *Hypothesis: What is your current guess as to what you will find?
</p>

2g. *Steps*: List the appropriate steps for your investigation; for example, read literature, write simulator, do mathematical analysis, estimate properties from simulation, write up report, and so on.

2h. *Time*: Estimate how long each step will take. Can you complete the project within one month?

(2a)

I'm working on a dynamic model of phylogenetic space: What does the tree of life look like as it grows? The model consists of a system of Galton-Watson trees contracted on the unit square. The system is a random multifractal, and mathematically tractable as a non-deterministic IFS. Graphic 1 gives a visualization of the system (excuse the size and frame rate -- it is a bit dizzying).

In [4]:
```
import matplotlib.pyplot as plt
from matplotlib import animation
from IPython.display import display, Image

from IPython.display import Image
with open('Random_Multifractal_GW_Zoom.gif','rb') as file:
    display(Image(file.read()), format='gif')
```

Out[4]:

In this model, downscale trees are the convex hulls of upscale nodes, and we have stochastically occurring transitions along system paths (Figure 1).

In [9]:
```
from IPython.display import Image
with open('Convex_Hull.png','rb') as file:
    display(Image(file.read()), format='png')

print "Figure 1 Transition region. Scale contraction expands nodes to trees."
```

Out[9]:

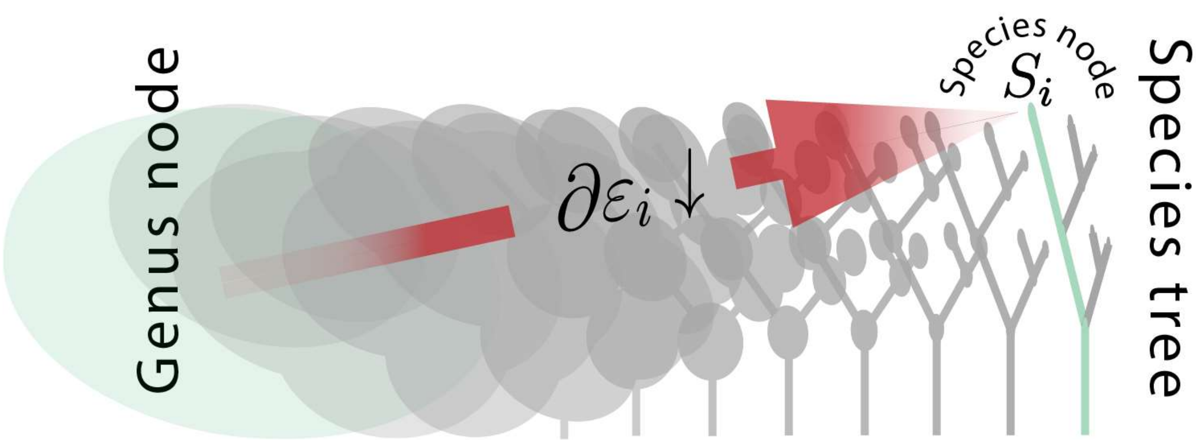


Figure 1 Transition region. Scale contraction expands nodes to trees.

I am using a slightly modified version of Laurent Nottale's scale dilation equation to return branch lengths within the system. His equation is:

$$\mathcal{L} = \mathcal{L}_0\{1 \pm \left(\frac{\lambda}{\varepsilon}\right)^{D_F - D_T}\}$$

In my model branch lengths give time intervals, so I choose to write the scale dilation equation as:

$$\frac{\Delta t_i}{\Delta t_j} = \{1 \pm \left(\frac{\lambda_{ij}}{d_\varepsilon(i,j)}\right)^{D_F - D_T}\}$$

Where Δt_i is elapsed time for organism i, Δt_j is elapsed time for organism j, $d_\varepsilon(i,j)$ is the scale distance between species i and j, λ_{ij} is the scale location of the transition between species i and j, D_F is the fractal dimension of the tree system in the region containing i and j, and D_T is the topological dimension in the region containing i and j.

