Modeling The Dynamics of Large Molecules

Brenton LeMesurier http://blogs.cofc.edu/lemesurierb/

George Mason University, July 30, 2012

Contents

1	Introduction	1
2	Some Physics (Classical Mechanics Only)	1
3	Nearest-Neighbor Interaction Approximation	2
4	Approximating By Linear Forces (Hooke's Law Springs)	4
5	Continuum Limits: PDE Models	5
6	The Fermi-Pasta-Tsingau-Ulam Equations	7
7	Continuum Limits For The FPTU System	8

1 Introduction

Mathematical studies of large molecules such as proteins and DNA can not be done by solving exact quantum mechanical equations.

Instead a variety of approximations are used to get more manageable systems of ODE's.

Sometimes, a second level of simplification is to approximate a large "chain" of ODE's by a single PDE.

I will describe a few of these derivations, and with the help of numerical experiments, show how such approximations make useful predictions about the dynamics of large molecules ... if handled with care.

To narrow this huge field, I will focus on examples of vibrations and wave motion in protein molecules, while introducing a few techniques that are useful in many other situations.

2 Some Physics (Classical Mechanics Only)

Conservative Forces and Potentials

One key to molecular modeling is that the only significant forces are electrostatic, and so are a collection of energy conserving interactions between each pair of particles. Which such forces, the interaction of two charged particles (atomic nuclei, electrons, etc) with position vectors \mathbf{x}_i and \mathbf{x}_j is described by a potential energy

 $V_{i,j}(r_{i,j})$

where $r_{i,j} = \|\mathbf{x}_i - \mathbf{x}_j\|$ is the distance between them.

Then the force on particle i due to the effect of particle j is

$$F_{i,j} = -\nabla_{\mathbf{x}_i} V_{i,j} = -V'_{i,j}(r_{i,j}) \frac{\mathbf{x}_i - \mathbf{x}_j}{r_{i,j}}.$$

Molecules as Chains of Particles

Many biologically important molecules such as DNA and many proteins are un-branched chains of molecular subunits: the subunits are nucleic acid residues in DNA, amino acid residues in proteins.

These subunits are mostly in minimum energy quantum mechanical bound states, so when modeling the dynamics of molecules, their internal details can be ignored and they can be treated as "particles" connected in a chain. The dynamic of this chain is determined by the longer range interactions between these new "particles", and is adequately described by classical physics of electrostatic interaction.

Thus our first model of a long biological molecule is a chain of N particles interacting through forces given by potentials.

Leonard-Jones Forces Between Molecular Subunits

The force between two of these molecular subunits is the net affect of attraction and repulsion between their various electrons and nuclei, so it is not the simple inverse square law.

Instead, it becomes highly repulsive when the particles are very close (so that the electron repulsion is dominant), and is attractive and rapidly weakening when they are further apart.

One commonly used approximation for this interaction is

The Leonard-Jones potential, a.k.a "12-6"

$$V(r) = E_0 \left[\left(\frac{h}{r}\right)^{12} - 2\left(\frac{h}{r}\right)^6 \right]$$

Here, h is the equilibrium distance between the particles, where potential energy is at its minimum and there is no force because V'(h) = 0.

3 Nearest-Neighbor Interaction Approximation

Uniform Nearest-Neighbor Interaction Approximation

The rapid 6^{th} power decay rate leads to the approximation that particles only interact with their nearest neighbors along the chain.

A further simplification is to assume that the particles and interactions are the same, which is a good approximation when they are all amino acid resides in a protein, or all nucleic acids in DNA or RNA.

Thus the only potential we need is

$$V_{i,i+1} = V_{i+1,i} = V(|x_i - x_{i+1}|),$$

= $V(x_{i+1} - x_i)$ since the particles stay "in order".

The force on particle i is

$$F_{i,i+1} + F_{i,i-1} = V'(x_{i+1} - x_i) - V'(x_i - x_{i-1})$$

so its acceleration is given by

$$m_0 \frac{d^2 x_i}{dt^2} = V'(x_{i+1} - x_i) - V'(x_i - x_{i-1})$$

Analogy to a Chain of Masses and Springs

Thus in basic models large molecules behave like a chain of masses connected by springs, albeit with a force law different from the common Hooke's law:



Displacements as New Coordinates

Since only stretching of the distance between particles matters, it is convenient to measure positions via displacements from their rest positions, which are $x_i = uh$. So we use the new variables

$$y_i = x_i - ih.$$

The potential can be shifted to a function V(s) of the stretching $s = y_{i+1} - y_i$, with the rest position now at value s = 0.

