M442 Lecture Notes

P. Howard

Fall 2003

Contents

1 Overview

All modeling projects begin with the identification of a situation of one form or another that appears to have at least some aspect that can be described mathematically. The first two steps of the project, often taken simultaneously, become: (1) gain a broad understanding of the situation to be modeled, and (2) collect data. Depending on the project, (1) and (2) can take minutes, hours, days, weeks, or even years. Asked to model the rebound height of a tennis ball, given an initial drop height, we immediately have a fairly broad understanding of the problem and suspect that collecting data won't take more than a few minutes with a tape measure and a stopwatch. Asked, on the other hand, to model the progression of Human Immunodeficiency Virus (HIV) as it attacks the body, we might find ourselves embarking on lifetime careers.[1]

[1] At one point or another this semester, we will study each of these problems.

2 Curve Fitting and Parameter Estimation

Often, the first step of the modeling process consists of simply looking at data graphically and trying to recognize trends. In this section, we will study the most standard method of curve fitting and parameter estimation: the method of least squares.

Example 2.1. Suppose the Internet auctioneer, eBay, hires us to predict its net income for the year 2003, based on its net incomes for 2000, 2001, and 2002. They provide us with the following table:

Year	Net Income
2000	48.3 million
2001	90.4 million
2002	249.9 million

Table 2.1: Yearly net income for eBay.

We begin by simply plotting this data as a *scatterplot* of points. In MATLAB, we develop Figure 2.1 through the commands,

```
year=[0 1 2];
income=[48.3 90.4 249.9];
plot(year,income,'o')
axis([-.5 2.5 25 275])
```

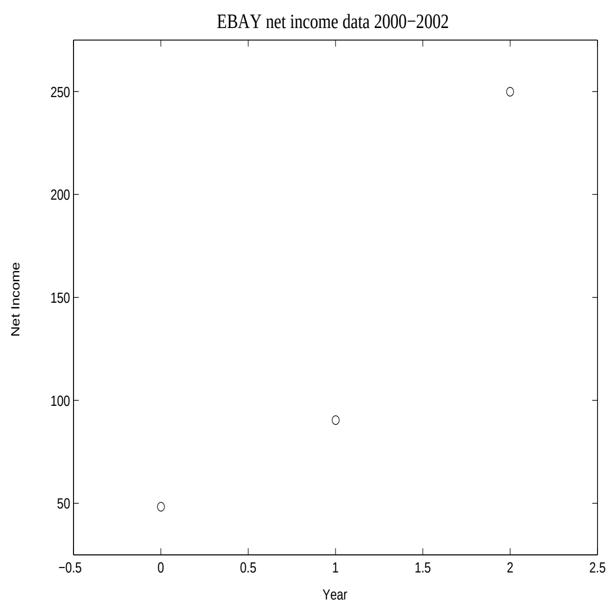

Figure 2.1: Net Income by year for eBay.

Our first approach toward predicting eBay's future profits might be to simply find a curve that best fits this data. The most common form or curve fitting is *linear least square regression.* △

2.1 Polynomial Regression

In order to develop an idea of what we mean by "best fit" in this context, we begin by trying to draw a line through these three points in such away that the distance between the points and the line is minimized (see Figure 2.2).

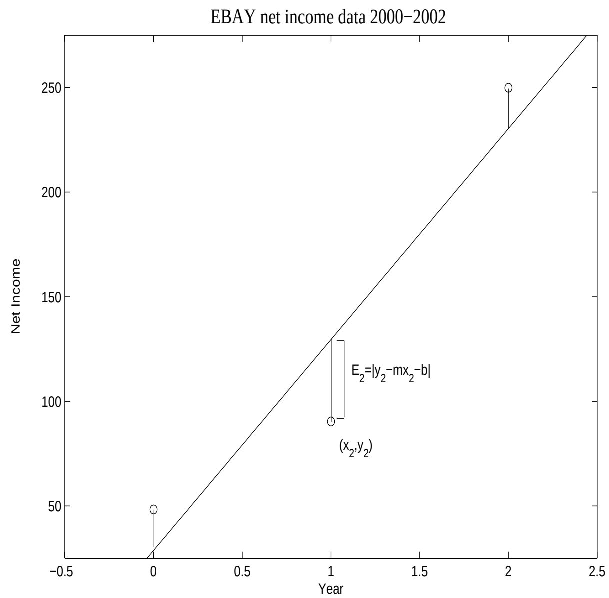

Figure 2.2: Least squares vertical distances.