However, since the dynamic tree of life is of interest, differential form is preferred:

$$\frac{\partial t_i}{\partial t_j} = \{1 \pm \left(\frac{\lambda_{ij}}{d_\varepsilon(i,j)}\right)^{D_F - D_T}\}$$

My primary goal is to apply the information measures to the regions where nodes transition to trees. Here, the topological dimension D_F goes from 0 (that of a point) to 1 (that of a line). I am hoping that the information measures give me a way of systematically calculating and comparing λ_{ij} across the system. It seems possible that the information measures may provide a way to calculate the fractal dimension spectrum and so obtain scale dependent values of D_F. Currently, I am using a random walk procedure to try and estimate this spectrum, but I am open to an information measure approach if it proves better.

(2b)

The model is set up so that the path dependent scale variable ε_i represents biological form. The quantity of interest in the dynamic case is the change in biological form along path i with respect to the change in time along path i as measured from species j:

$$\left(\frac{\partial \varepsilon_i}{\partial t_i}\right)^j \equiv \left(\frac{\partial \varepsilon}{\partial t}\right)_i^j$$

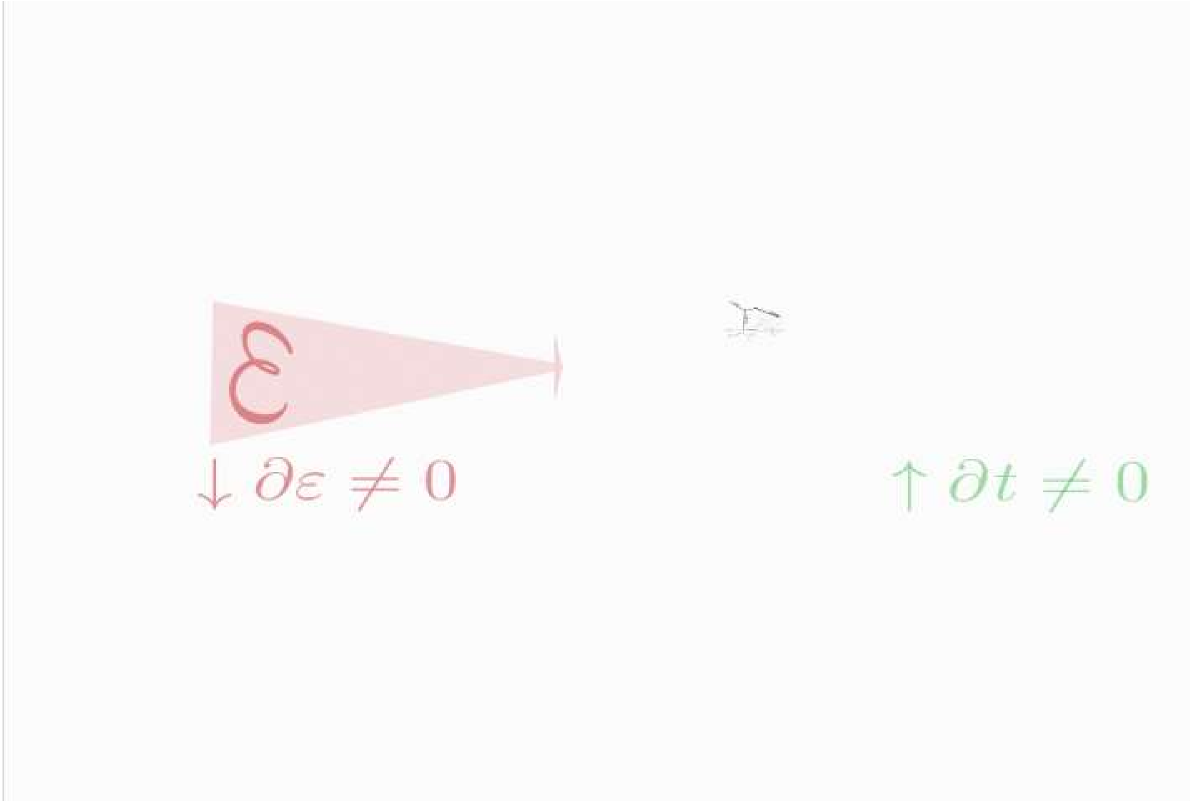
The system is extremely nonlinear since every quantity in the scale dilation equation is a function of both ε_i and ε_j (i.e., a function of the species type of organism i and organism j respectively). This means $\left(\frac{\partial \varepsilon}{\partial t}\right)_i^j$ is also a function of both ε_i and ε_j.

