Understanding in Computational Sciences

Imbunm Kim[*]

March 30, 2015

.

1 What is Computational Science(Scientific computing)

- *http://en.wikipedia.org/wiki/Computational_science*
- Interdisciplinary subject in nature involving physical and/or biological, social sciences, mathematics, and computer science.
- The field of study concerned with constructing mathematical models and quantitative analysis techniques and using computers to analyze and solve scientific problems.
- In practical use, it is typically the application of computer simulation and other forms of computation to problems in various scientific disciplines.
- The approach is to gain understanding, mainly through the analysis of mathematical models implemented on computers.

1.1 Motivation

How would you describe

- The flickering of a flame
- The texture of an oil painting
- Highway traffic during a rush hour
- Twinkling stars
- Breaking glasses
- The flight of a paper airplane
- The sound of a violin

[*] Department of Mathematics, Seoul National University, Seoul 151-747,Korea

- The diffusion of opinions
- Search engines (google, facebook)

Answers

- Analytical approach -theory
- Numerical approach -computing
- Observational approach -experiment

1.2 Science investigation

- Theory
- Experiment
- Computing:
  - Can often stimulate the insight and understanding that theory and experiment alone cannot achieve. With computers, scientists can study problems that previously would have been too difficult, time consuming, or hazardous; and, virtually instantaneously, they can share their data.
  - Stimulates insight and understanding
  - Sometimes leads to new theories, suggests new experiments
  - Is used to test new theories

Examples

- Spread of disease - AIDS
- Weather prediction
- Analysis of Space Shuttle disaster
- Air-quality model
- Creation of a galaxy simulation
- Potential for earthquake damage
- Modeling heart for treatment

1.3 Modeling process

Application of methods to analyze complex, real-world problems in order to predict what might happen with some course of action.
In order to develop modeling, we need some information: knowledge of natural laws, economical or social laws, as well as scientific data, i.e.

1. Numerical data describing real phenomena
2. Data collected according to certain strict rules

Model Classifications

- Specific vs general
- Stochastic(probabilistic) vs deterministic
  stochastic processes:share prices vs differential equations: Kepler's law of planet movement
  A system exhibits probabilistic or stochastic behavior if an element of chance exists. Otherwise, it exhibits deterministic behavior. A probabilistic or stochastic model exhibits random effects, while a deterministic model does not.

  Sources of random behaviour

  1. Sensitivity to or randomness of initial conditions: weather forecast, the path of a hurricane is probabilistic.
  2. Incomplete information:economic modeling

  MC(Monte Carlo method) -useful numerical methods for solving probabilistic or stochastic model.

- Microscopic vs macroscopic
  cell vs body in human, nano scale vs m scale,
  position, velocity for each vehicle vs a mathematical model that formulates the relationships among traffic flow characteristics like density, flow, mean speed of a traffic stream, etc. Such models are conventionally arrived at by integrating microscopic traffic flow models and converting the single-entity level characteristics to comparable system level characteristics in traffic flow
- Discrete (variable changes in incremental steps)vs continuous (variable changes continuously)
- Qualitative vs quantitative
- Dynamic (changes with time) vs static (does not consider time)
- Numerical vs analytical

- Model estimation(inverse problem) vs first principles models(forward problem)
  e.g.) forward pb : for given initial data and given parameter values, find the solution that obeys certain equations.
  inverse pb : for given solution or observed data, find the initial data and unknown parameter values by minimizing the error btw data and numerical solution.

Steps of developing modeling Process: Cyclic

- 1. Analyze the problem
  Determine problem's objective, decide on the pb's classification(deterministic or stochastic).

- 2. Formulate a model
  Design the model, forming an abstraction of the system we are modeling.

  1. Gather data
     Collect relevant data to gain information about system's behaviour.
  2. Make simplifying assumptions and document them
     As simple as possible, ignoring factors that do not seem as important.
  3. Determine variables and units
     Decide independent and dependent variables, in most cases, time is an independent var.
  4. Establish relationships among variables and submodels
     If possible, draw a diagram of model, break it into submodels and indicate relations among variables.
  5. Determine equations and functions

- 3. Solve model
  Some techniques and tools include algebra, calculus, graphs, computer programs and computer packages. If the model is too complex to solve, return to Step 2 to make additional simplifying assumptions or to Step 1 to reformulate the pb.

- 4. Verify and interpret model's solution
  Once we have a solution, we should carefully examine the results to make sure that they make sense (verification) and the solution solves the original problem (validation) and is usable.
  Verification determines if solution works correctly, solving the problem correctly
  Validation establishes if the system satisfies the problem's requirements, thus solving the right problem

- 5. Report on the model

  1. Analysis of problem
  2. Model design
  3. Model solution

4. Results and conclusions

- 6. Maintain model
  Necessary or desirable to make corrections, improvements, or enhancements

e.g., simplifications made in underlying model :

- Nonlinear becomes linear
  Hooke's law in mechanics: Displacement is profortional to force
  Ohm's law in networks : Current is profortional to voltage difference
  Scaling law in economics: Output is profortional to input
  Linear regression in statistics : A straight line or a hyperplane can fit the data
- Continuous becomes discrete
- Multidimensional becomes one-dimensional
- Variable coefficients become constants

1.4 Errors

- Data error-equipment error, calibration error, misread error, record incorrectly, etc.
- Modeling error-simplifying assumptions, determines incorrect eqs that cause the models results to deviate from reality.
  Lord Kelvin computed the age of the earth as between 20 and 40 million yrs old based on the assumption that the earth was cooling from molten mass with the sun being its only source of energy. (In reality, decaying radioactive elements in the earth's crust also generate heat that helps to maintain the temp of the planet, 12 billion yrs) In 1896, it was corrected.
- Implementation Errors
  In 1999, NASA's Mars Climate Orbiter was lost because the builder of the spacecraft programmed it to use English units(pounds, feet) and NASA's Jet propulsion scientists employed metric units(newtons, meters).
- Precision : the number of significant digits
  computer calculations-a float point number, significant digits, precision
  single precision number, is about 6 or 7 decimal digits, while the magnitude ranges from about 10^{-38} to 10^{38}.
  double precision number, is about 14 or 15 decimal digits, while the magnitude ranges from about 10^{-308} to 10^{308}.
- Absolute and Relative Errors
  $absoluteerror = |exact - approximated|$
  $relativeerror = \frac{absoluteerror}{exact}$

- Round-off Error Problem of not having enough bits to store entire floating point number(round up, round down) and approximating the result to the nearest number that can be represented
- Overflow and Underflow Overflow - error condition that occurs when not enough bits to express value in computer Underflow - error condition that occurs when result of computation is too small for computer to represent
- Arithmetic Errors : Algebra rules do not necessarily hold

  Suppose the machine rounds to three significant digits. Then calculate $\left(\frac{x}{y}\right) z$ and $\frac{xz}{y}$ with $x = 2.41$, $y = 9.75$, $z = 1.54$
  x/y = 0.247179 rounds to 0.247, (x/y)z=0.247*1.54 = 0.38038, after rounding, 0.380
  xz = 3.7114, 3.71 in rounding form, xz/y = 3.71/9.75 = 0.380513, 3.81 after rounding.
- Error Propagation
  Repeatedly executing $t = t + dt$, better to repeatedly increment i and calculate $t = i * dt$
  In looping, whenever possible, avoid accumulating floating pt values through repeated addition or subtraction
- Truncation Error
  Error that occurs when a truncated, or finite, sum is used as approximation for sum of an infinite series
- To reduce numerical errors
  Use maximum number of significant digits(multiple precision),
  If there are big differences in the magnitude of numbers, add from the smallest to largest numbers

2 Analytical Models for ODEs

2.1 Ordinary Differential and Difference Equations

2.1.1 Several types of equations

- Equation with unknown x

$$ax + b = 0$$

- Function equation with unknown function $u(x)$

$$u(x + y) = u(x)u(y) \quad \forall x, y \in \mathbb{R}.$$

- Differential equation with unknown function $u(x)$

$$\frac{du}{dt} = a(t)u + f(t, u)$$

- Integral equation with unknown function $u(x)$

$$u(t) = \int_0^t a(s)u(s)ds + f(t,u)$$

- Integro-differential equation with unknown function $u(x)$

$$\frac{du}{dt} + u(t) = \int_0^t a(t-s)u(s)ds + f(t,u)$$

2.1.2 Class of DEs

- Order of DEs

  If a differential equation contains the nth maximum order of differentiation of unknown function, we call the DE is of "nth order," or the equation is an nth-order DE. Thus, an nth order DE can be given in general as follows:

$$\sum_{k=0}^{n} a_k \left(t, u, \cdots, \frac{d^k u}{dt^k}\right) \frac{d^k u}{dt^k} = f(t), \tag{2.1}$$

  where $a_k \left(t, u, \cdots, \frac{d^k u}{dt^k}\right)$ is called the coefficient of the kth order derivative of unknown $u(t)$ for $k = 0, \cdots, n$. Here, n is the order of the DE.

- Linear vs. nonlinear DEs If, $k = 0, \cdots$, the maximum order n, the coefficients of the kth derivative of unknown function, say u, does not contain any of u and its derivatives, we call the DE linear. Otherwise, it is called nonlinear.

- Homogeneous vs. nonhomogeneous DEs. If every term in the DE contains the unknown function u or its derivatives, we say that the DE is "homogeneous". Otherwise, it is said to be "nonhomogeneous".

Example 1. *1.* $uu' + u = 1$

2. $u'' + uu' + u = 0$

3. $u'' + u' + u = t$

4. $u'' + u' + u = 0$

2.1.3 Some applications of differential equations

1. falling stone $y^{''} = g = const.$
2. parachutist $mv' = mg - bv^2$.
3. outflowing water $h' = -k\sqrt{h}$.

4. vibrating mass on a spring $my'' + ky = 0$.

5. Current I in a RLC circuit $LI'' + RI' + I/C = E'$.

6. pendulum $L\theta'' + gsin\theta = 0$.

7. Lotka-Volterra predator prey model $y_1' = ay_1 - by_1y_2, y_2' = ky_1y_2 - ly_2$.

2.1.4 Linear Differential Equations

Let $p(t)$ denote the population of a certain species at time t. The simplest population model is described by the relationship between the growth rate of the population and the current population. That is,

$$\frac{dp}{dt} = bp(t) - dp(t), \tag{2.2}$$

where b and d means the birth and death rates. Let us denote by $r = b - d$. Assume that the population at time t_0 is known to be p_0 and we would like to find the population at later time $t > t_0$.

This is one of the easiest examples of differential equations. The solution of this differential equation is given as follows.

$$\frac{\frac{dp}{dt}}{p(t)} = r. \tag{2.3}$$

Since the LHS is the derivative of $\log|p(t)|$, integrating both sides from t_0 to t yields

$$\int_{t_0}^{t} \frac{\frac{dp}{dt}}{p(s)}\, ds = \int_{t_0}^{t} r\, ds. \tag{2.4}$$

Thus,

$$\log|p(t)| - \log|p(t_0)| = r(t - t_0), \tag{2.5}$$

from which it follows that

$$p(t) = p_0 e^{r(t-t_0)}. \tag{2.6}$$

Notice that the sign of p_0 determines that of the absolute value $|p(t)|$.

Example 2. *In Sep. 1991, the famous mummy(Oetzi) from the Neolithic period of the Stone Age had been found. When did Oetzi approximately live and die if the ratio of carbon $6C^{14}$ to carbon $6C^{12}$ in this mummy is 52.5% of that of a living organism?*

physical information In the atmosphere and in living organisms, the ratio of radioactive carbon $6C^{14}$ to ordinary carbon $6C^{12}$ is constant. When an organism dies, its absorption of $6C^{14}$ by breathing and eating terminatess. Hence one can estimate the age of a fossil by comparing the radioactive carbon ratio in the fossil with that in atmosphere. To do this, needs to know the half-life(the period of time that it takes for a radioactive substance to decay to half of its original amount) of $6C^{14}$, which is 5715 yrs.
Experiments show that at each instant a radioactive substance decomposes at a rate proportional to the amount present.

modeling ans solve Radioactive decay is goverened by the ODE $y' = ky, k < 0$, then $y = y_0 e^{kt}$, y_0 is the initial ratio of $6C^{14}$ to carbon $6C^{12}$. Next, we use half-life(H) =5715 to decide $k(y_0 e^{kH} = 0.5y_0, k = -0.0001213)$. Finally, use the ratio 52.5% to determine the time t when Oetzi died.$(e^{kT} = 0.525 --> T = 5312)$

2.1.5 General first-order linear ODEs: Use an integrating factor

Consider the general first-order linear ODEs:

$$\frac{du}{dt} = a(t)u + f(t). \tag{2.7}$$

Let us write

$$\frac{du}{dt} - a(t)u = f(t). \tag{2.8}$$

Notice that it is not immediate to see primitive functions of the LHS. Observe that if we multiply the both sides by the following integrating factor

$$\mu(t) = e^{-\int_{t_0}^{t} a(s)\, dt}, \tag{2.9}$$

Now we have

$$\mu(t)\left(\frac{du}{dt} - a(t)u\right) = \mu(t)f(t), \tag{2.10}$$

whose LHS is the derivative of $\mu(t)u(t)$. This is the main contribution of the integrating factor $\mu(t)$. Then, by integrating both sides of the above equation, we get

$$\mu(t)u(t) - \mu(t_0)u(t_0) = \int_{t_0}^{t} \mu(s)f(s)\, ds. \tag{2.11}$$

Since $\mu(t_0) = 1$ and $\frac{1}{\mu(t)} = e^{\int_{t_0}^{t} a(s)\, dt}$ Hence we find

$$\begin{aligned} u(t) &= e^{\int_{t_0}^{t} a(s)\, ds}\left(u(t_0) + \int_{t_0}^{t} \mu(s)f(s)\, ds\right) && \text{(2.12a)} \\ &= e^{\int_{t_0}^{t} a(\tau)\, d\tau}\left(u(t_0) + \int_{t_0}^{t} e^{-\int_{t_0}^{s} a(\tau)\, d\tau} f(s)\, ds\right) && \text{(2.12b)} \\ &= e^{\int_{t_0}^{t} a(\tau)\, d\tau} u(t_0) + \int_{t_0}^{t} e^{\int_{s}^{t} a(\tau)\, d\tau} f(s)\, ds. && \text{(2.12c)} \end{aligned}$$

Example 3. *1.* $\frac{du}{dt} + u(t) = \frac{1}{1+t^2}, \quad u(0) = 1.$

2. $\frac{du}{dt} = 2tu(t) + e^{-t}\sin t.$

3. $(1+t^2)\frac{du}{dt} = -tu(t) + (1+t^2)^{3/2}.$

2.1.6 Nonlinear DEs; separable differential equations

So far we treated the linear differential equations of first order. However, most real world problems are modeled as nonlinear DEs of the form

$$\frac{du}{dt} = f(u,t), \quad u(t_0) = u_0 \tag{2.13}$$

These are usually not easily solvable in exact formula. Most of these nonlinear DEs are calculated numerically. One needs to employ possibly the most efficient and stable numerical methods among all available numerical schemes. This is the point at which Computational Sciences play an important role.