Adding a constant to the potential has no effect on the forces, so we can have

$$V(0) = 0, V'(0) = 0, V''(0) > 0.$$

The equations of motion are now

$$m_0 \frac{d^2 y_i}{dt^2} = V'(y_{i+1} - y_i) - V'(y_i - y_{i-1})$$

Viewing Displacements as Vertical

With all the particles and motions in a single horizontal line, it can be hard to visualize the motion.

So it is convenient to interpret the same equations differently: treat the displacements y_i as vertical movements, perpendicular to the chain. Then we can also think of the horizontal positions as fixed, equally spaced *x*-coordinates $x_i = ih$.



This view will also help with the mathematics.

4 Approximating By Linear Forces (Hooke's Law Springs)

Approximating By Linear Forces Hooke's Law Springs

Even these simplified equations are very hard to solve, so the first studies (by the chemist Debye) made another simplifying assumption: if the energy is low enough, the stretching s stays small, so the only part of the potential that matters is near s = 0. The Taylor series approximation is

$$V(s) = V(0) + V'(0)s + \frac{1}{2}V''(0)s^2 + \frac{1}{6}V'''(0)s^3 + \cdots$$
$$= \frac{1}{2}ks^2 + \frac{1}{6}V'''(0)s^3 + \cdots \qquad (k > 0)$$

so for small s, we can use just the first non-zero term, and approximate with

$$V(s) = \frac{1}{2}ks^2.$$

This corresponds to a spring force of -V'(s) = -ks, the famous Hooke's Law.

Debye's Linear Mass-Spring Model

The equations with this small amplitude motion, and linear force approximation are

$$m_0 \frac{d^2 y_i}{dt^2} = k(y_{i+1} - y_i) - k(y_i - y_{i-1})$$

or

The Linear Mass-Spring Chain Model

$$\frac{d^2 y_i}{dt^2} = c^2 (y_{i+1} - 2y_i + y_{i-1}), \quad c = \sqrt{\frac{k}{m_0}}.$$

5 Continuum Limits: PDE Models

Continuum Limit of This Particle Chain

The picture and animation above resembles a vibrating string, except that here there is only a finite collection of x positions, not a continuous x variable in the equations.

But we already know how to solve the vibrating string equations, and *it is often far easier to work with a single partial differential equation than a large collection of ordinary differential equations.*

So let us look at yet another approximation: the **continuum limit** of this model.

Continuum Limits

The basic idea is to

• treat all the values $y_i(t)$ as approximated by values of a single function of continuous position variable x:

$$y_i(t) \approx y(x_i, t), \quad x_i = ih,$$

and then

• use Taylor polynomial approximations.

The big assumption here is that the quantities $y_i(t)$ vary slowly along the chain, so that this function y(t,x) fitting them can be a nice smooth, differentiable function, and thus the Taylor polynomial approximation makes sense.

This must not only be true for the initial data at time t = 0, (which we can maybe enforce), but the equations must keep it true.

Continuum Limits: Taylor Polynomial Approximation

Look at any position x_i and call this x, so that $x_{i\pm 1} = x \pm h$. The Taylor polynomial approximations for $y_{i\pm 1} \approx y(x \pm h)$ are

$$y_{i+1} \approx y + hy_x + \frac{h^2}{2}y_{xx} + \frac{h^3}{6}y_{xxx} + \frac{h^4}{24}y_{xxxx} + \cdots$$
$$y_{i-1} \approx y - hy_x + \frac{h^2}{2}y_{xx} - \frac{h^3}{6}y_{xxx} + \frac{h^4}{24}y_{xxxx} + \cdots$$

So the linear mass-spring equation

$$\frac{d^2 y_i}{dt^2} = c^2 (y_{i+1} - 2y_i + y_{i-1})$$

is approximated by

$$\frac{1}{c^2}y_{tt} = h^2 \left[y_{xx} + \frac{h^2}{12}y_{xxxx} + \cdots \right].$$

Continuum Limit: Time Rescaling

The factor h^2 throughout the right-hand side shows that the time scale is slow for small h, so we measure time in more appropriate units,

$$\tau = ht$$

Then

$$\frac{1}{c^2}y_{\tau\tau} = y_{xx} + \frac{h^2}{12}y_{xxxx} + \cdots$$

Continuum Limit: Ignoring All Small h Terms

If the spacing h is small, and the solution is nice enough that the higher derivatives like y_{xxxx} never get too big, we can approximate with the limit as $h \to 0$:

$$\frac{1}{c^2}y_{\tau\tau} = y_{xx}$$

which is the familiar wave equation, also used to describe a vibrating string.