Labeling our three points (x_1, y_1), (x_2, y_2), and (x_3, y_3), we observe that the distance between the line and the point (x_2, y_2) is given by the error $E_2 = |y_2 - mx_2 - b|$. The idea behind the least squares method is to sum these vertical distances and minimize the total error. In practice, we square the errors both to keep them positive and to avoid possible difficulty with differentiation (recall that absolute values can be subtle to differentiate), which will be required for minimization. Our total least squares error becomes

$$E(m, b) = \sum_{k=1}^{n} (y_k - mx_k - b)^2.$$

In our example, $n = 3$, though the method remains valid for any number of data points. In order to maximize or minimize a function of multiple variables, we compute the partial derivative with respect to each variable and set them equal to zero. Here, we compute

$$\begin{aligned}\frac{\partial}{\partial m}E(m,b) =&0\\ \frac{\partial}{\partial b}E(m,b) =&0.\end{aligned}$$

We have, then,

$$\begin{aligned}\frac{\partial}{\partial m}E(m,b) =& -2\sum_{k=1}^{n} x_k(y_k - mx_k - b) = 0,\\ \frac{\partial}{\partial b}E(m,b) =& -2\sum_{k=1}^{n} (y_k - mx_k - b) = 0,\end{aligned}$$

which we can solve as a linear system of two equations for the two unknowns m and b. Rearranging terms and dividing by 2, we have

$$\begin{aligned} m\sum_{k=1}^{n} x_k^2 + b\sum_{k=1}^{n} x_k &= \sum_{k=1}^{n} x_k y_k, \\ m\sum_{k=1}^{n} x_k + b\sum_{k=1}^{n} 1 &= \sum_{k=1}^{n} y_k. \end{aligned} \tag{2.1}$$

Observing that $\sum_{k=1}^{n} 1 = n$, we multiply the second equation by $\frac{1}{n}\sum_{k=1}^{n} x_k$ and subtract it from the first to get the relation,

$$m\Big(\sum_{k=1}^{n} x_k^2 - \frac{1}{n}(\sum_{k=1}^{n} x_k)^2\Big) = \sum_{k=1}^{n} x_k y_k - \frac{1}{n}(\sum_{k=1}^{n} x_k)(\sum_{k=1}^{n} y_k),$$

or

$$m = \frac{\sum_{k=1}^{n} x_k y_k - \frac{1}{n}(\sum_{k=1}^{n} x_k)(\sum_{k=1}^{n} y_k)}{\sum_{k=1}^{n} x_k^2 - \frac{1}{n}(\sum_{k=1}^{n} x_k)^2}.$$

Finally, substituting m into equation (2.1), we have

$$b = \frac{1}{n}\sum_{k=1}^{n} y_k - (\sum_{k=1}^{n} x_k)\frac{\sum_{k=1}^{n} x_k y_k - \frac{1}{n}(\sum_{k=1}^{n} x_k)(\sum_{k=1}^{n} y_k)}{n\sum_{k=1}^{n} x_k^2 - (\sum_{k=1}^{n} x_k)^2} = \frac{(\sum_{k=1}^{n} y_k)(\sum_{k=1}^{n} x_k^2) - (\sum_{k=1}^{n} x_k)(\sum_{k=1}^{n} x_k y_k)}{n\sum_{k=1}^{n} x_k^2 - (\sum_{k=1}^{n} x_k)^2}.$$

Observe that we can proceed similarly for any polynomial. For second order polynomials with general form $y = a_0 + a_1 x + a_2 x^2$, our error becomes

$$E(a_0, a_1, a_2) = \sum_{k=1}^{n} (y_k - a_0 - a_1 x_k - a_2 x_k^2)^2.$$

In this case, we must compute a partial derivative of E with respect to three parameters, and consequently (upon differentiation) solve three linear equations for the three unknowns.

The MATLAB command for polynomial fitting is *polyfit(x,y,n)*, where x and y are vectors and n is the order of the polynomial. For the eBay data, we have

```
>>polyfit(year,income,1)
ans =
100.8000 28.7333
```

Notice particularly that, left to right, MATLAB returns the coefficient of the highest power of x first, the second highest power of x second etc., continuing until the y-intercept is given last. Alternatively, for polynomial fitting up to order 10, MATLAB has the option of choosing it directly from the graphics menu. In the case of our eBay data, while Figure 1 is displayed in MATLAB, we choose **Tools, Basic Fitting**. A new window opens and offers a number of fitting options.

Example 2.2. (Crime and Unemployment.) Suppose we are asked to model the connection between unemployment and crime in the United States during the period 1994–2001. We might suspect that in some general way increased unemployment leads to increased crime ("idle hands are the devil's playground"), but our first step is to collect data. First, we contact the Federal Bureau of Investigation and study their Uniform Crime Reports (UCR), which document, among other things, the United States' crime rate per 100,000 citizens (we could also choose to use data on violent crimes only or gun-related crime only etc., each of which is a choice our model will make). Next, we contact the U.S. Bureau of Labor and obtain unemployment percents for each year in our time period. Summarizing our data we develop the following table.
Proceeding as with Example 2.1, we first look at a scatterplot of the data, given in Figure 2.3.

Year	Crime Rate	Percent Unemployment
1994	5,373.5	6.1%
1995	5,277.6	5.6%
1996	5,086.6	5.4%
1997	4,922.7	4.9%
1998	4,619.3	4.5%
1999	4,266.8	4.2%
2000	4,124.8	4.0%
2001	4,160.5	4.8%

Table 2.2: Crime rate and unemployment data.