Before we discuss numerical methods, we show two types of nonlinear DEs of first order that can be easily solved in closed form. They are separable DEs and exact DEs.

If the nonlinear DE can be written as

$$\frac{du}{dt} = \frac{g(t)}{f(u)}, \quad u(t_0) = u_0 \tag{2.14}$$

it is called a separable DE. Usually it can be solved by separating the terms containing independent and dependent variables t and u.

$$f(u)\frac{du}{dt} = g(t), \tag{2.15}$$

The it can be written in the form

$$\frac{d}{dt}F(u(t)) = g(t) \tag{2.16}$$

where $F(u)$ is any anti- derivative of $f(u)$;, $F(u) = \int f(u)du$. Consequently,

$$F(u(t)) = \int g(t)dt + c. \tag{2.17}$$

OR

$$f(u)\ du = g(t)\ dt \tag{2.18}$$

Integrate both sides with respect to u from u_0 to u and t from t_0 to t.

$$\int_{u_0}^{u} f(u)\ du = \int_{y_0}^{t} g(s)\ ds. \tag{2.19}$$

In case the LHS is explicitly integrable, we can integrate both sides and solve for $u(t)$.

Example 4. *1.* $\frac{du}{dt} = au - bu^2, \quad u(0) = u_0.$

2. $e^t \frac{du}{dt} = 1 + t + t^2, \quad u(1) = 1.$

3. $\frac{du}{dt} = \frac{\sin t}{1+e^u}, \quad u(0) = 0.$

2.1.7 Nonlinear DEs; exact differential equations

The other class of DEs we will treat is of the form

$$M(t,u) + N(t,u)\frac{du}{dt} = 0. \tag{2.20}$$

If $\frac{M(t,u)}{N(t,u)}$ is separable, we can solve the DE by separate the variables. Otherwise, we can solve it if it can be written as

$$\frac{d\phi(t,u)}{dt} = 0, \tag{2.21}$$

for some smooth function $\phi(t,u)$ of two variables. In this case, we can integrate both sides to have

$$\phi(t,u) = c, \tag{2.22}$$

where c is determined uniquely if we have an initial condition.

By the chain rule, (2.21) is equivalent to

$$\frac{\partial \phi(t,u)}{\partial u}\frac{du}{dt} + \frac{\partial \phi(t,u)}{\partial t} = 0, \tag{2.23}$$

which we hope to be of the same form as (2.20). Thus, if

$$M(t,u) = \frac{\partial \phi(t,u)}{\partial t}, \quad \text{and } N(t,u) = \frac{\partial \phi(t,u)}{\partial u}.$$

Since

$$\frac{\partial}{\partial u}\frac{\partial \phi(t,u)}{\partial t} = \frac{\partial}{\partial t}\frac{\partial \phi(t,u)}{\partial u},$$

the condition is equivalent to have

$$\frac{\partial M}{\partial u} = \frac{\partial N}{\partial t}. \tag{2.24}$$

If this condition is satisfied, we call the DE (2.20) is *exact*. The solution procedure for exact DEs is as follows: we have

$$M(t,u) = \frac{\partial \phi(t,u)}{\partial t} \quad \text{and} N(t,u) = \frac{\partial \phi(t,u)}{\partial u}.$$

From

$$\int M(t,u)\,dt = \int \frac{\partial \phi(t,u)}{\partial t}\,dt \quad \text{and} \quad \int N(t,u)\,du = \int \frac{\partial \phi(t,u)}{\partial u}\,du,$$

choose any LHS which you can integrate easily. Suppose the first. Then

$$\phi(t,u) = \int M(t,u)\,dt + f(u), \tag{2.25}$$

where $f(t)$ is an integral constant with respect to the t-variable. Next, differentiate both sides with respect to u so that

$$\frac{\partial\phi(t,u)}{\partial u} = \frac{\partial}{\partial u}\int M(t,u)\,dt + f'(u), \tag{2.26}$$

Since

$$\frac{\partial\phi(t,u)}{\partial u} = N(t,u),$$

$f(t)$ can be determined by integrating with respect to u

$$f'(u) = N(t,u) - \frac{\partial}{\partial u}\int M(t,u)\,dt$$

Then $u(t)$ can be found by solving (2.25) in terms of t.

Instead of integrating $\int M(t,u)\,dt$, if $\int N(t,u)\,du$ is easily integrable, a similar procedure will solve the exact DE.

Example 5. *1.* $u + e^t + (t + \sin u)\frac{du}{dt} = 0.$

2. $3t^2u + 8tu^2 + (t^3 + 8t^2u + 12u^2)\frac{du}{dt} = 0.$

3. $\cos(u) + e^t + t\sin u\frac{du}{dt} = 0, \quad u(0) = 1.$

2.1.8 Existence theory

A simple geometric interpretation of a first order ODE

$$\frac{du}{dt} = f(t,u), \quad u(t_0) = u_0. \tag{2.27}$$

A solution curve that passes through (t_0, u_0) must have at that pt the slope $u'(t_0)$ equal to the value of f at that pt; that is $u'(t_0) = f(t_0, u_0)$. It follows that you can indicate directions of solution curves by drawing short straight-line segments in the tu plane and then fitting (approximate) solution curves through the direction field(or slope field) thus obtained.

This is important for two reasons.

1.Need not solve DE, This is essential becz many ODEs have complicated solution formulas or none at all.

2.This method shows, in graphical form, the whole family of solutions and their typical properties, accuracy is somewhat limited, but most cases this does not matter.

Consider the initial-value problem

$$\frac{du}{dt} = f(t,u), \quad u(t_0) = u_0 \tag{2.28}$$

Theorem 1. *Let* $R = [t_0, t_0 + a] \times [u_0 - b, u_0 + b]$*. Suppose* $f \in C^0(R)$ *and* $\frac{\partial f}{\partial u} \in C^0(R)$*. Let*

$$M := \max_{(t,u)\in R} |f(t,u)|, \quad \alpha = \min(a, \frac{b}{M}). \tag{2.29}$$

Then the initial-value problem has a unique solution $u = u(t)$ *for* $t \in [t_0, t_0 + \alpha]$*.*

In practice, the IVP does not have a closed form solution. Usually we have to solve the problem using numerical methods which will covered in later.

2.2 Higher order linear differential equations

Theorem 2. Fundamental theorem for the homogeneous linear ODE
For homogeneous linear ODE (2.30)*, any linear combination of solutions on an ope interval I is again a solution of* (2.30)*.In particular, for such an equation, sums and constant multiples of solutions are again solutions.*

2.2.1 Constant coefficients case: homogeneous differential equations

We consider

$$\sum_{k=0}^{n} a_k \frac{d^k u}{dt^k} = 0. \tag{2.30}$$

Consider the solution of the form $u(t) = e^{rt}$. Then we get

$$(\sum_{k=0}^{n} a_k r^k) e^{rt} = 0. \tag{2.31}$$

Thus r should satisfy the nth degree polynomial equation:

$$\sum_{k=0}^{n} a_k r^k = 0, \tag{2.32}$$

Notice that the above polynomial equation has the factorization (by the "Fundamental Theorem of Algebra") as follows:

$$a_n \Pi_{k=1}^{n} (r - r_k) = 0, \quad r_k \in \mathbb{C}, \tag{2.33}$$

where r_k may be repeated. Moreover, if any r_k is in $\mathbb{C} \setminus \mathbb{R}$, then there exists a r_j such that r_j is the complex conjugate $\overline{r_k}$ of r_k. That is, if $r_k = \alpha_k + i\beta_k$ with $\alpha_k, \beta_k \in \mathbb{R}$, then $r_j = \alpha_k - i\beta_k$. Grouping real and complex roots with repetition, we write (2.33) as follows:

$$\Pi_{j=1}^{J_R} (r - r_j)^{p_j} \Pi_{j=1}^{J_I} [(r - (\alpha_j + i\beta_j))(r - (\alpha_j - i\beta_j))]^{q_j} = 0, \tag{2.34}$$

where the powers p_j, q_j are positive integer satisfying

$$\sum_{j=1}^{J_R} p_j + 2\sum_{j=1}^{J_I} q_j = n \tag{2.35}$$

Hence by linearity, the solution of (2.30) is given by

$$\begin{aligned} u(t) \quad = \quad & \sum_{j=1}^{J_R}\left[\sum_{k=1}^{p_j} B_{j,k} t^{k-1} e^{r_j t}\right] \\ & + \sum_{j=1}^{J_I}\left[\sum_{k=1}^{q_j} e^{\alpha_j t}\left(C_{j,k} t^{k-1}\cos\beta_j t + D_{j,k} t^{k-1}\sin\beta_j t\right)\right] \end{aligned}$$

with the n real coefficients $B_{j,k}$'s, $B_{j,k}$'s, and $D_{j,k}$'s may be determined uniquely once linear independent n initial and/or boundary conditions are imposed.

Example 6. *1.* $u'' + \omega^2 u = 0.$

2. $u^{(4)} + u = 0.$

3. $u''' - 3u'' + 3u' - 1 = 0.$

4. $u''' - 3u'' + 3u' - 1 = 0. \quad u(0) = 1, u'(0) = 2, u''(0) = 0.$

5. $u'' + \omega^2 u = 0. \quad u(0) = 1, u(1) = 0.$

2.2.2 Constant coefficients case: nonhomogeneous differential equations

We consider the nonhomogeneous n order linear DE with constant coefficients:

$$\sum_{k=1}^{n} a_k \frac{d^k u}{dt^k} = f(t). \tag{2.36}$$

All nonhomogeneous linear DE can be solved in two steps. The first step is to solve the homogeneous linear DE

$$\sum_{k=0}^{n} a_k \frac{d^k u_h}{dt^k} = 0, \tag{2.37}$$

which is given in (2.36).

The second step is to find a particular solution

$$\sum_{k=0}^{n} a_k \frac{d^k \psi}{dt^k} = f(t). \tag{2.38}$$

Then the general solution is decomposed by $u(t) = u_h(t) + \psi(t)$. Thus,

$$\begin{aligned} u(t) &= \sum_{j=1}^{J_R} \left[\sum_{k=1}^{p_j} B_{j,k} t^{k-1} e^{r_j t} \right] \\ &\quad + \sum_{j=1}^{J_I} \left[\sum_{k=1}^{q_j} e^{\alpha_j t} \left(C_{j,k} t^{k-1} \cos \beta_j t + D_{j,k} t^{k-1} \sin \beta_j t \right) \right] + \psi(t). \end{aligned}$$

The particular solution can be found by using (1) The method of variation of parameters, (2) The method of judicious guessing

2.2.3 modeling:Free Oscillations

Consider an ordinary spring that resists compression as well extension and suspend it vertically from a fixed support. At the lower end of the spring, attach a body of mass m. Assume m be so large that we can neglect the mass of the spring. If we pull the body down a certain distance and then release it, it starts strictly moving vertically. How can we obtain the motion of the body, say the displacement $y(t)$ as function of time t?

Physical Information

Newton's second law-Force is the resultant of all the forces acting on the body.

Hooke's law-In spring stretching system, $F_0 = -ks_0$, where k : spring constant, s_0:the stretch, F_0 : an upward force.

Choose the downward direction as the positive direction, thus regarding downward forces as positive and vice versa.

From the position $y = 0$ we pull the body downward. This stretches the spring by some amount $y > 0$. By Hooke's law, $F_1 = -ky$, the restoring force to pull the body back to $y = 0$.

Undamped System If the damping effect is negligible,

$$my'' + ky = 0. \tag{2.39}$$

Then the general solution is $y(t) = Acosw_0t + Bsinw_0t = Ccos(w_0t - \delta)$,$w_0 = \sqrt{k/m}, C = \sqrt{A^2 + B^2}, \delta = arctan(B/A)$ which is called harmonic oscillation. It describes the body executes with frequency $w_0/2\pi$ Hz.