For a chain of length L (endpoints at x = 0 and x = L) that is fixed at these endpoints, solutions to this are well known:

$$y = \sum_{j=1}^{\infty} \left[a_j \cos\left(\frac{c\pi j\tau}{L}\right) + b_j \sin\left(\frac{c\pi j\tau}{L}\right) \right] \sin\left(\frac{j\pi x}{L}\right),$$
$$= \sum_{j=1}^{\infty} \left[a_j \cos\left(\frac{c\pi jht}{L}\right) + b_j \sin\left(\frac{c\pi jht}{L}\right) \right] \sin\left(\frac{j\pi x}{L}\right)$$

Checking That This Approximation Works

In this very special case of linear equations, we can validate this PDE approximation by brute force: we can solve the mass-spring chain ODE system:

$$y_i = \sum_{j=1}^{\infty} \left[a_j \cos\left(\omega_j t\right) + b_j \sin\left(\omega_j t\right) \right] \sin\left(\frac{j\pi x_i}{L}\right),$$
$$\omega_j = c \sqrt{2\left(1 - \cos\frac{j\pi h}{L}\right)}.$$

In the approximation of small h so that $\frac{j\pi h}{L}$ is small, the Taylor polynomial approximation $\cos \theta \approx 1 - \frac{1}{2}\theta^2$ gives

$$\omega_j \approx \frac{c\pi jh}{L}$$

giving a match to the above solutions for the PDE,

$$y = \sum_{j=1}^{\infty} \left[a_j \cos\left(\frac{c\pi jht}{L}\right) + b_j \sin\left(\frac{c\pi jht}{L}\right) \right] \sin\left(\frac{j\pi x}{L}\right)$$

Checking That A Continuum Approximation Works, Without Explicit Solutions

However explicit solutions like these will not usually be possible when using with more accurate models that are not linear.

Instead a more generally useful validation is that the solutions of the PDE continue to vary slowly between successive values of x_i , confirming the assumption used to get this approximation.

Failure to Explain Thermalization, Part I

For Debye, these solutions made up of separate vibration modes were a disappointment: he was looking for an explanation of the physical phenomenon of *thermalization*, which means that in time, the energy of a system like a crystal is eventually evenly divided between all vibrational modes. That does not happen here: if the initial vibration has just the lowest frequency mode j = 1, that is all one ever gets.

(Debye started with a simple "one-dimensional crystal model", which is the same as our molecular chain.)

It is a problem for us too, since it suggests that the above solutions are not what really happens in a molecule.

Debye suggested that the strange outcome was an artifact of the approximation with linear equations, and this is at least partially right as we will see soon.

Unfortunately, it was not possible to solve a more accurate nonlinear model, so progress waited until computers were available (for other than military tasks) in the early 1950's.

6 The Fermi-Pasta-Tsingau-Ulam Equations

Enrico Fermi Revisits Debye's Suggestion

In the early 1950's Enrico Fermi was working at Los Alamos National Lab, and had the idea of using its early computer (the MANIAC I) to experiment with computations as a way to seek new scientific ideas, rather than just using them to apply known physical laws for quantitative simulations of specific physical processes.

He took up Debye's idea, and just went the next few steps in the Taylor polynomial approximation of the interaction potential:

$$V(s) = \frac{1}{2}ks^{2} + \frac{a}{3}s^{3} + \frac{b}{4}s^{4} + \dots \qquad (k > 0)$$

so that the force is given by

$$V'(s) = -ks - as^2 - bs^3 + \cdots$$

The Fermi-Pasta-Tsingau-Ulam Experiments

This lead to the first and still perhaps the most famous of computational experiments in mathematics.

Fermi worked with senior mathematician Stan Ulam and two young mathematicians who were then about the only people who could program this primitive early computer: Mary Tsingau and Joe Pasta. (Specifically, it was Mary Tsingau who devised the algorithm.)

They started with the simplest modification, with just the cubic potential energy term, so b = 0

$$m_0 \frac{d^2 y_i}{dt^2} = k[y_{i+1} - 2y_i + y_{i-1}] + a[(y_{i+1} - y_i)^2 - (y_i - y_{i-1})^2].$$

and various simple sine forms for the initial data.

Due to a historical quirk, this is usually known as the Fermi-Pasta-Ulam system or just FPU.

Numerical Solutions of the FPTU System

The results of their experiments can now be easily reproduced, for example with a free java applet from http://www.compadre.org/osp/items/detail.cfm?ID=6891 part of the Open Source Physics Project at http://www.compadre.org/osp/

Failure to Explain Thermalization, Part II Coherent Traveling Pulses Observed Instead

As the results from the applet suggest, the solutions still do not divide energy equally between all the modes (34 in this simulation; one for each position x_i .)