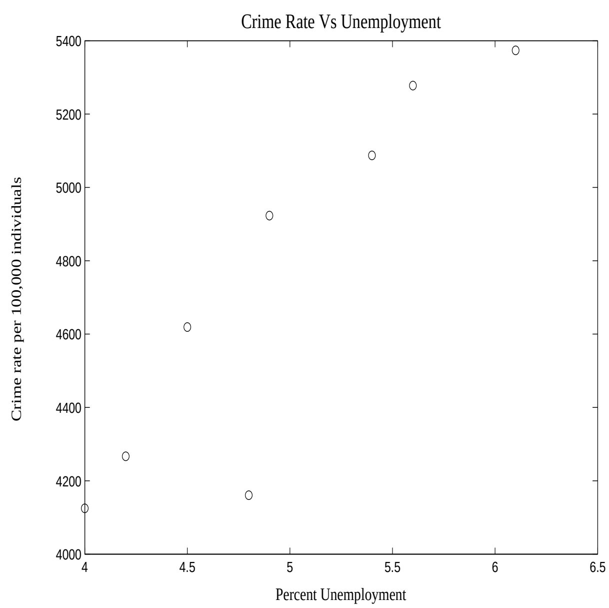

Figure 2.3: Scatterplot of crime rate versus unemployment.

Certainly the first thing we observe about our scatterplot is that there does seem to be a distinct connection: as unemployment increases, crime rate increases. In fact, aside from the point for 2001, the trend appears fairly steady. In general, we would study this point at the year 2001 very carefully and try to determine whether this is an anomaly or a genuine shift in the paradigm. For the purposes of this example, however, we're going to treat it as an *outlyer*—a point that for one reason or another doesn't follow an otherwise genuinely predictive model. The important point to keep in mind is that discarding outlyers when fitting data is a perfectly valid approach, *so long as you continue to check and reappraise your model as future data becomes available.*

Discarding the outlyer, we try both a linear fit and a quadratic fit, each shown on Figure 2.4.

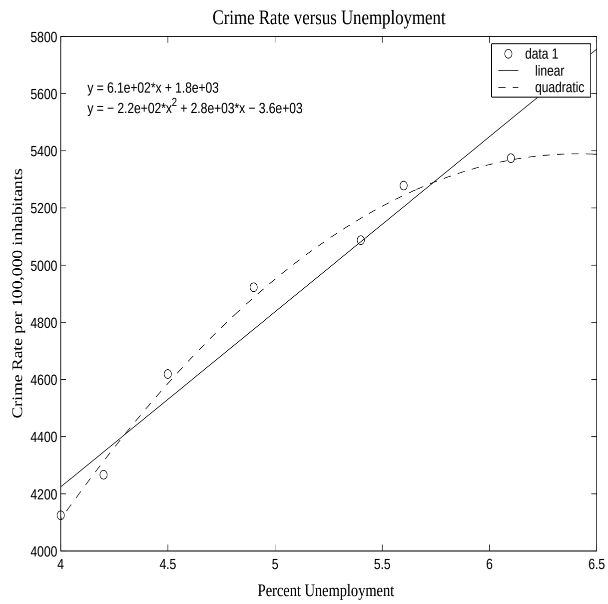

Figure 2.4: Best fit curves for crime–unemployment data.

Clearly, the quadratic fit is better, and though we haven't yet quantitatively developed the socioeconomic reasons for this particular relation,[2] we do have a genuinely predictive model that we can test against future data. For example, the percent unemployment in 2002 was 5.8%, which our quadratic model would predict should be associated with a crime rate of 5,239.2 crimes per 100,000 inhabitants. The actual crime rate for 2002 (not yet finalized) was 4,214.6. At this point, we are led to conclude that our model is not sufficiently accurate. In this case, the problem is most likely a lag effect: while a short-term rise in the unemployment rate doesn't seem to have much effect on crime, perhaps a sustained rise in unemployment would. Though for now we will leave this as a question for someone else to grapple with. △

2.2 Regression with more general functions

Example 2.3. Yearly temperature fluctuations are often modeled by trigonometic expressions, which lead to more difficult regression analyses. Here, we'll consider the case of monthly average maximum temperatures in Big Bend National Park. In Table 2.3 the first column lists raw data of average maximum temperatures each month. In order to model this data with a simple trigonometic model, we'll subtract the mean (which gives Column 3) and divide by the maximum absolute value (which gives Column 4) to arrive at a column of dependent variables that vary like sin and cos between -1 and +1.

[2] And don't worry, we're not going to.