Damped System If we now add the damping force $F_2 = -cy'$ by considering the friction.

$$my'' + cy' + ky = 0. \tag{2.40}$$

The solution is

- Overdamping $y(t) = c_1exp(-(\alpha-\beta)t)+c_2exp(-(\alpha+\beta)t), \alpha = c/2m, \beta = 1/2m\sqrt{c^2 - 4mk}$, damping takes out energy so quickly that the body does not oscillate.

- Critical damping $y(t) = (c_1 + c_2t)exp(-\alpha t), \alpha = c/2m, \beta = 0$, this sol can pass ti the equilibrium $y = 0$ at most once. If c_1, c_2 are both positive, it can't pass $y = 0$.

- Underdamping - damped oscillations
  $y(t) = exp(-\alpha t)(Acos\omega^*t + Bsin\omega^*t) = Cexp(-\alpha t)cos(\omega^*t - \delta)$,
  $\beta = i\omega^*, \omega^* = \sqrt{k/m - c^2/(4m^2)}, C^2 = A^2 + B^2, tan\delta = B/A$.

2.2.4 modeling:Forced Oscillations

In addition to free oscillations, add an external force, $r(t)$, say resonance.

$$my'' + cy' + ky = r(t) = F_0 cos\omega t, \quad F_0 > 0, \omega > 0. \tag{2.41}$$

This is a nonhomogeneous linear 2nd ODE. To find the particular solution $y_p(t)$, start from $y_p(t) = acos\omega t + bsin\omega t$. After finding a, b, the general sol becomes $y(t) = y_h(t) + y_p(t)$.

2.2.5 Variable coefficients case

Even for second order DEs, if the coefficients are variable it is not usually easy to solve the DE.

One of the easiest cases may be Euler's equation of the form:

$$a_2 t^2 \frac{d^2u}{dt^2} + a_1 t \frac{du}{dt} + a_0 u = 0. \tag{2.42}$$

In this case, one can find a solution of the form $u(t) = t^r$ for some r. Indeed, since

$$\frac{du}{dt} = rt^{r-1}, \quad \frac{d^2u}{dt^2} = r(r-1)t^{r-2}, \tag{2.43}$$

one then has

$$[a_2 r(r-1) + a_1 r + a_0]\, t^r = 0. \tag{2.44}$$

This reduces to a quadratic equation

$$a_2 r(r-1) + a_1 r + a_0 = 0. \tag{2.45}$$

There are three types of solutions for the above quadratic equation:

1. Two distinct real roots: $r = r_1, r_2$ with $r_1 \neq r_2$:

$$u(t) = C_1 t^{r_1} + C_2 t^{r_2}. \tag{2.46}$$

2. Double real roots: $r = r_1$:

$$u(t) = C_1 t^{r_1} + C_2 t^{r_1} \ln t. \tag{2.47}$$

3. Complex roots: $r_1 = \alpha + i\beta$, $r_2 = \overline{r_1}$

$$u(t) = C_1 t^{\alpha} \cos(\beta \ln t) + C_2 t^{\alpha} \sin(\beta \ln t). \tag{2.48}$$

Example 7. *1.* $t^2 \frac{d^2u}{dt^2} + t\frac{du}{dt} - 4u = 0.$

In this case, try to find a solution of form $u(t) = t^r$*. Then, we have* $r(r-1) + r - 4 = 0$*. Hence,* $r = \pm 2$*. Thus the general solution is* $u(t) = C_1 t^2 + C_2 \frac{1}{t^2}$*.*

2. $t^2\frac{d^2u}{dt^2} - t\frac{du}{dt} + u = 0.$

   In this case, try to find a solution of form $u(t) = t^r$. *Then, we have* $r(r-1) - r + 1 = 0$. *Hence,* $r = 1$ *is a double root. Thus the general solution is* $u(t) = C_1 t + C_2 t \ln t$.

3. $t^2\frac{d^2u}{dt^2} - t\frac{du}{dt} + 4u = 0.$

   We have $r(r-1) - r + 5 = 0$. *Hence,* $r = 1 \pm 2i$ *a complex root. The general solution is* $u(t) = C_1 t\cos(2\ln t) + C_2 t\sin(2\ln t)$.

2.2.6 Method of Frobenius

In general, consider

$$\sum_{j=0}^{n} a_j(t)\frac{d^j u}{dt^j} = f(t). \tag{2.49}$$

The DE is has a singular point if the leading coefficient $a_n(t)$ vanishes at $t = t_*$. In this case, seek a solution of the form

$$u(t) = t^r \sum_{j=0}^{\infty} A_j t^j, \tag{2.50}$$

and plug this into (2.49) to have relationship for A_j's and r. The solution procedure is extremely long and tedious.

Consider Bessel's equation of order ν of the form:

$$t^2\frac{d^2u}{dt^2} + t\frac{du}{dt} + (t^2 - \nu^2)u = 0. \tag{2.51}$$

When $\nu = 1/2$, we have $u_1(t) = \frac{sint}{\sqrt{t}}$ and $u_2(t) = \frac{cost}{\sqrt{t}}$.
When $\nu = 0$, we have $u_1(t) = J_0(t) = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{2^{2n}(n!)^2}$ is referred to the first kind of order zero and $u_2(t) = u_1(t)lnt + \sum_{n=0}^{\infty} \frac{(-1)^{n+1} H_n t^{2n}}{2^{2n}(n!)^2}$.

2.2.7 Systems of Differential Equations

Let us say that we have a simple pendulum, where a particle of mass m is supported by an inelastic string of length l. Suppose that the motion of the pendulum satisfies a second-order nonlinear DE:

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\sin\theta = 0 \tag{2.52}$$

Let $x_1(t) = \theta(t)$, $x_2(t) = \frac{d\theta}{dt}(t)$. Then we have

$$\frac{d}{dt}\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} x_2(t) \\ -\frac{g}{l}\sin x_1(t) \end{bmatrix}, \tag{2.53}$$

a system of first-order nonlinear DEs.

There are two trivial solutions:
$\{x_1(t)=0, x_2(t)=0\}$ and $\{x_1(t)=\pi, x_2(t)=0\}$, which are stable and
unstable, respectively.

- Solvability
- Stability
- Approximation: numerical methods
- Autonomous vs. nonautonomous
  If the system does not contain a coefficient which has ”t” variable explicitly, we call the system autonomous; otherwise it is called nonautonomous

Example 8. *Linear case*
$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} (= A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix})$
We try to find a solution of the form

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = e^{\lambda t}\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}, \ \xi_1,\ \xi_2 \ \textit{constants}, \overrightarrow{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}. \tag{2.54}$$

Then

$$\lambda e^{\lambda t}\overrightarrow{\xi} = Ae^{\lambda t}\overrightarrow{\xi}, \ or \ A\overrightarrow{\xi} = \lambda\overrightarrow{\xi} \tag{2.55}$$

λ *: eigenvalue (characteristic value)*
$\overrightarrow{\xi}$ *: eigenvector (characteristic vector)*

$[\lambda I - A]\overrightarrow{\xi} = \overrightarrow{0}$
$\lambda I - A$ *should be singular or noninvertible if* $\overrightarrow{\xi} \neq \overrightarrow{0}$ *exists*

$\Rightarrow det(\lambda I - A) = 0,$

$det\begin{bmatrix} \lambda-3 & 2 \\ -1 & \lambda-1 \end{bmatrix} = (\lambda-3)(\lambda-1)+2 = \lambda^2 - 4\lambda + 5$
$= (\lambda-2)^2 + 1 = 0$

Therefore, $\lambda = 2 \pm i$.
Now we need to find the eigenvectors:

1. $\lambda = 2+i : \begin{bmatrix} 2+i-3 & 2 \\ -1 & 2+i-1 \end{bmatrix} = \begin{bmatrix} -1+i & 2 \\ -1 & 1+i \end{bmatrix}$

$$\begin{bmatrix} -1+i & 2 \\ -1 & 1+i \end{bmatrix}\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \xi_1 = 1+i,\ \xi_2 = 1$$

2. $\lambda = 2-i:$ Complex conjugate $\xi_1 = 1-i,\ \xi_2 = 1$

Thus we have two solutions of the form $e^{\lambda t}\begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}$.

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = e^{(2+i)t}\begin{bmatrix} 1+i \\ 1 \end{bmatrix} \text{and } \ e^{(2-i)t}\begin{bmatrix} 1-i \\ 1 \end{bmatrix},$$

which are complex conjugates.

A complex-valued solution: $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = C_1 e^{(2+i)t}\begin{bmatrix} 1+i \\ 1 \end{bmatrix} + C_2 e^{(2-i)t}\begin{bmatrix} 1-i \\ 1 \end{bmatrix}$
with $C_1, C_2 \in \mathbb{C}$.

For a real-valued solution: Look at

$$\begin{aligned} e^{(2+i)t}\begin{bmatrix} 1+i \\ 1 \end{bmatrix} &= e^{2t}(\cos t + i\sin t)\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} + i\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \\ &= e^{2t}\left(\cos t\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \sin t\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + ie^{2t}\left(\cos t\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \sin t\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) \\ &= e^{2t}\begin{bmatrix} \cos t - \sin t \\ \cos t \end{bmatrix} + ie^{2t}\begin{bmatrix} \cos t + \sin t \\ \sin t \end{bmatrix} \end{aligned}$$

Thus, a general real-valued solution is given by

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = c_1 e^{2t}\begin{bmatrix} \cos t - \sin t \\ \cos t \end{bmatrix} + c_2 e^{2t}\begin{bmatrix} \cos t + \sin t \\ \sin t \end{bmatrix} \tag{2.56}$$

with $c_1, c_2 \in \mathbf{R}$

Predator-prey system (Vito Volterra)

$$\frac{dx}{dt} = ax - bxy \ : \text{prey}$$

$$\frac{dy}{dt} = -cy + bxy \ : \text{predator}$$

Equilibrium point (solution)

$$x(a-by) = 0 \qquad y(-c+bx) = 0$$

A closed formula

$$x^c y^b = K e^{dx+by}$$

$$\frac{dy}{dx} = \frac{y(-c+dx)}{x(a-by)} \Rightarrow \frac{a-by}{y}dy = \frac{-c+dx}{x}dx$$

Theorem 3. *An* $n^{th} - order ODE y^{(n)} = F(t, y, y', \cdots, y^{(n-1)})$can be converted to a system of n first-order ODEs by setting

$$y_1 = y, \quad y_2 = y', \quad y_3 = y'', \quad , \cdots, y_n = y^{(n-1)}).$$

This system is of the form

$$\begin{aligned} y_1^{'} &= y_2 \\ y_2^{'} &= y_3 \\ y_{n-1}^{'} &= y_n \\ y_n^{'} &= F(t, y_1, y_2, \cdots, y_n). \end{aligned}$$

2.2.8 Laplace -Fourier transformation

Laplace transform(ation)

$$f(t) \to [Lf](s) := \int_0^\infty f(t) e^{-st} dt, s \in \mathbb{C} \tag{2.57}$$

Inversion formula

$$f(t) = \frac{1}{2\pi i} \int_\gamma [Lf](z) e^{sz} dz,$$

where γ is a complex contour.
Fourier transform(ation)

$$f(t) \to [Ff](w) = \int_{-\infty}^\infty f(t) e^{-iwt} dt \tag{2.58}$$

Inversion formula

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^\infty [Ff](w) e^{iwt} dw,$$

where $w \in \mathbf{R}$

The Laplace transform and Fourier transform exist if $\int_0^\infty |f(t)| e^{-Re(s)t} dt < \infty$ and $\int_{-\infty}^\infty |f(t)| dt < \infty$, respectively.

Suppose f(t) is periodic with period $T > 0$. That is,

$$f(t+T) = f(t) \quad \forall t \in \mathbf{R}$$

In this case,

$$\int_{-\infty}^{\infty}|f(t)|dt=\sum_{k=-\infty}^{\infty}\int_{kT}^{(k+1)T}|f(t)|dt=\infty\cdot\int_{0T}^{T}|f(t)|dt=\infty=\sum_{k=-\infty}^{\infty}\int_{(k-\frac{1}{2})T}^{(k+\frac{1}{2})T}|f(t)|dt \tag{2.59}$$

whenever

$$\int_0^T|f(t)|dt>\infty$$

Therefore the Fourier transform (2) will not exist.