This surprising observation took about twenty years to explain, and one key step was again a continuum limit, leading to a PDE that, with considerable effort, was solved by Norm Zubusky, Martin Kruskal and others.

Today, I just want to show the careful steps that Zabusky and Kruskal took in 1965 to get the right continuum limit approximation.

I will also show one naive but wrong approach, as a cautionary tale.

7 Continuum Limits For The FPTU System

Continuum Limit of The FPTU System

Using the same approach as above, one first gets

$$\frac{1}{c^2}y_{\tau\tau} = y_{xx} + (2ah)y_xy_{xx} + \left(\frac{h^2}{12}\right)u_{xxxx} + \cdots$$

Next a common strategy can be used. For the basic wave equation $\frac{1}{c^2}y_{tt} = y_{xx}$, the "traveling wave" change of variables

$$\xi = x - ct, \quad u(\xi, t) = y(x, t)$$

gives the very simple equation

$$u_{tt} - 2cu_{t\xi} = 0$$

Some obvious solutions are ones that do not depend on t at all: $u(\xi, t) = f(\xi)$, so

$$y(x,t) = f(x-ct).$$

This gives a traveling wave: a constant "shape" given by function f, moving right at speed c.

The idea here is that: if a transformation converts one equation into a form that is easy to solve, the same transformation on a similar but more difficult equation might still simplify things.

Traveling Wave Transformation

Using a similar traveling wave change of variables on the above continuum limit equation

$$\frac{1}{c^2}y_{\tau\tau} = y_{xx} + (2ah)y_xy_{xx} + \left(\frac{h^2}{12}\right)u_{xxxx} + \cdots$$

gives:

$$u_{\tau\xi} - \left(\frac{ah}{2}\right)u_{\tau\tau} = -u_{\xi}u_{\xi\xi} - \left(\frac{h}{24a}\right)u_{\xi\xi\xi\xi} + \cdots$$

The delicate question is which of the infinite number of terms can we discard and still get a good approximation?

A First, Failed Attempt At a Continuum Limit

One naive idea is to do what worked before: assume that h is very small and so set h = 0. This gives

 $u_{\tau\xi} + u_{\xi}u_{\xi\xi} = 0.$

Introducing the new quantity $v = u_{\xi}$ gives

Burger's Equation

 $v_{\tau} + vv_{\xi} = 0.$

Shock Waves: The Continuum Limit Assumptions Fail

Unfortunately, the solutions of Burger's equation often do not stay slowly varying in position, even if the initial data is. In fact they often develop *shock fronts*, where all derivatives become infinite at a finite time:

Continuum Limit of The FPTU System, Done Right

The above problem is fixed by returning to the equation

$$u_{\tau\xi} - \left(\frac{ah}{2}\right)u_{\tau\tau} = -u_{\xi}u_{\xi\xi} - \left(\frac{h}{24a}\right)u_{\xi\xi\xi\xi} + \cdots$$

and keeping one more term: the fourth derivative one.

The argument is that both h and a are small, so the coefficient

$$\frac{ah}{2}$$

is small enough to neglect, but due to the ratio of small quantities, the term

$$\frac{h}{24a}u_{\xi\xi\xi\xi}$$

might not be, especially if $u_{\xi\xi\xi\xi}$ grows in time like it did with Burger's equation.

The Korteveg-de Vries Equation as a Molecular Model

This new continuum limit approximation is the famous

Korteveg - de Vries [KdV] Equation

$$v_\tau + vv_\xi + \frac{h}{24a}v_{\xi\xi\xi} = 0.$$

As discovered by Zabusky and Kruskal, this has robust traveling wave solutions: changing units (multiplying the space and time variables by constants) the solutions look like

$$v(x,t) = \frac{C}{2}\operatorname{sech}^{2}\left[\frac{\sqrt{C}}{2}(x-Ct-D)\right], \ C > 0, \ D \ \text{arbitrary}$$



the Korteveg - de Vries Traveling Wave

Traveling Waves in Long Molecules

What is important to us is that again, energy does not spread out equally over all vibration modes, because it can stay in the same shape, just changing position.

This helps to explain the surprising observations of the FPTU experiment.

Although this was disappointing to Debye's objective, it leads to some interesting ideas about molecular dynamics: traveling waves like this are potentially involved in the transfer of energy within protein molecules, and in the dynamics of DNA, such as its recoiling after getting straightened out during duplication.