Month	Average Max Temp	Minus Mean	Scaled
Jan.	60.9	-18.0	-1.00
Feb.	66.2	-12.7	-.71
Mar.	77.4	-1.5	-.08
Apr.	80.7	1.8	.10
May	88.0	9.1	.51
June	94.2	15.3	.85
July	92.9	14.0	.78
Aug.	91.1	12.2	.68
Sept.	86.4	7.5	.42
Oct.	78.8	-.1	-.01
Nov.	68.5	-10.4	-.58
Dec.	62.2	-16.7	-.93

Table 2.3: Average maximum temperatures for Big Bend National Park.

A reasonable model for this data is $T(m) = \sin(m-a)$ (T represents scaled temperatures and m represents a scaled index of months ($m \in [0, 2\pi]$)), where by our reductions we've limited our analysis to a single parameter, a. Proceeding as above, we consider the regression error

$$E(a) = \sum_{k=1}^{n} (T_k - \sin(m_k - a))^2.$$

Computing $\partial_a E(a) = 0$, we have

$$2 \sum_{k=1}^{n} (T_k - \sin(m_k - a)) \cos(m_k - a) = 0,$$

a nonlinear equation for the parameter a. Though nonlinear algebraic equations are typically difficult to solve analytically, they can certainly be solve numerically. In this case, we will use MATLAB's $fzero()$ function. First, we write an M-file that contains the function we're setting to zero, listed below as *bigbend.m.*

```
function value = bigbend(a);
%BIGBEND: M-file containing function for
%fitting Big Bend maximum temperatures.
scaledtemps = [-.08 .10 .51 .85 ...
.78 .68 .42 -.01 -.58 -.93 -1.00 -.71];
value = 0;
for k=1:12
l=2*pi*k/12;
value = value + scaledtemps(k)*cos(l-a)-sin(l-1)*cos(l-a);
end
```

Finally, we solve for a and compare our model with the data, arriving at Figure 2.5.[3]

```
>>months=1:12;
>>temps=[60.9 66.2 77.4 80.7 88.0 94.2 92.9 91.1 86.4 78.8 68.5 62.2];
```

[3]In practice, it's better to write short M-files to carry out this kind of calculation, rather than working at the Command Window, but for the purposes of presentation the Command Window prompt (> >) helps distinguished what I've typed in from what MATLAB has spit back.

```
>>fzero(@bigbend,1)
ans =
1.7169
>>modeltemps=18*sin(2*pi*months/12-1.72)+78.9;
>>plot(months,temps,'o',months,modeltemps)
```

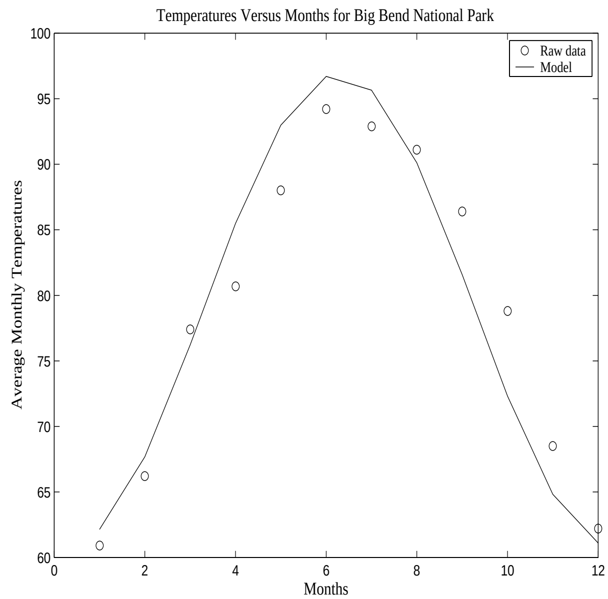

Figure 2.5: Trigonometric model for average monthly temperatures in Big Bend National Park.

For multiple parameters, we must solve a system of nonlinear algebraic equations, which can generally be quite difficult even computationally. For this, we will use the MATLAB function *lsqcurvefit()*.

Example 2.4. Let's consider population growth in the United States, beginning with the first government census in 1790 (see Table 2.4).

Year	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900
Pop	3.93	5.31	7.24	9.64	12.87	17.07	23.19	31.44	39.82	50.16	62.95	75.99

Year	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
Pop	91.97	105.71	122.78	131.67	151.33	179.32	203.21	226.50	249.63	281.42

Table 2.4: Population data for the United States, 1790–2000, measured in millions.

If we let $p(t)$ represent the population at time t, the logistic model of population growth is given by

$$\frac{dp}{dt} = rp(1 - \frac{p}{K}); \quad p(0) = p_0 \tag{2.2}$$

where p_0 represents the initial population of the inhabitants, r is called the "growth rate" of the population, and K is called the "carrying capacity." Observe that while the rate at which the population grows is assumed to be proportional to the size of the population, the population is assumed to have a maximum possible number of inhabitants, K. (If $p(t)$ ever grows larger than K, then $\frac{dp}{dt}$ will become negative and the population will decline.) Equation (2.2) can be solved by separation of variables and partial fractions, and we find

$$p(t) = \frac{p_0 K}{(K - p_0)e^{-rt} + p_0}. \tag{2.3}$$

Though we will take year 0 to be 1790, we will assume the estimate that year was fairly crude and obtain a value of p_0 by fitting the entirety of the data. In this way, we have three parameters to contend with, and carrying out the full regression analysis would be tedious.