In this case, we define the Fourier series as follows:

$$f(t)\to\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}f(t)e^{-\frac{2\pi i}{T}kt}dt:=\widehat{f}(k),\quad \text{for all integer}k \tag{2.60}$$

Inversion formula

$$f(t)=\sum_{k=-\infty}^{\infty}\widehat{f}(k)e^{\frac{2\pi i}{T}kt} \tag{2.61}$$

where $w=2\pi/Ti$.
Indeed,

$$\begin{aligned}f(t)=\sum_{k=-\infty}^{\infty}\left(\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}f(s)e^{-\frac{2\pi i}{T}ks}ds\right)e^{\frac{2\pi i}{T}kt} &= \sum_{k=-\infty}^{\infty}\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}f(s)e^{\frac{2\pi i}{T}k(t-s)}ds\\ &= \int_{-\frac{T}{2}}^{\frac{T}{2}}f(s)\sum_{k=-\infty}^{\infty}\frac{1}{T}e^{\frac{2\pi i}{T}k(t-s)}ds\\ &= f(t)\end{aligned}$$

$$\begin{aligned}\sum_{k=-\infty}^{\infty}\frac{1}{T}e^{\frac{2\pi i}{T}k(t-s)} &= \frac{1}{T}\sum_{k=1}^{\infty}\left[e^{\frac{2\pi i}{T}k(t-s)}+e^{-\frac{2\pi i}{T}k(t-s)}\right]\\ &= \frac{1}{T}\sum_{k=1}^{\infty}2\cos\frac{2\pi k}{T}(t-s)+\frac{1}{T}.\end{aligned}$$

Note that

$$\begin{aligned}\int_{-\frac{T}{2}}^{\frac{T}{2}}2\cos\frac{2\pi k}{T}(t-s)ds &= \frac{T}{2\pi k}\left[\sin\left(\frac{2\pi k}{T}(t-\frac{T}{2})\right)-\sin\left(\frac{2\pi k}{T}(t+\frac{T}{2})\right)\right]\\ &= \frac{T}{2\pi k}\left[\sin\left(\frac{2\pi k}{T}t-\pi k\right)-\sin\left(\frac{2\pi k}{T}+\pi k\right)\right]\\ &= 0\end{aligned}$$

$T = 2\pi$ case: From eq. (3.58) we have

$$f(t) = \sum_{k=-\infty}^{\infty} \widehat{f}(k) e^{ikt} \tag{2.62}$$

$$\widehat{f}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-ikt}dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\cos ktdt - i\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\sin ktdt$$

$$\begin{aligned} f(t) = \sum_{k=-\infty}^{\infty} \widehat{f}(k)[\cos kt + i\sin kt] &= \sum_{k=-\infty}^{\infty} (A_k - iB_k)(\cos kt + i\sin kt) \\ &= \sum_{k=-\infty}^{\infty} (A_k \cos kt + B_k \sin kt) + i\sum (A_k \sin kt - B_k \cos kt) \end{aligned}$$

If $f(t)$ is a real-valued function,

$$f(t) = A_0 + \sum_{k=1}^{\infty}[A_k \cos kt + B_k \sin kt]$$

since $A_{-k} = A_k$, $B_{-k} = -B_k$.

Notice that we have

$$A_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos ktdt \ , \ B_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin ktdt$$

$$A_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)dt$$

Convolution

$$(f*g)(t) = \int_{-\infty}^{\infty} f(s)g(t-s)ds = (g*f)(t)$$

$$(f*g)(t) = \int_{0}^{t} f(s)g(t-s)ds \ \text{ if } \ f(s) = g(s) = 0 \ \text{ for } \ s < 0$$

$$(f*g)(t) = (g*f)(t) \tag{2.63}$$

$$L[f*g](s) = [Lf](s)\ [Lg](s) \tag{2.64}$$

$$F[(f*g)](w) = [Ff](w)\ [Fg](w) \tag{2.65}$$

Proof:

$$\begin{aligned} \int_{-\infty}^{\infty}(f*g)(t)e^{-iwt}dt &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(s)g(t-s)ds\right)e^{-iwt}dt \\ &= \int_{-\infty}^{\infty} f(s)\left[\int_{-\infty}^{\infty} g(t-s)e^{-iw(t-s)}dt\right]e^{-iws}ds \\ &= [Fg](w)\int_{-\infty}^{\infty} f(s)e^{-iws}ds \\ &= [Ff](w)\ [Fg](w) \end{aligned}$$

Parseval's relation

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \int_{-\infty}^{\infty} |[Ff](w)|^2 \, dw$$

Proof:

$$\begin{aligned}
\int_{-\infty}^{\infty} |[Ff](w)|^2 \, dw &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(t)e^{-iwt}dt\right]^2 dw \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t)e^{-iwt}dt \int_{-\infty}^{\infty} f(s)e^{+iws}dsdw \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t)f(s)e^{-iwt}e^{+iws}dwdsdt \\
&= \int_{-\infty}^{\infty} f(t) \int_{-\infty}^{\infty} f(s) \int_{-\infty}^{\infty} e^{-iw(t-s)}dwdsdt \\
&= \int_{-\infty}^{\infty} f(t) \int_{-\infty}^{\infty} f(s)\delta(t-s)dsdt \\
&= \int_{-\infty}^{\infty} f(t)f(t)dt \\
&= \int_{-\infty}^{\infty} |f(t)|^2 \, dt
\end{aligned}$$

Differentiation and Integration of transforms

Theorem 4. *If $L(f) = F(s)$, then $L\{tf(t)\} = -F'(s)$, i.e.,$L^{-1}\{F'(s)\} = -tf(t)$.*
$L\{f(t)/t\} = \int_s^{\infty} F(s)ds$, hence $L^{-1}\{\int_s^{\infty} F(s)ds\} = f(t)/t$.

3 Numerical Models for ODEs

3.1 Numerics in General

3.1.1 Solutions of Equations by iteration

- Fixed point iteration for solving $f(x) = 0$.
  In one way or another we transform $f(x) = 0$ algebraically to the form $x = g(x)$. Then we choose an x_0 and compute $x_{n+1} = g(x_n)$, (n=0,1,...).

  Theorem 5. Convergence of fixed point iteration
  Let $x = s$ be a sol of $x = g(x)$, suppose g has a continuous derivative in some interval J containing s. Then if $\|g'(x)\| \leq K < 1$ in J, the iteration process defined by $x_{n+1} = g(x_n)$, (n=0,1,...) converges for any x_0 in J, and $lim x_n = s$.

  Example 9. *Set up an iteration process for $f(x) = x^3 + x + 1 = 0$ by fixed point iteration.*

- Newton's method(Newton-Raphson) for solving $f(x) = 0$.

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad (n = 0, 1, ...)$$

with properly chosen x_0.

Example 10. *Set up an iteration process for $f(x) = x^3 + x + 1 = 0$ by Newton's method.*

Remark 1. *Convergence of Iteration by Newton's method is much more rapid than that of fixed pt iteration and this motivates the concept of the order of an iteration process.*

Let $x_{n+1} = g(x_n)$ define an iteration method, x_n approximate a solution s of $x = g(x)$. Then $x_n = s - \epsilon_n$, where ϵ_n is the error of x_n. Assume g is differentiable enough, then

$$\begin{aligned} x_{n+1} = g(x_n) = g(s) + g^{'}(s)(x_n - s) + 1/2g^{''}(s)(x_n - s)^2 + \cdots \\ = g(s) - g^{'}(s)\epsilon_n + 1/2g^{''}(s)\epsilon_n^2 + \cdots \end{aligned}$$

The exponent of ϵ_n in the first nonvanishing term agter $g(s)$ is called the order of the iteration defined by g and this measures the spd of convergence.

To see this subtract $g(s) = s$ on both sides.Then $x_{n+1} - s = -\epsilon_{n+1}$. If $\epsilon_{n+1} \approx g^{'}(s)\epsilon_n$, the first order of convergence, If $\epsilon_{n+1} \approx 1/2g^{''}(s)\epsilon_n^2$, the second order of convergence.

Theorem 6. 2nd order Convergence of Newton's method
If $f(x)$ is three times diff, f', f'' are not zero at a sol s of $f(x) = 0$, then for x_0 suff close to s, Newton's method is of 2nd order.

This is true due to $f(s) = 0, g^{'}(s) = 0, g^{''}(s) \neq 0$

- Secant method for solving $f(x) = 0$.

$$x_{n+1} = x_n - \frac{f(x_n)(x_n - x_{n-1})}{f(x_n) - f(x_{n-1})}, \quad (n = 1, ...)$$

with properly chosen x_0.
The convergence is superlinear, about 1.62 .

- Bisection method:always converge but often slowly
The input for the method is a continuous function f, an interval $[a, b]$, and the function values f(a) and f(b). The function values are of opposite sign (there is at least one zero crossing within the interval). Each iteration performs these steps:
1. Calculate c, the midpoint of the interval, c = 0.5 * (a + b).
2. Calculate the function value at the midpoint, f(c).
3. If convergence is satisfactory (that is, a - c is sufficiently small, or f(c) is sufficiently small), return c and stop iterating.
4. Examine the sign of f(c) and replace either (a, f(a)) or (b, f(b)) with (c, f(c)) so that there is a zero crossing within the new interval.

3.1.2 Interpolation

Interpolation means to find approximate values of a function $f(x)$ for an x between different x-values $x_0, x_1, \cdots, x_n$ at which the values of $f(x)$ are given. A standard ides is find a polynomial $p_n(x)$ of degree n or less that assumes the given values;

$$p_n(x_0) = f_0, p_n(x_1) = f_1, \cdots, p_n(x_n) = f_n. \tag{3.1}$$

Theorem 7. *Weierstrass approximation theorem For any continuous $f(x)$ on an interval $J : a \leq x \leq b$ and error bound β, there exist a unique polynomial $p_n(x)$ (of sufficiently figh degree n) such that $|f(x) - p_n(x)| < \beta$ for all x on J.*

- Lagrange Interpolation
  Given $(x_0, f_0), ..., (x_n, f_n)$, find a polynomial that is 1 at x_j and 0 at other nodes, and then take the sum of these n+1 polynomials to get the unique interpolation pol of degree $\leq n$.

  1. linear int
     Given $(x_0, f_0), (x_1, f_1), p_1 = L_0 f_0 + L_1 f_1$ with $L_0(x) = (x - x_1)/(x_0 - x_1), L_1(x) = (x - x_0)/(x_1 - x_0)$, this gives

     $$p_1(x) = L_0(x) f_0 + L_1(x) f_1.$$

  2. Quadratic int
     Given $(x_0, f_0), (x_1, f_1), (x_2, f_2), p_2 = L_0 f_0 + L_1 f_1 + L_2 f_2$ with $L_0(x) = (x - x_1)(x - x_2)/(x_0 - x_1)(x_0 - x_2), L_1(x) = (x - x_0)(x - x_2)/(x_1 - x_0)(x_1 - x_2), L_2(x) = (x - x_0)(x - x_1)/(x_2 - x_0)(x_2 - x_1)$, this gives

     $$p_1(x) = L_0(x) f_0 + L_1(x) f_1 + L_2(x) f_2.$$

  3. general int
     Given $(x_0, f_0), (x_1, f_1), ..., (x_n, f_n), p_n = L_0 f_0 + L_1 f_1 + L_2 f_2 + ... + L_n f_n$

     $$f(x) \approx p_n(x) = \sum_{k=0}^{n} L_k(x) f_k = \sum \frac{l_k(x)}{l_k(x_k)} f_k$$

     where $L_k(x_k) = 1, L_k(x_j) = 0, j \neq k, l_0(x) = (x - x_1)(x - x_2)...(x - x_n), ..., l_k(x) = (x - x_0)...(x - x_{k-1})(x - x_{k+1})...(x - x_n), l_n(x) = (x - x_0)...(x - x_{n-1})$. Then $p_n(x_k) = f_k$,

     $$\epsilon_n(x) = f(x) - p_n(x) = (x - x_0)...(x - x_n) \frac{f^{n+1}(t)}{(n+1)!}.$$

     Practically, if the error is difficul;t or impossible to obtain, take another node and find Lagrange poly $p_{n+1}(x)$ and regard $p_{n+1}(x) - p_n(x)$ as a crude error estimate for $p_n(x)$.

- Splines

  For some functions f(x0, poly interpolation may oscillate as n increases, numerical instability. Instead of a single high degree polynomial over the whole interval, we subdivide whole interval and use several low-degree polynomials(cannot oscillate much), one over each subinterval, and get an interpolating ft,called spline, by fitting them together into a single curve. i.e., spline interpolation is piecewise polynomial interpolation.

  Example 11. *Consider cubic spline. If f(x) is given on $a \leq x \leq b$ and subdivide it to $a = x_0 < x_1 < ... < x_n = b$, we obtain a cubic splibne $g(x)$ that approximates $f(x)$ by requiring that $g(x_0) = f(x_0) = f_0, ..., g(x_n) = f(x_n) = f_n$. Assume $g^{'(x_0)=k_0, g^{'(x_n)=k_n}}$, then we get the unique cubic spline interpolation.*

3.1.3 Numerical Integration and Differentiation

Numerical Integration means the numeric evaluation of integrals

$$J = \int_a^b f(x)dx,$$

where a, b are given, f is a given ft analytically by a formula or empirically by a table of values. If there exist $F(x)$ such that $F'(x) = f(x)$,

$$J = \int_a^b f(x)dx = F(b) - F(a).$$

Numerical integration methods are obtained by approximating the integran f by fts that can be easily integrated.

- Rectangular Rule

  $$J = \int_a^b f(x)dx = h[f(x_1^*) + f(x_2^*) + \cdots + f(x_n^*)] - \frac{(b-a)hf^{'}(\widehat{x})}{6}, \quad (h = (b-a)/n)$$

  where x_j^* is the midpoint of the jth subinterval.

- Trapezoidal Rule

  $$J = \int_a^b f(x)dx = h[0.5f(a) + f(x_1) + \cdots + f(x_{n-1}) + 0.5f(b)] - \frac{(b-a)h^2 f^{''}(\widehat{x})}{12},$$
  $$(h = (b-a)/n)$$

  In a single interval, $f(x) - p_1(x) = (x - x_0)(x - x_1)f^{''}(t)/2$ with a suitable t between x_0, x_1. Then

  $$\begin{aligned} \int_{x_0}^{x_0+h} f(x)dx - \frac{h}{2}[f(x_0) + f(x_1)] &= \int_{x_0}^{x_0+h} (x - x_0)(x - x_0 - h)\frac{f^{''}(t(x))}{2}dx \\ &= \int_0^h v(v-h)dv\frac{f^{''}(\tilde{t})}{2} = -\frac{h^3}{12}f^{''}(\tilde{t}). \end{aligned}$$

This is the error for the trapezoidal rule with n=1, the local error. The global error is local error*number of intervals, thus $-\frac{(b-a)h^2 f''(\widehat{x})}{12}$.