I don't currently have these variables illustrated for the system of Galton-Watson trees, but gif 2 illustrates the dynamic $\frac{\partial \varepsilon_i}{\partial t_i}$ for a hypothetical tree of life.

In [11]:
```
from IPython.display import Image
with open('Zoom_Published_Phylogenies_Path_Highlighted.gif','rb') as file:
    display(Image(file.read()), format='gif')

print "Gif 2: Scale contraction and time expansion along a path. For conceptua
l illustration only."
```

Out[11]:




```
Gif 2: Scale contraction and time expansion along a path. For conceptual illu
stration only.
```

The question of state space is not a simple one to answer, and that is to me the most interesting part of this system. The state space is every unit of biological form. So states are topologically contained within states. But even more, due to the multifractal structure, whether any given state i exists depends both on ε_i and ε_j (i.e., on the species of the organism being observed and on the species of the organism doing the observing). So scale space changes depending on which species is considered the observer.

It seems the state space is finite but I hesitate even to posit that (Are there infinitely many descriptions of Earth's biological system over the entirety of its history?). The state space is astronomical in size though, that much is certain. In terms of the Galton-Watson model, most simulations have on the order of a few hundred states after four levels.

(2c)

I'm investigating how λ_{ij} behaves under the dynamic $\left(\frac{\partial\varepsilon}{\partial t}\right)_i^j$.

(2d)

I am hoping that studying the intrinsic computation properties of the node-tree transition region will present a meaningful way to obtain parameter estimates of λ_{ij} for the system of Galton-Watson trees introduced above.

(2e)

There are three random variables along every path in the multifractal Galton-Watson tree system above. They are:

(1) Offspring (2) Generations (3) Scale contraction

Along each path in the system, each random variable can be taken as Markovian. So the number of possible offspring for a current node depends on the number of possible offspring of its parent node, and likewise for the number of possible generations. For scale contraction, we must have that a tree birthed from a node is at a scale less than the node that birthed it. Within these boundaries set by the Markov conditions, we can sample from a uniform distribution or any distribution we like.

The stochasticity of the system makes the information measures approach appropriate, and the Markov condition makes the approach more easily executed.

(2f)

I think I will find that the information measures give a region in which λ_{ij} is well defined and nonzero. Over these regions fractal effects have to be considered when measuring time intervals. The information measures will then indicate from which species a given species i exists, and from which species i does not exist. Assuming we have estimates for the other variables in the scale dilation equation, we can estimate the dynamic quantity $\left(\frac{\partial \varepsilon}{\partial t}\right)_i^j$

(2g)

1) Read a few papers on Galton-Watson branching as a Markov process. 2) Review information measures and descriptively apply the measures to the model. 3) Code the determination of information measures from simulations, using CMPY package when possible. 4) Gather and analyze data from simulations. 5) Write report.

(2h)

1) Read a few papers on Galton-Watson branching as a Markov process. (~ 1 week) 2) Review information measures and descriptively apply the measures to the model. (~ 1/2 week) 3) Code the determination of information measures from simulations, using CMPY package when possible. (~ 1 week) 4) Gather and analyze data from simulations. (~1 week) 5) Write report. (~ 1/2 week)

It seems the project is doable within one month.

In [0]:

```
In [0]:
```