The first step in finding values for our parameters with MATLAB consists of writing our function $p(t)$ as an M-file, with our three parameters p_0, r, and K stored as a parameter vector $p = (p(1), p(2), p(3))$:

```
function P = logistic(p,t);
%LOGISTIC: MATLAB function file that takes
%time t, growth rate r (p(1)),
%carrying capacity K (p(2)),
%and initial population P0 (p(3)), and returns
%the population at time t.
P = p(2).*p(3)./((p(2)-p(3)).*exp(-p(1).*t)+p(3));
```

MATLAB's function *lsqcurvefit()* can now be employed at the Command Window, as follows.

```
>>decades=0:10:210;
>>pops=[3.93 5.31 7.24 9.64 12.87 17.07 23.19 31.44 39.82 50.16 62.95 75.99...
91.97 105.71 122.78 131.67 151.33 179.32 203.21 226.5 249.63 281.42];
>>p0=[.01 1000 3.93];
>>options = optimset('MaxFunEvals',5000);
>>[p,error]=lsqcurvefit(@logistic,p0,decades,pops)
Maximum number of function evaluations exceeded;
increase options.MaxFunEvals
p =
0.0206 485.2402 8.4645
error =
464.1480
>>sqrt(error)
ans =
21.5441
>>modelpops=logistic(p,decades);
>>plot(decades,pops,'o',decades,modelpops)
```

After defining the data, we have entered our initial guess as to what the parameters should be, the vector p_0. (Keep in mind that MATLAB is using a routine similar to *fzero()* to solve the system of nonlinear algebraic equations, and typically can only find roots reasonably near your guesses.) In this case, we have guessed a small value of r corresponding very roughly to the fact that we live 70 years and have on average (counting men and women) 1.1 children per person[4] ($r \cong 1/70$), a population carrying capacity of 1 billion, and an initial population equal to the census data. In the next line, we have used an optional command *optimset*, which gives us control over the maximum number of iterations *lsqcurvefit()* will do. Finally, we use *lsqcurvefit()*, entering repectively our function file, our intial parameter guesses, and our data. In this case, MATLAB has not closed its iteration completely, but checking our model, we will find that the parameters fit fairly well as they are. The function *lsqcurvefit()* renders two outputs: our parameters and a sum of squared errors, which we have called error. For the parameters, we observe that r has remained small (roughly 1/50),

[4] According to census 2000.

our carrying capacity is 485 million people, and our initial population is 8.46 million. Though the error looks enormous, keep in mind that this is a sum of all errors squared,

$$\text{error} = \sum_{\text{decades}} (\text{pops(decade)-modelpops(decade)})^2.$$

A more reasonable measure of error is the square root of this, from which we see that over 22 decades our model is only off by around 21.54 million people. In the last two lines of code, we have created Figure 2.6, in which our model is compared directly with out data. △

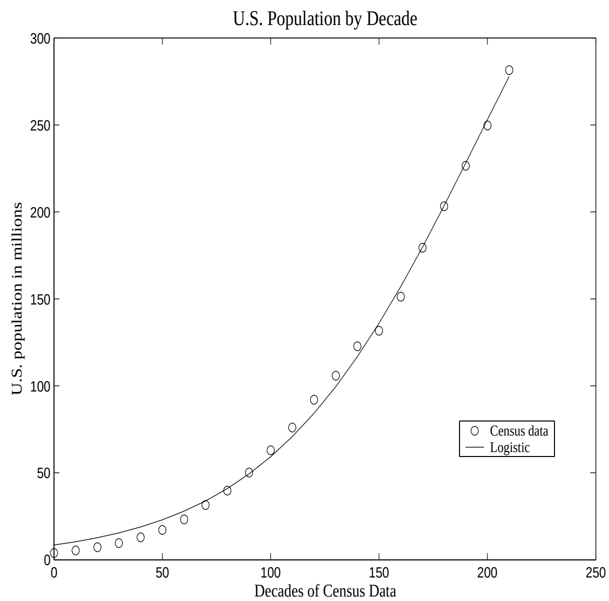

Figure 2.6: U.S. Census data and logistic model approximation.

2.3 Multivariate Regression

Often, the phenomenon we would like to model depends on more than one independent variable. (Keep in mind the following distinction: While the model in Example 2.4 depended on three parameters, it depended on only a single independent variable, t.)