- Simpson's Rule

$$J = \int_a^b f(x)dx = \frac{h(f_0 + 4f_1 + 2f_2 + 4f_3 + \cdots + 2f_{2m-2} + 4f_{2m-1} + f_{2m})}{3}]$$
$$-\frac{(b-a)h^4 f^4(\widehat{x})}{180}, \quad (h = (b-a)/(2m))$$

- Gauss Integration formulas(Gauss quadrature formula)

$$\int_{-1}^{1} f(x)dx \approx \sum_{j=1}^{n} w_j f_j, \quad f_j = f(x_j).$$

Then we can determine the n coefficients(w_j) and n nodes (x_j) so that Gauss Integration formulas give exact results for polynomials of degree k ($\leq 2n-1$) as high as possible. If the integration interval is $[a, b]$, can convert to $[-1, 1]$ by $t = 0.5[a(x-1) + b(x+1)]$

3.2 Numerical Methods for ODE

Start from HERE

3.2.1 Initial Value Problem

Initial value problem of the first-order differential equation

- A differential equation and a condition the solution must satisfy:

$$y' = \frac{dy}{dt} = f(y,t), \quad y(t_0) = y_0$$

All numerical methods are to seek the solution at any discrete time t_n with IC.

Step-by Step approach

- Start from the given $y_0 = y(t_0)$ and proceed stepwise
- Compute approximate values of the solution $y(t)$ at discrete time levels

$t_1 = t_0 + h,\ t_2 = t_0 + 2h,\ t_3 = t_0 + 3h,\ \cdots$. (step size h is a fixed number)

Taylor series

$$y(t+h) = y(t) + hy'(t) + \frac{h^2}{2}y''(t) + \cdots \Rightarrow y(t+h) \approx y(t) + hy'(t) = y(t) + hf(y,t).$$

by assuming the higher order terms are small enough.

3.2.2 Euler Method

Explicit Euler's Method

- Simplest numerical method for the integration of the first-order ODE

$$y_{n+1} = y_n + hf(y_n, t_n),\ n = 0, 1, 2\cdots.$$

- Geometrically it is an approximation of the curve $y(t)$ by a polygon whose first side is tangent to the curve at t_0
- INSERT FIGURE

$$\begin{aligned} y_1 &= y_0 + hf(y_0, t_0) \\ y_2 &= y_1 + hf(y_1, t_1) \\ &. \\ &. \\ &. \end{aligned}$$

$$y' = \frac{dy}{dt} = f(y,t) \Rightarrow y_{n+1} = y_n + hf(y_n, t_n)$$

Hardly ever used in practice

equivalent rectangle integration rule with left e

Example used for error analysis

Truncation error (Local vs. Global)

- Local error

One-step error caused by numerical algorithm starting from the known initial condition

In Euler's method, the truncation error in each step is proportional to h^2, written $O(h^2)$.

$$y(t+h) = y(t) + hy'(t) + \frac{h^2}{2!}y''(t) + \frac{h^3}{3!}y'''(t)\cdots$$

- Global error

Accumulation of local errors

Error at some finite time of the problem

$$e_N = \sum_{n=1}^{N} e_n = \sum_{n=1}^{N} \frac{1}{2} y_n^{''}(\xi(t_n))h^2$$

where $t_{n-1} < \xi(t_n) < t_n$.

$$e_N = \sum_{n=1}^{N}(\frac{1}{2}y_n^n(\xi(t_n))h)h = \sum_{n=1}^{N} g(t_n)h = \overline{g}\cdot L = \frac{1}{2}\overline{y}_n^n(\tau)hL$$

where $t_0 < \tau < t_N$. Thus, $e_N = O(h)$: first-order accuracy in globally.

Stability of the explicit Euler's method

- Consider the following ODE

$$y' = \frac{dy}{dt} = \lambda y \text{ with } \mathrm{Re}(\lambda) < 0$$

Applying Euler's method leads to

$$y_{n+1} = y_n + \lambda h y_n = y_n(1 + \lambda h).$$

Thus, the solution at time step n is

$$y_n = y_0(1 + \lambda h)^n.$$

For complex λh, we have

$$y_n = y_0(1 + \lambda_R h + i\lambda_I h)^n = y_0\sigma^n, \text{ where } \sigma = (1 + \lambda_R h + i\lambda_I h) : \text{ amplification factor.}$$

The numerical solution is called 'stable' if

$$|\sigma| \leq 1.$$

The region of stability for Euler's method is

$$|\sigma|^2 = (1 + \lambda_R h)^2 + (\lambda_I h)^2 = 1.$$

If λ is real and negative, then the maximum step size for stability is

$$h \leq \frac{2}{|\lambda|} \rightarrow \text{ stability condition 'limits' the step size.}$$

- Explicit Euler's method is *conditionally stable*

Implicit Euler's method

equivalent rectangle integration rule w right end pt

$$y_{n+1} = y_n + hf(y_{n+1}, t_{n+1}), \quad n = 0, 1, 2 \cdots$$

- If f is non-linear, we must solve a non-linear algebraic equation at each time step to obtain y_{n+1}. → an 'iterative' or 'local linearization' algorithm
- Computational cost per time step is higher than the explicit method, but ...

Stability of the implicit Euler's method

- Consider the following ODE

$$y' = \frac{dy}{dt} = \lambda y \text{ with } \mathrm{Re}(\lambda) < 0$$

Applying the backward Euler scheme, we obtain

$$y_{n+1} = y_n + \lambda h y_{n+1} \to y_{n+1} = (1-\lambda h)^{-1} y_n \text{ or } y_n = \sigma^n y_0, \text{ where } \sigma = \frac{1}{(1-\lambda h)}.$$

Considering complex λ,

$$\sigma = \frac{1}{(1-\lambda_R h - i\lambda_I h)} = \frac{1}{(Ae^{i\theta})}, \text{ where } A = \sqrt{(1-\lambda_R h)^2 + \lambda_I^2 h^2}, \ \theta = -\tan^-1(\frac{\lambda_1 h}{1-\lambda_R h}).$$

For stability, $|\sigma| = \frac{|e^{-i\theta}}{A} = \frac{1}{A} \leq 1.$ → *Unconditionally stable*

Ex 1) Solve the following ODE using Euler's method

$$y' + 0.5y = 0$$

$$y(0) = 1 \quad 0 \leq t \leq 20. \text{ (step size } h = 1, 4, 2)$$

- Explicit Euler's method

  λ is real and negative. The explicit Euler's method should be stable for $h \leq 4$. The solution is advanced by

$$y_{n+1} = y_n - 0.5hy_n.$$

- Implicit Euler's method

  The implicit Euler's method should be unconditionally stable. The solution is advanced by

$$y_{n+|} = \frac{y_n}{1+0.5h}.$$

3.2.3 Improved Euler Method

Weakness of Euler's method

- Simple but low accuracy

  By taking more terms in following equation into account, we can obtain a higher order method.

$$y(t+h) = y(t) + hy'(t) + \frac{h^2}{2}y''(t) + \cdots$$

$$\text{If } y' = f(t, y(t)), \text{ then } y(t+h) = y(t) + hf + \frac{h^2}{2}f' + \frac{h^3}{6}f'' \cdots$$

since y in f depends on t,

$$f' = \frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial y}\frac{dy}{dt} = f_t + f_y f,$$

$$f'' = \frac{\partial}{\partial t}[f_t + f_y f] + \frac{\partial}{\partial y}[f_t + f_y f]f = f_{tt} + 2f_{yt}f + f_t f_y + f_y^2 f + f_{yy}f^2$$

The number of terms increases rapidly, and the method is not practical for higher than third order

General strategy for higher order method

- Optimize computational cost
- Making the order of the method as high as possible

Trapezoidal method

- The formal solution of $y' = f(y,t),\ y(0) = y_0$ at $t = t_{n+1}$ is

$$y_{n+1} = y_n + \int_{t_n}^{t_{n+1}} f(y,t)dt$$

- Approximating the integral with the trapezoidal method leads to

$$y_{n+1} = y_n + \frac{h}{2}[f(y_n, t_n) + f(y_{n+1}, t_{n+1})].$$

- Order of accuracy

one of improved euler

Applying the trapezoidal method to $y' = \lambda y$ yields

$$y_{n+1} - y_n = \frac{h}{2}[\lambda y_n + \lambda y_{n+1}] \text{ or } y_{n+1} = \frac{1+\lambda\frac{h}{2}}{1-\lambda\frac{h}{2}}y_n.$$

Expanding the amplification factor σ leads to

$$\sigma = \frac{1+\lambda\frac{h}{2}}{1-\lambda\frac{h}{2}} = 1 + \lambda h + \frac{\lambda^2h^2}{2} + \frac{\lambda^3h^3}{4} + \cdots$$

The exact solution is

$$y(t) = y_0e^{\lambda t} = y_0e^{\lambda nh} = y_0(e^{\lambda h})^n$$

We compare σ with $e^{\lambda h}$ using Taylor series

$$e^{\lambda h} = 1 + \lambda h + \frac{\lambda^2h^2}{2} + \frac{\lambda^3h^3}{6} + \cdots$$

Stability of the trapezoidal method

- Amplification factor σ for complex $\lambda = \lambda_R + i\lambda_I$

$$\sigma = \frac{1 + \frac{\lambda h}{2}}{1 - \frac{lambdah}{2}} = \frac{1 + \frac{\lambda_R h}{2} + i\frac{\lambda_I h}{2}}{1 - \frac{\lambda_R h}{2} - i\frac{\lambda_I h}{2}} = \frac{A}{B} e^{i(\theta - \alpha)},$$

  where

$$A = \sqrt{(1 + \frac{\lambda_R h}{2})^2 + \frac{\lambda_I^2 h^2}{4}},\ B = \sqrt{(1 - \frac{\lambda_R h}{2})^2 + \frac{\lambda_I^2 h^2}{4}}$$

  Thus,

$$|\sigma| = \frac{A}{B}.$$

- Since we are only interested in cases where $\lambda_R < 0$, and for these cases $A < B$, it follows that $|\sigma| < 1$.

- Trapezoidal method is *unconditionally stable*, which is expected since it is an implicit method.

Ex 2) Solve the following second-order equation using the explicit Euler, implicit Euler, and trapezoidal methods.

$$y'' + \omega^2 y = 0,\ t > 0$$

$$y(0) = 1,\ y'(0) = 0,\ \omega = 4,\ h = 0.15.$$

- We can convert this second-order equation to two first-order equations.

$$y_1' = y_2.\ y_2' = -\omega^2 y_1\ \rightarrow$$

For explicit Euler:
For implicit Euler:
For trapezoidal:

3.2.4 Runge Kutta Method

Runge-Kutta method

- One of the most popular (explicit) methods to solve $y' = \frac{dy}{dt} = f(y,t),\ y(t_0) = y_0$

- Introduce intermediate points between t_n and t_{n+1} to accurately evaluate f

Second-order R-K method

- General form of the second-order R-K formula

$$\underline{y_{n+1} = y_n + \gamma_1 k_1 + \gamma_2 k_2}$$

Solution at time step t_{n+1} is obtained from

$$k_1 = hf(y_n, t_n)$$

$$k_2 = hf(y_n + \underline{\beta k_1}, t_n + \underline{\alpha h})$$

where k_1 and k_2 are defined sequentially

Choice of the interior point

- Taylor series expansion of y_{n+1}

$$y_{n+1} = y_n + hy'_n + \frac{h^2}{2} y''_n + \cdots \text{ with } y'_n = f(y_n, t_n)$$

$$y'' = f' = f_t + f_y f,$$

$$y_{n+1} = y_n + hf(y_n, t_n) + \frac{h^2}{2}(f_{t_n} + f_{y_n} f_{t_n}) + \cdots$$

- Taylor series expansion for second-order R-K formula

Taylor series expansion of k_2 2variable taylor expansion

$$k_2 = h[f(y_n, t_n) + \beta k_1 f_{y_n} + \alpha h f_{t_n} + O(h^2)].$$

Substituting in $y_{n+1} = y_n + \gamma_1 k_1 + \gamma_2 k_2$ yields

$$y_{n+1} = y_n + (\gamma_1 + \gamma_2) hf + \gamma_2 \beta h^2 f_{t_n} f_{y_n} + \gamma_2 \alpha h^2 f_{t_n} + \cdots$$

- Comparison of the two Taylor series

Matching coefficients of similar terms leads to

$$\gamma_1 + \gamma_2 = 1 \; \gamma_2 \alpha = \frac{1}{2}, \; \gamma_2 \beta = \frac{1}{2} \Rightarrow \gamma_2 = \frac{1}{2}\alpha, \; \beta = \alpha, \; \gamma_1 = 1 - \frac{1}{2\alpha}$$

- A one-parameter family of second-order R-K formula

$$k_1 = hf(y_n, t_n)$$

$$k_2 = hf(y_n + \alpha k_1, t_n + \alpha h)$$

$$y_{n+1} = y_n + \left(1 - \frac{1}{2\alpha}\right) k_1 + \frac{1}{2\alpha} k_2$$

Choice of $\alpha = \frac{1}{2}$ is most popular

y_(n+1) = y_n + hf(y_n + 1/2hf(y_n,t_n), t_n +1/2h)

- R-K formulas expressed in a different but equivalent form

  Popular form of the second-order R-K formula $(\alpha = \frac{1}{2})$ is

$$y*_{n+\frac{1}{2}} = y_n + \frac{h}{2} f(y_n, t_n)$$

$$y_{n+1} = y_n + hf(y*_{n+\frac{1}{2}}, t_{n+\frac{1}{2}})$$

→ Predictor-corrector method (or Heun's method)

One calculates the *predicted* value $y*_{n+\frac{1}{2}}$ which is then used $f(y*_{n+\frac{1}{2}}, t_{n+\frac{1}{2}})$ to obtain the *corrected* value, y_{n+1}.

Fourth-order Runge-Kutta method

- The most widely used explicit R-K method
- The fourth-order R-K formula after four Taylor series expansions followed by matching coefficients

$$y_{n+1} = y_n + \frac{1}{6}k_1 + \frac{1}{3}(k_2 + k_3) + \frac{1}{6}k_4,$$

$$\begin{cases} k_1 = hf(y_n, t_n) \\ k_2 = hf(y_n + \frac{1}{2}k_1, t_n + \frac{h}{2}) \\ k_3 = hf(y_n + \frac{1}{2}k_2, t_n + \frac{h}{2}) \\ k_4 = hf(y_n + k_3, t_n + h). \end{cases}$$

- Applying the method to the model equation, $y' = \lambda y$, leads to

$$y_{n+1} = \left(1 + \lambda h + \frac{\lambda^2 h^2}{2!} + \frac{\lambda^3 h^3}{3!} + \frac{\lambda^4 h^4}{4!}\right) y_n$$

  The exact solution of $y' = \lambda y$

$$y(t) = y_0 e^{\lambda t} = y_0 e^{\lambda nh} = y_0 (e^{\lambda h})^n \text{ with } e^{\lambda h} = 1 + \lambda h + \frac{\lambda^2 h^2}{2!} + \frac{\lambda^3 h^3}{3!} + \frac{\lambda^4 h^4}{4!} \cdots$$

Ex 3) Solve the following equation using second and fourth-order R-K methods

$$y'' + \omega^2 y = 0 \; t > 0$$

$$y(0) = 1,\ y'(0) = 0,\ \omega = 4,\ h = 0.15.$$

- Second-order R-K advancement

$$y_1' = y_2,\ y_2' = -\omega^2 y_1$$

$$y_1^{(n+\frac{1}{2})} * = y_1^{(n)} + \frac{h}{2} y_2^{(n)}$$

$$y_2^{(n+\frac{1}{2})}* = y_2^{(n)} - \frac{h}{2}\omega^2 y_1^{(n)}$$

$$\Rightarrow \begin{cases} y_1^{(n+1)} = y_1^{(n)} + hy_2^{(n+\frac{1}{2})*} \\ y_2^{(n+1)} = y_2^{(n)} - h\omega^2 y_1^{(n+\frac{1}{2})*} \end{cases}$$

- Similarly for Fourth-order R-K advancement

3.2.5 Multi-Step Method

Characteristics of multi-step methods

- Multi-step method is to use information from prior to step n to construct a polynomial that approximates the derivative function
- this method is not self-starting. → usually using another method such as the explicit Euler method

The leapfrog method

- This method derived by applying the central difference formula.

$$\begin{cases} y_{n+1} = y_n + hy'_n + \frac{h^2}{2}y''_n + \cdots \\ y_{n-1} = y_n - hy'_n + \frac{h^2}{2}y''_n - \cdots \end{cases} \rightarrow y_{n+1} - y_{n-1} = 2hy'_n + 2\frac{h^3}{6}y'''_n + \cdots$$

$$y_{n+1} = y_{n-1} + 2hf(y_n, t_n) + O(h^3) \ \rightarrow \ \text{second-order accurate globally}$$

Starting with an initial condition y_0, a self-starting method like Euler's or R-K method is used to obtain y_1

Then the leapfrog method is used for steps two and higher

Adams-Bashforth method

- Taylor series expansion of y_{n+1}

$$y_{n+1} = y_n + hy'_n + \frac{h^2}{2}y''_n + \cdots$$

Substituting

$$y'_n = f(y_n, t_n)$$

and a first-order finite difference approximation for y''_n

$$y''_n = \frac{f(y_n, t_n) - f(y_{n-1}, t_{n-1})}{h} + O(h)$$

leads to

$$y_{n+1} = y_n + \frac{3h}{2}f(y_n, t_n) - \frac{h}{2}f(y_{n-1}, t_{n-1}) + O(h^3)$$

→ second-order accurate globally

3.2.6 System of First-Order ODEs

Conversion of higher order ODEs

- Higher order ordinary differential equations can be converted to a system of first-order ODEs
- System of ODEs appear in many physical situations

  Chemical reactions among several species

  Vibration of a complex structure with several elements

- Generic form

$$\mathbf{y}' = \mathbf{f}(\mathbf{y}, \mathbf{t}), \ \ \mathbf{y}(\mathbf{0}) = \mathbf{y_0}$$

  where $\mathbf{y}$ is a vector with elements y_i and $\mathbf{f}(\mathbf{y_1}, \mathbf{y_2}, \mathbf{y_3}, \cdots, \ \mathbf{y_m}, \mathbf{t})$ is a vector function with elements $f_i(y_1, y_2, y_3, \cdots, \ y_m, t), \ i = 1, 2, 3, \cdots, \ m$.

- Example

  The explicit Euler's method

$$y_i^{(n+1)} = y_i^{(n)} + h f_i \left(y_1^{(n)}, y_2^{(n)}, \cdots, \ y_m^{(n)}, t_n \right), \ i = 1, 2, 3, \cdots, \ m$$

3.2.7 Boundary Value Problems

Boundary value problem

- The data associated with a differential equation are prescribed at more than one value of the independent variable
- To have a BVP, we must have at least a second-order differential equation

$$y'' = f(x, y, y'), \ y(0) = y_0, \ y(L) = y_L,$$

  where f is an arbitrary function.

  Here, the data are prescribed at $x = 0$ and at $x = L$.

Techniques for solving BVP

- Shooting method

  Iterative technique which uses the standard method for IVP such as R-K methods

- Direct methods

  These methods are based on straightforward finite differencing of the derivatives in the differential equation.

3.2.8 Shooting Methods

Approach

- Convert a BVP into an IVP
- Solve the resulting problem iteratively (trial and error)
- Linear ODEs allow a quick linear interpolation
- Non-linear ODEs will require an iterative approach similar to our root finding techniques

For linear problems

- Consider the general second-order linear equation

$$y''(x) + A(x)y'(x) + B(x)y(x) = f(x),$$

$$y(0) = y_0,\ y(L) = y_L$$

- Convert a BVP to an IVP

$$u = y,\ v = y'$$

$$\Big\{ u' = v\ v' = f(x, u, v) \quad , u(0) = y_0,\ u(L) = y_L$$

Let's denote two solutions of the equation as $u_1(x)$ and $u_2(x)$, which are obtained using $u_1(0) = u_2(0) = u(0) = y_0$, and two different initial guesses for $v(0)$.

Since the differential equation is linear, the exact solution can be formed as a linear combination of y_1 and y_2

$$u(x) = c_1u_1(x) + c_2u_2(x)$$

provided that

$$c_1 + c_2 = 1,\ c_1u_1(L) + c_2u_2(L) = y_L\ \rightarrow\ c_1 = \frac{y_L - u_2(L)}{u_1(L) - u_2(L)},\ c_2 = \frac{u_1(L) - y_L}{u_1(L) - u_2(L)}.$$

Substitution for c_1 and c_2 into $u(x) = c_1u_1(x) + c_2u_2(x) = y(x)$ gives the desired solution

For non-linear problems

- Linear interpolation between two solutions will not necessarily result in a good estimated of the required boundary conditions
- Recast the problem as a root finding problem
- We may have to perform several iterations to obtain the solution at L within an acceptable tolerance.

- Solution procedure

  Guess an initial value of $v(0)$ just as in the linear problem

  Using R-K or some other ODE method, we will obtain a solution at L.

  Denote the difference between the result from the integration and boundary conditions

$$m = utrue(L) - u_{guess}(L)$$

  Check to see if m is within an acceptable tolerance.

$$\varepsilon \leq \varepsilon_{acceptabletolerance}$$

$$\varepsilon = |\frac{m_i - m_{i-1}}{m_i}|$$

  If not, then use the secant method to determine our next guess

- The secant method

$$u = y, \ v = y'$$

$$\Big\{ u' = v \ v' = f(x, u, v) \quad , u(0) = y_0, \ u(L) = y_L$$

  Suppose that we use two initial guesses, $v_1(0), v_2(0) \rightarrow$ get $u_1(L), \ u_2(L)$

  The straight line between the points $(v_1(0), u_1(L))$ and $(v_2(0), u_2(L))$

$$v(0) = v_2(0) + m[u(L) - u_2(L)],$$

  where

$$m = \frac{v_1(0) - v_2(0)}{u_1(L) - u_2(L)}$$

  Next guess is the value for $v_3(0)$ at the above line.

$$v_3(0) = v_2(0) + m[y_L - u_2(L)].$$

  General form

$$v_{\alpha+1}(0) = v_\alpha(0) + m_{\alpha-1}[y_L - u_\alpha(L)],$$

  where $\alpha = 1, 2, 3, \cdots$ is the iteration index and

$$m\alpha - 1 = \frac{v_\alpha(0) - v_{\alpha-1}(0)}{u_\alpha(L) - u_{\alpha-1}(L)}$$

Algorithm

3.2.9 Direct Methods

Finite difference approximation

- Second-order approximation to the linear differential equation yields

$$y''(x) + A(x)y'(x) + B(x)y(x) = f(x),\ y(0) = u_0,\ y(L) = y_L \rightarrow$$

$$\frac{y_{j+1} - 2y_j + y_{j-1}}{h^2} + A_j\frac{y_{j+1} - y_{j-1}}{2h} + B_jy_j = f_j,\ y_{(j=0)} = y_0,\ y_{(j=L)} = y_L$$

Rearranging the terms yields

$$\alpha_jy_{j+1} + \beta_jy_j + \gamma_j\gamma_{j-1} = f_j,$$

where

$$\alpha_j = \left(\frac{1}{h^2} + \frac{A_j}{2h}\right),\ \beta_j = \left(B_j + \frac{2}{h^2}\right),\ \gamma_j = \left(\frac{1}{h^2} + \frac{A_j}{2h}\right)\quad j = 1, 2, \cdots N - 1.$$

Tridiagonal system Solve using standard linear system solver

4 Analytic methods for PDEs

A PDE is an equation involving one or more partial derivatives of an (unknown) function, called it u that depends on two or more variables, often time t and one or several variables in space. The order of the highest derivetive is called 'order' of the PDE. Also, 2nd order PDEs are the most important ones in applications. Linear, homogeneous are similary defined as in ODEs.

Theorem 8. *If u_1 and u_2 are solutions of a homogeneous linear PDE in some region R, then $u = c_1u_1 + c_2u_2$ is also a sol of the PDE in R.*

4.1 Main Model Problems - Partial Differential Equations

4.1.1 First-order partial differential equations

- linear PDE

$$\frac{\partial u(x,t)}{\partial t} = 0 \quad \text{for all } x \in R, \quad t > 0; u(x,0) = f(x) \tag{4.1}$$

Solution
Integrate $\frac{\partial u}{\partial t}(x,t)$ with respect to t.

$$0 = \int_0^t \frac{\partial u}{\partial t}(x,s)ds = u(x,t) - u(x,0) = u(x,t) - f(x)$$

Hence

$$u(x,t) = f(x) \quad \forall x \in R; \quad \forall t > 0$$

- one way wave equation

$$\frac{\partial u(x,t)}{\partial t} + c\frac{\partial u(x,t)}{\partial x} = 0, \text{for all } x \in R, t > 0; u(x,0) = f(x) \tag{4.2}$$

If $c = c(x,t)$, still linear, need to use the Method of characteristics which is out of scope.
Solution
Notice that if there exists a function $\phi(x)$ such that $u(x,t) = \phi(x-ct)$ which is continuously differentiable, $\phi(x)$ satisfies the PDE. Indeed by the chain rule

$$\frac{\partial u}{\partial t} = \phi'(x-ct)\cdot(-c)\ , \ \ \frac{\partial u}{\partial x} = \phi'(x-ct).$$

Hence,

$$\frac{1}{c}\frac{\partial u}{\partial t}(x,t) + \frac{\partial u}{\partial x} = 0 \quad \forall x \in \mathbf{R}, \ \ \forall \mathbf{t} > \mathbf{0}.$$

Next, trim with the initial condition

$$u(x,0) = f(x) = \phi(x), \ \forall x \in \mathbf{R}$$

Hence such a function$\phi(x)$ must be $f(x)$.
Therefore,

$$u(x,t) = f(x-ct)$$

is the solution.
Notice that this is the wave propagating to the right at the speed of c. Indeed,

$$u(x + c\Delta t, t + \Delta t) = f((x + c\Delta t) - c(t + \Delta t)) = f(x - ct) = u(x,t)$$

Any information at x at time t propagates to $(x + c\Delta t)$ at time $t + \Delta t$.
Remark If there is a change in the sign in the PDE as

$$\frac{1}{c}\frac{\partial u}{\partial t}(x,t) - \frac{\partial u}{\partial x}(x,t) = 0, \ (x,t) \in \mathrm{Re}_x \ \mathrm{Re}_t$$

the solution is given by

$$u(x,t) = f(x+ct).$$

This wave propagates to the left with the same speed.