Example 2.5. Film production companies such as Paramount Studios and MGM employ various techniques to predict movie ticket sales. In this example, we will consider the problem of predicting final sales based on the film's first weekend. The first difficulty we encounter is that first weekend sales often depend more on hype than quality. For example, *Silence of the Lambs* and *Dude, Where's My Car?* had surprisingly similar first-weekend sales: \$13,766,814 and \$13,845,914 respectively.[5] (*Dude* did a little better.) Their final sales weren't so close: \$130,726,716 and \$46,729,374, again respectively. Somehow, our model has to *numerically* distinguish between movies like *Silence of the Lambs* and *Dude, Where's My Car?* Probably the easiest way to do this is by considering a second variable, the movie's rating. First weekend sales, final sales and TV Guide ratings are listed for ten movies in Table 2.5.[6]

Letting S represent first weekend sales, F represent final sales, and R represent ratings, our model will take the form

$$F = a_0 + a_1 S + a_2 R,$$

[5]All sales data for this example were obtained from http://www.boxofficeguru.com.

[6]TV Guide's ratings were easy to find, so I've used them for the example, but they're actually pretty lousy. FYI.

Movie	First Weekend Sales	Final Sales	Rating
Dude, Where's My Car?	13,845,914	46,729,374	1.5
Silence of the Lambs	13,766,814	130,726,716	4.5
We Were Soldiers	20,212,543	78,120,196	2.5
Ace Ventura	12,115,105	72,217,396	3.0
Rocky V	14,073,170	40,113,407	3.5
A.I.	29,352,630	78,579,202	3.0
Moulin Rouge	13,718,306	57,386,369	2.5
A Beautiful Mind	16,565,820	170,708,996	3.0
The Wedding Singer	18,865,080	80,245,725	3.0
Zoolander	15,525,043	45,172,250	2.5

Table 2.5: Movie Sales and Ratings.

where a_0, a_1, and a_2 are parameters to be determined. For each set of data points from Table 2.5 (S_k, R_k, F_k) $(k = 1, ..., 10)$ we have an equation

$$F_k = a_0 + a_1 S_k + a_2 R_k.$$

Combining the ten equations (one for each k) into matrix form, we have

$$\begin{pmatrix} 1 & S_1 & R_1 \\ 1 & S_2 & R_2 \\ \vdots & \vdots & \vdots \\ 1 & S_{10} & R_{10} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_{10} \end{pmatrix}, \tag{2.4}$$

or $Ma = F$, where

$$M = \begin{pmatrix} 1 & S_1 & R_1 \\ 1 & S_2 & R_2 \\ \vdots & \vdots & \vdots \\ 1 & S_{10} & R_{10} \end{pmatrix}, \quad a = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}, \quad \text{and} \quad F = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_{10} \end{pmatrix}.$$

In generally, equation (2.4) is over-determined—ten equations and only three unknowns. In the case of over-determined systems, MATLAB computes $a = M^{-1}F$ by linear least squares regression, as above, starting with the error $E = \sum_{k=1}^{10}(F_k - a_0 - a_1 S_k - a_2 R_k)^2$. Hence, to carry out regression in this case on MATLAB, we use the following commands:

```
>>S=[13.8 13.8 20.2 12.1 14.1 29.4 13.7 16.6 18.9 15.5];
>>F=[46.7 130.7 78.1 72.2 40.1 78.6 57.4 170.7 80.2 45.2];
>>R=[1.5 4.5 2.5 3.0 3.5 3.0 2.5 3.0 3.0 2.5];
>>M=[ones(size(S))' S' R'];
>>a=M\F'
a =
-6.6986
0.8005
25.2523
```

Notice that a prime (') has been set after each vector in the definition of M to change the row vectors into column vectors. MATLAB's notation for $M^{-1}F$ is $M\backslash F$.

Though scatterplots can be difficult to read in three dimensions, they are often useful to look at. In this case, we simply type *scatter3(S,R,F)* at the MATLAB Command Window prompt to obtain Figure 2.7.

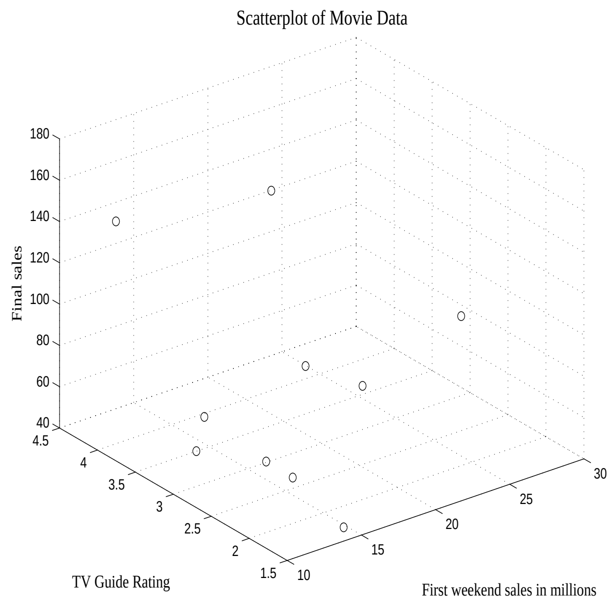

Figure 2.7: Scatterplot for movie sales data.