- nonlinear Burger's equation

$$\frac{\partial u(x,t)}{\partial t} + u(x,t)\frac{\partial u(x,t)}{\partial x} = 0, \text{for all } x \in R, t > 0; u(x,0) = f(x) \tag{4.3}$$

- The general nonlinear PDE of first order is of the form;

$$f(t,x,u,\frac{\partial u(x,t)}{\partial t}, \frac{\partial u(x,t)}{\partial x}) = 0. \tag{4.4}$$

- The semilinear first order PDE is of the form;

$$a(x,t,u)\frac{\partial u(x,t)}{\partial t} + b(x,t,u)\frac{\partial u(x,t)}{\partial x} = c(x,t,u). \tag{4.5}$$

4.1.2 Second-order partial differential equations

Types of PDE (mathematically) in quasilinear form

$$Au_{xx} + 2Bu_{xy} + Cu_{yy} = F(x, y, u, u_x, u_y) \tag{4.6}$$

is called

1. Hyperbolic if $AC - B^2 < 0$, wave equation, $u_{tt} = c^2 u_{xx}$
2. Parabolic if $AC - B^2 = 0$, heat equation, $u_t = c^2 u_{xx}$
3. Elliptic if $AC - B^2 > 0$, laplace equation, $u_{xx} + u_{yy} = 0$, poisson equation, $u_{xx} + u_{yy} = f$.

- Wave equation

  Modelling small transverse vibrations of an elastic string, such as violin string

  We place the string along x-axis, stretch it to length L, and fasten it at the ends $x = 0$ and $x = L$. Then distort it, and at some instant, call it $t = 0$, release it and allow it to vibrate. The problem is to determine the vibrations of the string, (to fine its deflection $u(x, t)$ at any pt and time x, t, resp.)

  Physical Assumptions:

  - The mass of the string per unit length is constant. The string is perfectly elastic and does not offer any resistance to bending.
  - The tension caused by the stretching the string before fastening it at the ends is so large that the action of the gravitational force on tghe string can be neglected.
  - The string performs small transverse motions in a vertical plane; that is,every particle of the string moves strictly vertically and so that the deflection and the slope at every pt of the string always remain small in absolute value.

  Derivation of the PDE of the wave equation from forces

  Since the string offers no resistance to bending, the tension is tangential to the curve of the string at each pt. Let T_1, T_2 are the tension at the endpts. Since the string move vertically, no motion in the horizontally. Hence the horizontal components of the tension is constant,

  $$T_1 cos\alpha = T_2 cos\beta = T. \tag{4.7}$$

  In the vertical direction, we have $-T_1 sin\alpha, T_2 sin\beta$ of tensions applied on the endpts of the segement. By Newton's second law, the resultant of these two forces is equal to the mass $\rho\Delta x$ times the acceleration $\partial^2 u/\partial t^2$, evaluated at some pt btw x and $x + \Delta x$. Therefore

  $$T_2 sin\beta - T_1 sin\alpha = \rho\Delta x \frac{\partial^2 u}{\partial t^2}.$$

Using (4.7), we have

$$tan\beta - tan\alpha = \frac{\rho\Delta x}{T}\frac{\partial^2 u}{\partial t^2}. \tag{4.8}$$

Observe that $tan\alpha, tan\beta$ are the slopes of the string at x and $x + \Delta x$. Dividing (4.8) by Δx,we have

$$\frac{1}{\Delta x}\left[(\frac{\partial u}{\partial x})|_{x+\Delta x} - (\frac{\partial u}{\partial x})|_{x}\right] = \frac{\rho}{T}\frac{\partial^2 u}{\partial t^2}. \tag{4.9}$$

Let Δx approach zero, we obtain the linear PDE

$$u_{tt} = c^2 u_{xx}, \quad c^2 = T/\rho. \tag{4.10}$$

Solution by seperating variables, Use of Fourier Series

Consider the model of a vibrating elastic string in one dimension.

$$u_{tt} = c^2 u_{xx}, \quad c^2 = T/\rho. \tag{4.11}$$

with ICs and BCs.

$$u(0,t) = 0, u(L,t) = 0 \quad \text{for all t}, \qquad u(x,0) = f(x), u_t(x,0) = g(x) \quad (0 \le x \le L). \tag{4.12}$$

We will solve this in three steps,

1. Method of separating variables by setting $u(x,t) = E(x)G(t)$, we obtain 2 ODEs for $E(x)$ and $G(t)$.
2. Determine the solutions by using ICs and BCs.
3. Using Fourier Series, compose the solution for 1D wave equation.

D'Alembert's solution of the wave equation, characteristics

Introducing by $v = x + ct, w = x - ct$, then we obtain $u(x,t) = \phi(x+ct) + \psi(x-ct)$, which is d'Alembert's solution of the wave eq. Once ICs are given as above, we have a more intuitive final form

$$u(x,t) = \frac{1}{2}[f(x+ct) + f(x-ct)] + \frac{1}{2c}\int_{x-ct}^{x+ct} g(s)ds. \tag{4.13}$$

- Heat equation

Physics and derivation 1

The heat equation is derived from Fourier's law and conservation of energy. By Fourier's law, the flow rate of heat energy through a surface is proportional to the negative temperature gradient across the surface,

$$\mathbf{q} = -k\nabla u \tag{4.14}$$

where k is the thermal conductivity and u is the temperature. In one dimension, the gradient is an ordinary spatial derivative, and so Fourier's law is

$$\mathbf{q} = -ku_x. \tag{4.15}$$

In the absence of work done, a change in internal energy per unit volume in the material, ΔQ, is proportional to the change in temperature, Δu. That is,

$$\Delta Q = c_p \rho \Delta u, \tag{4.16}$$

where c_p is the specific heat capacity and ρ is the mass density of the material. Choosing zero energy at absolute zero temperature, this can be rewritten as

$$Q = c_p \rho u. \tag{4.17}$$

By the conservation of energy, the increase in internal energy in a small spatial region of the material $x - \Delta x \leq \xi \leq x + \Delta x$ over the time period $t - \Delta t \leq \tau \leq t + \Delta t$ is given by

$$c_p \rho \int_{x-\Delta x}^{x+\Delta x} [u(\xi, t+\Delta t) - u(\xi, t-\Delta t)]\, d\xi = c_p \rho \int_{t-\Delta t}^{t+\Delta t} \int_{x-\Delta x}^{x+\Delta x} \frac{\partial u}{\partial \tau}\, d\xi d\tau, \tag{4.18}$$

where the fundamental theorem of calculus was used. If no work is done and there are neither heat sources nor sinks, the change in internal energy in the interval $[x - \Delta x, x + \Delta x]$ is accounted for entirely by the flux of heat across the boundaries. By Fourier's law, this is

$$k \int_{t-\Delta t}^{t+\Delta t} \left[\frac{\partial u}{\partial x}(x + \Delta x, \tau) - \frac{\partial u}{\partial x}(x - \Delta x, \tau)\right] d\tau = k \int_{t-\Delta t}^{t+\Delta t} \int_{x-\Delta x}^{x+\Delta x} \frac{\partial^2 u}{\partial \xi^2}\, d\xi d\tau, \tag{4.19}$$

again by the fundamental theorem of calculus. By conservation of energy,

$$\int_{t-\Delta t}^{t+\Delta t} \int_{x-\Delta x}^{x+\Delta x} [c_p \rho u_\tau - k u_{\xi\xi}]\, d\xi d\tau = 0. \tag{4.20}$$

Therefore, for $[t - \Delta t, t + \Delta t] \times [x - \Delta x, x + \Delta x], c_p \rho u_t - k u_{xx} = 0$. Finally,

$$u_t = \frac{k}{c_p \rho} u_{xx}. \tag{4.21}$$

which is the heat equation, where the coefficient (often denoted) $\alpha = \frac{k}{c_p \rho}$ is called the thermal diffusivity.

Physics and derivation 2

In a metal rod with non-uniform temperature,heat(thermal energy) is transferred from regions of higher temperature to regions of lower temperature. Three physical principles are used here.

1. Heat(or thermal) energy of a body with uniform properties:

$$Heatenergy = cmu,$$

where m is the body mass, u is the temperature, c is the specific heat,

$$units[c] = L^2T^2U^1$$

. c is the energy required to raise a unit mass of the substance 1 unit in temperature.

2. Fouriers law of heat transfer:

rate of heat transfer proportional to negative temperature gradient,

$$\frac{Rateofheattransfer}{area} = -K\frac{\partial u}{\partial x},$$

where K isthethermal conductivity,$units[K0] = MLT^{-3}U^1$ .In other words, heat is transferred from areas of high temp to low temp.

3. Conservation of energy.

Consider a uniform rod of length l with non-uniform temperature lying on the x-axis from x =0 to x = l. By uniform rod, we mean the density ρ, specific heat c, thermal conductivity K, cross-sectional area A are ALL constant. Assume the sides of the rod are insulated and only the ends may be exposed. Also assume there is no heat source within the rod. Consider an arbitrary thin slice of the rod of width Δx between x and $x + \Delta x$. The slice is so thin that the temperature throughout the slice is $u(x, t)$. Thus,

$$Heatenergyofsegment = c \times \rho A\Delta x \times u = c\rho A\Delta xu(x, t).$$

By conservation of energy,

change of heat in from heat out from heat energy of segment in time Δt equals heat in from left boundary minus heat out from right boundary.

From Fouriers Law,

$$c\rho A\Delta xu(x, t + \Delta t) - c\rho A\Delta xu(x, t) = \Delta t(K\frac{\partial u}{\partial x})_x - \Delta t(K\frac{\partial u}{\partial x})_{x+\Delta x}.$$

Rearranging yields(recall ρ, c, A, K are constant),

$$\frac{u(x, t + \Delta t) - u(x, t)}{\Delta t} = \frac{K}{c\rho}\left(\frac{(\frac{\partial u}{\partial x})_{x+\Delta x} - (\frac{\partial u}{\partial x})_x}{\Delta x}\right).$$

Taking the limit $\Delta t, \Delta x$ goes to zero, gives the Heat Equation,

$$u_t = \frac{K}{c\rho}u_{xx}.$$

Since the slice was chosen arbitrarily,the Heat Equation applies throughout the rod.

Solution by Fourier Series

Consider the temperature in a long thin metal bar of constant cross section and homogeneous material, and is perfectly insulated laterally, so that heat flows in the x-direction only. Then u depends on x, t and becomes 1 dim heat equation.

$$u_t = c^2 u_{xx}, \quad c^2 = K/c_p\rho. \tag{4.22}$$

with ICs and BCs.

$$u(0,t) = 0, u(L,t) = 0\text{for all t}, \qquad u(x,0) = f(x) \quad (0 \leq x \leq L). \tag{4.23}$$

We will solve this in three steps as in wave equation,

1. Method of separating variables by setting $u(x,t) = F(x)G(t)$, we obtain 2 ODEs for $F(x)$ and $G(t)$.
2. Determine the solutions by using ICs and BCs.
3. Using Fourier Series, compose the solution for 1D heat equation.

- Laplace's equation

Physics and derivation

In wave equation or heat equation, consider the time independent steady problems, say $\partial u/\partial t = 0$. These reduce to Laplace's equation.

4.1.3 Third-order partial differential equations

A remarkable third order evolution equation that originally arose in the modeling of surface water waves serves to introduce yet further phenomena, both linear and nonlinear. The third-order derivative models dispersion, in which waves of different frequencies move at different speeds. Coupled with the same nonlinearity as in the inviscid and viscous Burgers, the result is one of the most remarkable equations in all of mathematics, with far-reaching implications, not only in fluid mechanics and applications, but also in complex function theory and physics.

$$u_t + u_{xxx} = 0 \qquad \text{for all } x \in R, t > 0; u(x,0) = f(x). \tag{4.24}$$

We can either solve this by using Fourier transform or try to substitute an exponential ansatz $u(x,t) = exp(iwt + ikx)$ representing a complex oscillatory wave of frequency w, which indicates the time vibrations, wave number k, which indicates the corresponding oscillations in space. We can obtain the dispersion relation($w = k^3$).

The Korteweg-deVris Equation-soliton

$$u_t + u_{xxx} + uu_x = 0 \quad \text{for all } x \in R, t > 0; u(x,0) = f(x). \tag{4.25}$$

5 Numerical Methods for PDE

6 Observational Models

6.1 Empirical Models

Sometimes it is difficult or impossible to develop a mathematical model that explains(models) a situation. However, if data exist, we can often use these data as the sole basis of an emphirical model. The emphirical consists of a function that fits the data. The graph of the function goes through the data points approximately. Thus although we can't employ an empirical model to explain a system, we can use such a model to predict behaviour where data do not exist. Data are crucial for an empirical model.

Data to

- suggest model
- estimate parameters
- test model

Empirical model

- based on data only
- to predict, not explain
- consists of a function that captures the trend of the data

Linear empirical model

Example 12. $xLst = 0, 2, 4, 5, 6, 8, yLst = 5.3, 7.0, 9.4, 11.1, 12.3, 14.2$ *How form a list of ordered pairs? Plot points? How do we find linear least squares fit for fitting function? Get* $y = 5.0449 + 1.16122x$

Linear regression finds m and b so that the sum of the squares of the errors is as small as possible. In other words, for n points, $(x_1, y_1), \cdots, (x_n, y_n)$,

minimize $f(m, b) = \sum_{i=1}^{n}(mx_i+b-y_i)^2$. This is done by finding m, b such that $\frac{\partial f}{\partial m} = \frac{\partial f}{\partial b} = 0$, yield

$$m = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}, \quad b = \frac{n\sum x_i^2 \sum y_i - \sum x_i y_i \sum x_i}{n\sum x_i^2 - (\sum x_i)^2}$$

Example 13. *In 1976 study: Population (P in millions) effects on quality of life,*

P = 341948, 1092759, 5491, 49375, 1340000, 365, 2500, 78200, 867023, 14000, 23700, 70700, 304500, 138000, 2602000

v = 4.81, 5.88, 3.31, 4.90, 5.62, 2.76, 2.27, 3.85, 5.21, 3.70, 3.27, 4.31, 4.42, 4.39, 5.05

Measure mean time (t) to walk 50 ft on main streets(Pace of life: velocity (ft/sec) = v = 50/t)

Form list of ordered pairs, plot points, how to pull pts to left to make graph appear linear ? Plot pts of form $(\sqrt{P}, v)$ or $(log(P), v)$

Fit linear line with independent variable x to transformed pts, get $y = 0.0409809 + 0.3739x$, plot line and transformed pts.

Always , eventually plot actual and predicted data together!!

and more

- Can we use pace of life to predict population?
- Can we use for city planning ?
- Can only trust to within an order of magnitude
- If “outliers”, may discard or repeat experiment

Strategies

- when concave down, try substituting for x by $\sqrt{x}, ln(x), -\frac{1}{\sqrt{x}}, -\frac{1}{x}, -\frac{1}{x^2}$(to have increasing, take the negatives) or for y by y^2, y^3.
- when concave up, try substituting for x to stretch to right by x^2, x^3, or for y by $\sqrt{y}, ln(y), -\frac{1}{\sqrt{y}}, -\frac{1}{y}, -\frac{1}{y^2}$.

7 monte carlo simulations

A monte carlo simulation is a probabilistic model involving an element of chance, thus not deterministic and uses random number generator. Disadvantages are

- The simulation maybe expensive in time or money to develop
- Bcz it is impossible to test every alternative, we can provide good solution but not a best solution
- Bcz a simulation is a probabilistic model involving an element of chance, we should be careful of our conclusions.
- Results may be difficult to verify bcz often we don not have real-workd data
- cannot be sure we understand what the simulation sctually does

7.1 Random number generator

Example 14. *Multiplicative Linear Congruential Method*

*random(0) = 10, random(n) = (7 * random(n - 1)) mod 11, for n > 0*

Generates? 10, 4, 6, 9, 8, 1, 7, 5, 2, 3

The general form for the Multiplicative Linear Congruential Method to generate pseudorandom number integers from 0 upto modulus is

random(0) = seed, random(n) = (multiplier * random(n - 1)) mod modulus, for n $>$ 0
modulus often largest integer comp. can store, such as $2^{31} - 1 = 2,147,483,647$, One multiplier: 16,807

Random floating point number

For random number, rand, with $0.0 \leq r < 1.0$

rand = random/modulus

Example r(0) = 10, r(n) = (7 * r(n - 1)) mod 11, for n $>$ 0

Integers: 10, 4, 6, 9, 8, 1, 7, 5, 2, 3,

Floating point: 10/11, 4/11, , 3/11

Random floating point number $min \leq r < max$

For $0 \leq rand < 1$,r = (max - min) rand + min

Example: random floating point number between 20.0 and 26.3? 6.3rand + 20.0

Random integer $min \leq r \leq max$

For 0 rand ¡ 1,n = int((max - min + 1) rand + min)

Example: random integer between 20 and 26, inclusively? int(7rand + 20)

7.2 Area through MCM

Although MC simulation is probabilistic, the technique can model determinstic behavior, such as area under a curve by providing a stochastic numeric integration technique.

Consider the problem of finding the area of $f(x)$ from $x = 0$ to $x = 2$.

Throwing Darts for Area

if random y $<$ f(random x between 0 and 2) 1; else 0

questions

How to calculate fraction (fractionUnder) of darts that hit under curve?

How to calculate area (rectArea) of rectangular dart board?

How to estimate for area under curve between 0 and 2?

better estimates

Better estimate if more darts

Better result if perform many times and obtain mean area

Standard deviation gives indication of how good

Perform many times and obtain mean area

7.3 Random numbers from various Distributions

MC requires the use of unbiased random numbers. The distribution of these numnbers is a description of portion of times each possible outcome or each possible range of outcomes occurs on the average over many trials. However, the distribution that simulation requires depends on the problem. Now we discuss the algorithms for generating random numbers from several types of distributions.

- Uniform distributions
  Generator is just as likely to return value in any interval

  In list of many random numbers, on the average each interval contains same number of generated values

  Methods for generating random numbers in other distributions depend on an ability to produce random numbers with a UD.

- Discrete distribution
  Distribution with discrete values

  Probability function (or density function or probability density function)

  Returns probability of occurrence of particular argument value

  To Generate Random Numbers in Discrete Distribution with Equal Probabilities for Each of n Events, Generate uniform random integer from a sequence of n integers, where each integer corresponds to an event

- Continuous distribution
  Distribution with continuous values

  Probability function (or density function or probability density function)

  Indicates probability that given outcome falls inside specific range of values

- Normal or Gaussian distribution
  Probability density function, where μ is mean and σ is the standard deviation

  $$\frac{1}{\sqrt{2\pi\sigma}}exp(-\frac{(x-\mu)^2}{2\sigma})$$

  Box-Muller-Gauss Method for Normal dist. with μ, σ

  Compute $bsin(a) + \mu$ and $bcos(a) + \mu$, where

  a = uniform random number in $[0, 2\pi)$

  rand = uniform random number in $[0, 1)$

  b = $\sigma\sqrt{-ln(rand)}$

- Exponential distribution
  Probability density functions:

  $f(t) = |r|e^{rt}$ with r $<$ 0 and t $>$ 0 or $f(t) = |r|e^{rt}$ with r $>$ 0 and t $<$ 0

  compute ln(rand)/r, where rand is a URNG in $[0, 1)$.

7.4 Random walk

- Apparently random movement of entity
- Cellular automaton (plural, automata) is type of computer simulation that is dynamic computational model and is discrete in space, state, and time
- Space is grid, or one-, two-, or three-dimensional lattice, or array, of sites
- Cell of lattice has state
- Number of states is finite
- Rules, or transition rules, specifying local relationships and indicating how cells are to change state, regulate behavior of system
- Examples : Brownian Motion

Algorithm for random walk, where at each time step entity goes diagonally in NE, NW, SE, or SW direction, and to return the distance

```
seed random number generator
let x, x0, y, and y0 be 0
let n be number of steps
let lst be list containing origin
do the following n times:
    let rand be a random 0 or 1
    if rand is 0
        increment x by 1
    else
        decrement x by 1
    let rand be a random 0 or 1
    if rand is 0
        increment y by 1
    else
        decrement y by 1
    append point (x, y) onto end of lst
create and display graphics of random walk
return distance btw first and last pts, $\sqrt{(x-x0)^2+(y-y0)^2}$
```

Average distance coveded

How average over many runs?

How compute average distance in n steps, where n varies from 0 to 50?

How determine model fit to determine relationship of distance versus n, number of steps?

7.5 Diffusion-spreading of fire

This developes a 2 dim computer simulation for the spread of fire, can be extended to numerous other examples involving contagion, such as the propagation of diseases, heat diffusion, distribution of pollution.

- Initializing the system

  Grid-site values:

  probTree = probability of grid site occupied by tree; i.e., tree density

  probBurning = If tree, probability that tree is burning; i.e., fraction of burning trees

  How initialize n-by-n grid, forest?

  0 - empty, 1 - non-burning tree, 2 - burning tree

- Cell Initializing Algorithm

  If rand $<$ probTree

  If another rand <probBurning; Assign BURNING to the cell

  Else; Assign TREE

  Else ; Assign EMPTY

- updating rules

  von Neumann neighborhood

  Function spread(site, N, E, S, W) returns next value of site, based on values in locations site, N, E, S, and W

  If a site is empty, next ?

  If a site is burning, next ? burns down thus empty

  If a site is tree, next ? tree or burning

  For this purpose, introduce additional probabilities; probImmune, probLightning

  if site is tree and (N,E,S,w is burning)

  if rand $<$ probImmune ; tree

  else burning

  What about boundaries?

  Apply periodic boundary conditions in both x and y directions.

  extendLat1(lat)

  Function to accept a grid and to return a grid extended one cell in each dir with PBC

  Pre: $n \times n$ **grid**

  Algorithm

Post:$(n + 2) \times (n + 2)$ **grid** Main MC fire diffusion Algorithm

fire(n, probTree, probBurning, chanceLightning, chanceImmune, t)

Ft to return a list of grids in a simulation of thr spread of a fire in a forest, where a cell value of EMPTY, TREE and BURNING

Pre:

n is the size of the square grid

ProbTree is the prob that a site is initially occupied by tree

ProbBurning is the initial burning prob

chanceLightning is the prob of lightning hitting a site

t is the number of time steps

spread is the ft for the updating rules at each grid pts

Post: A list of grid at each time step

Algorithm:

```
global probLightning $< -$ chanceLightning, probImmune $< -$ chanceImmune
Initialize forest to be n-by-n grid of values, EMPTY (no tree), TREE (non-burning tree), or BURNING (burning tree), where probTree is probability of tree and probBurning is probability that tree is burning
grids $< -$list containing forest
do the following t times:
forestExtended $< -$ extendLat1(forest)
forest $< -$ call applyExtended
grids $< -$ list with forest appended onto grids
return grids
```

7.6 Movement of Ants

- Analysis of Problem

  Ants emitting pheromones when carrying food

  Simulate movements of ants in presence of chemical trail

  Group of ants as a whole can exhibit self-organizing behavior that makes group appear to have single consciousness

- Formulating a model

1. Gather data

   Employ empirical observations of ant species that leave peromone trails.

   With each step, an ant tends to turn to and move in the direction of the greatest amount of chemical.

   As time passes with no ant in a location, the amount of peromone diminishes there.

2. Make simplifyimg assumptions

   Assume an ant tends to move in the direction of the greatest amount of chemicalnand with that movement the ant deposits additional peromone.

   The chemical dissipates with time, start with a straight trail of increasing amounts of peromones. Do not consider food or a nest.

3. Determine variables

   Each cell of the gris contains value of ordered pair of integers 1st component - level of chemical attractant, nonnegative integer 2nd component (EMPTY - cell with no ant, NORTH - north-facing ant, EAST - east-facing ant, SOUTH - south-facing ant, WEST - west-facing ant)

   Initialize the grid with zero amount of chemical except for a trail with gradated amounts of chemical, the prob that an ant initially occupies a cell is probAnt, for a cell with an ant, we choose a random direction.

4. Establish relations and submodels

   Ant movement for one step consists of 2 actions, sensing and walking. First, ant tests the nbhr sites and turns to the maximum pheromone. Then if possible to do so wo colliding with another ant, the ant moves to that location.

5. Determine fts-Sensing

   Consider Von Neumann nhbd. sense(site, N,E,S,W) returns next value of site, based on values in parameter locations.

   Empty cell does not sense.

   Ant turns in direction of neighboring cell with greatest amount of chemical. In case of more than one neighbor having maximum amount, ant picks direction at random towards one of these cells

6. Determine fts-Walking w/o concern for collision

   Consider Moore nhbd. walk(site, N, E, S, W, NE, SE, SW, NW, Nn, Ee, Ss, Ww) returns next value of site, based on values in parameter locations.

   If no ant, decrease strength of scent by 1, but not lower than 0.

   If an ant leaves a cell, increase amount of chemical in site vacated by 1.

   If an ant stays in a cell, no change in chemical amount.

   If an ant moves into a cell, the cell continues to have its same amount of chemical.

   With an ant moves to a new location, the animal faces in the same direction it did before moving.

7. Determine fts-Walking w concern for collision

   If movement would cause collision, do not vacate site and do not change amount of chemical at site.

   For example, a collision can occur for a north facing ant at a site under any of the listed circumstances;(An ant is in the N cell. An east(south, west) facing ant is in the cell to NW(Nn, NE) of site)

   Another collision occurs if 2 ants want to move into site, which is empty. In this case, site remains empty(do not allow either to do so). The amount of chemical is positive, decreased by one. If no chemical is present, chemical remains 0.

   Periodic boundary conditions in this case-¿MIMIC INFINITE SYSTEM. Extend boundary by 1 for sense(ExtendLat1) and by 2 for walk(ExtendLat2).

- Solve a model

  Algirithm for Ants(grid,t)

```
Initialize gridList be list containing grid
do the following t times:
elat1 $<-$ call extendLat1 to get grid extended by 1 cell in each direction
gridSense $<-$ call applyExtended1 to apply sense to each internal cell of elat1
elat2 $<-$ call extendLat2 to get gridSense extended by 2 cells in each direction
grid $<-$ call applyExtended2 to apply walk to original internal cells of elat2
gridList $<-$ gridList with grid appended
return gridList
```

- Verifying and Interpreting -visualization

  The number of ants is conserved.

  A scientific visualization definitely help in interpreting and analying your result.