While in the case of a single variable, regression found the best-fit line, for two variables it finds the best-fit plane. Employing the following MATLAB code, we draw the best fit plane that arose from our analysis (see Figure 2.8).

```
>>hold on
>>x=10:1:30;
>>y=1:.5:5;
>>[X,Y]=meshgrid(x,y);
>>Z=a(1)+a(2)*X+a(3)*Y;
>>surf(X,Y,Z)
```

2.4 Parameter Estimation Directly from Differential Equations

Modeling a new phenomenon, we often find ourselves in the following situation: we can write down a differential equation that models the phenomenon, but we cannot solve the differential equation analytically. We would like to solve it numerically, but all of our techniques thus far for parameter estimation assume we know the exact form of the function.

Example 2.6. In population models, we often want to study the interaction between various populations. Probably the simplest interaction model is the *Lotka-Volterra* predator–prey model,

$$\begin{aligned}\frac{dx}{dt} &= ax - bxy \\ \frac{dy}{dt} &= -ry + cxy,\end{aligned} \tag{2.5}$$

where $x(t)$ represents the population of prey at time t and $y(t)$ represents the population of predators at time t. Observe that the interaction terms, -bxy and +cxy, correspond respectively with death of prey in the presence of predators and proliferation of predators in the presence of prey.

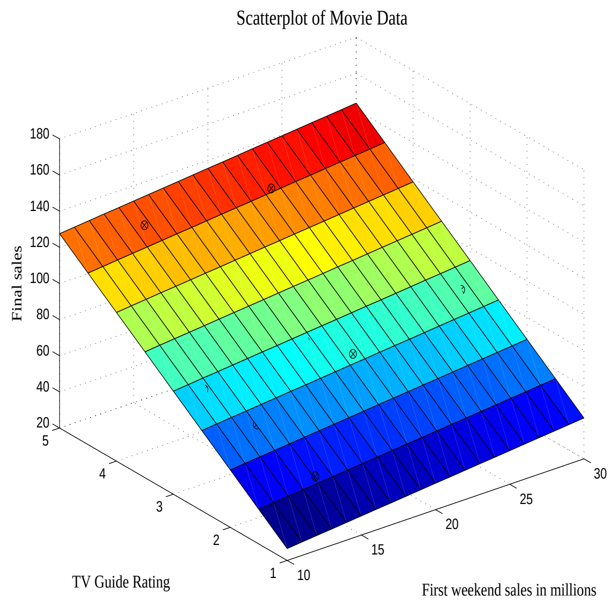

Figure 2.8: Movie sales data along with best-fit plane.

Though certainly instructive to study, the Lotka–Volterra model is typically too simple to capture the complex dynamics of species interaction. One famous example that it does model fairly well is the interaction between lynx (a type of wildcat) and hare (mammals in the same biological family as rabbits), as measured by pelts collected by the Hudson Bay Company between 1900 and 1920. Raw data from the Hudson Bay Company is given in Table 2.6.

Year	Lynx	Hare	Year	Lynx	Hare	Year	Lynx	Hare
1900	4.0	30.0	1907	13.0	21.4	1914	45.7	52.3
1901	6.1	47.2	1908	8.3	22.0	1915	51.1	19.5
1902	9.8	70.2	1909	9.1	25.4	1916	29.7	11.2
1903	35.2	77.4	1910	7.4	27.1	1917	15.8	7.6
1904	59.4	36.3	1911	8.0	40.3	1918	9.7	14.6
1905	41.7	20.6	1912	12.3	57.0	1919	10.1	16.2
1906	19.0	18.1	1913	19.5	76.6	1920	8.6	24.7

Table 2.6: Number of pelts collected by the Hudson Bay Company (in 1000's).

Our goal here will be to estimate values of a, b, r, and c without finding an exact solution to (2.5). Beginning with the predator equation, we first assume the predator population is not zero and re-write it as

$$\frac{1}{y}\frac{dy}{dt} = cx - r.$$

If we now treat the expression $\frac{1}{y}\frac{dy}{dt}$ as a single variable, we see that c and r are respectively the slope and intercept of a line. That is, we would like to plot values of $\frac{1}{y}\frac{dy}{dt}$ versus x and fit a line through this data. Since we have a table of values for x and y, the only difficulty in this is finding values of $\frac{dy}{dt}$. In order to do this, we first recall the definition of derivative,

$$\frac{dy}{dt}(t) = \lim_{h\to 0} \frac{y(t+h) - y(t)}{h}.$$

Following the idea behind Euler's method for numerically solving differential equations, we conclude that for h sufficiently small,

$$\frac{dy}{dt}(t) \cong \frac{y(t+h) - y(t)}{h},$$

which we will call the *forward difference* derivative approximation.

Critical questions become, how good an approximation is this and can we do better? To answer the first, we recall that the Taylor series for any function, $f(x)$, which admits a power series expansion, is given by

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + ...$$

Letting $x = t + h$ and $a = t$, we obtain the expansion,

$$f(t+h) = f(t) + f'(t)h + \frac{f''(t)}{2}h^2 + \frac{f'''(t)}{3!}h^3 + ...$$

Finally, we subtract $f(t)$ from both sides and divide by h to arrive at our approximation,

$$\frac{f(t+h) - f(t)}{h} = f'(t) + \frac{f''(t)}{2}h + \frac{f'''(t)}{3!}h^2 + ...,$$

from which we see that the error in our approximation is proportional to h. We will say, then, that the forward difference derivative approximation is an *order one* approximation, and write

$$f'(t) = \frac{f(t+h) - f(t)}{h} + \mathbf{O}(|h|); \quad t \text{ typically confined to some bounded interval, } t \in [a, b],$$

where $g(h) = \mathbf{O}(|h|)$ simply means that $|\frac{g(h)}{h}|$ remains bounded as $h \to 0$.

Our second question above was, can we do better? In fact, it's not difficult to show (see homework) that the *central difference* derivative approximation is second order:

$$f'(t) = \frac{f(t+h) - f(t-h)}{2h} + \mathbf{O}(h^2).$$

Returning to our data, we observe that h in our case will be 1, not particularly small. But keep in mind that our goal is to estimate the parameters, and we can always check the validity of our estimates by checking the model against our data. Since we cannot compute a central difference derivative approximation for our first year's data (we don't have $y(t-h)$) we begin in 1901, and compute

$$\frac{1}{y(t)}\frac{dy}{dt} \cong \frac{1}{y(t)}\frac{y(t+h) - y(t-h)}{2h} = \frac{1}{6.1}\frac{9.8 - 4.0}{2} = c \cdot 47.2 - r.$$

Repeating for each year up to 1919 we obtain the system of equations that we will solve by regression. In MATLAB, the computation becomes,

```
>>H=[30 47.2 70.2 77.4 36.3 20.6 18.1 21.4 22 25.4 27.1 ...
40.3 57 76.6 52.3 19.5 11.2 7.6 14.6 16.2 24.7];
>>L=[4 6.1 9.8 35.2 59.4 41.7 19 13 8.3 9.1 7.4 ...
8 12.3 19.5 45.7 51.1 29.7 15.8 9.7 10.1 8.6];
>>for k=1:19
dL(k)=(1/L(k+1))*(L(k+2)-L(k))/2;
Hs(k)=H(k+1);
end
>>plot(Hs,dL,'o')
```

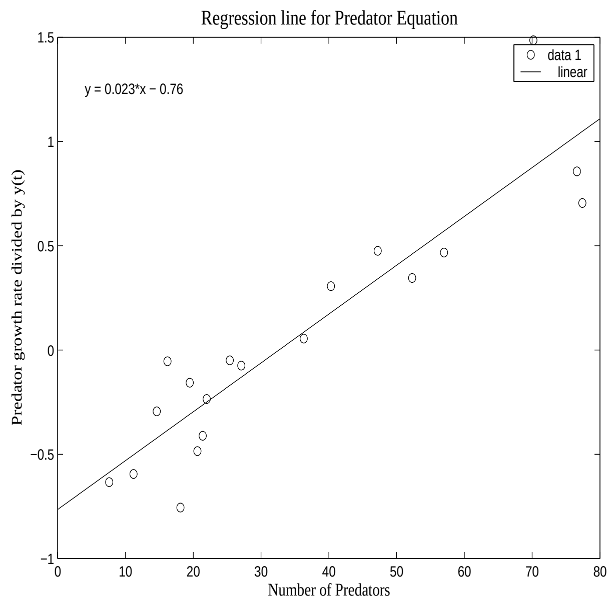

Figure 2.9: Linear fit for parameter estimation in prey equation.

From Figure 1, we can read off our first two parameter values $c = .023$ and $r = .76$. Proceeding similarly for the prey equation, we find $a = .47$ and $b = .024$, and define our model in the M-file *lv.m*.

```
function yprime = lv(t,y)
%LV: Contains Lotka-Volterra equations
a = .47; b = .024; c = .023; r = .76;
yprime = [a*y(1)-b*y(1)*y(2);-r*y(2)+c*y(1)*y(2)];
```

Finally, we check our model with Figure 2.10.

```
>>[t,y]=ode23(@lv,[1 21],[30 4]);
>>years=1:21;
>>subplot(2,1,1);
>>plot(t,y(:,1),years,H,'o')
>>subplot(2,1,2)
>>plot(t,y(:,2),years,L,'o')
```

3 Dimensional Analysis

If we want to compute the average velocity of an object, we typically measure two quantities: the distance the object traveled and the amount of time that passed while the object was in motion. We say, then, that the *dimensions* of velocity are length, L, divided by time, T, and write

$$\text{dimensions of velocity} = [v] = LT^{-1}. \tag{3.1}$$

We will refer to length, time and mass, M, as *fundamental dimensions.* Notice in particular that (3.1) holds true regardless of the units we choose—feet, meters, etc. Typical physical quantities and their associated dimensions are listed in Table 3.1.

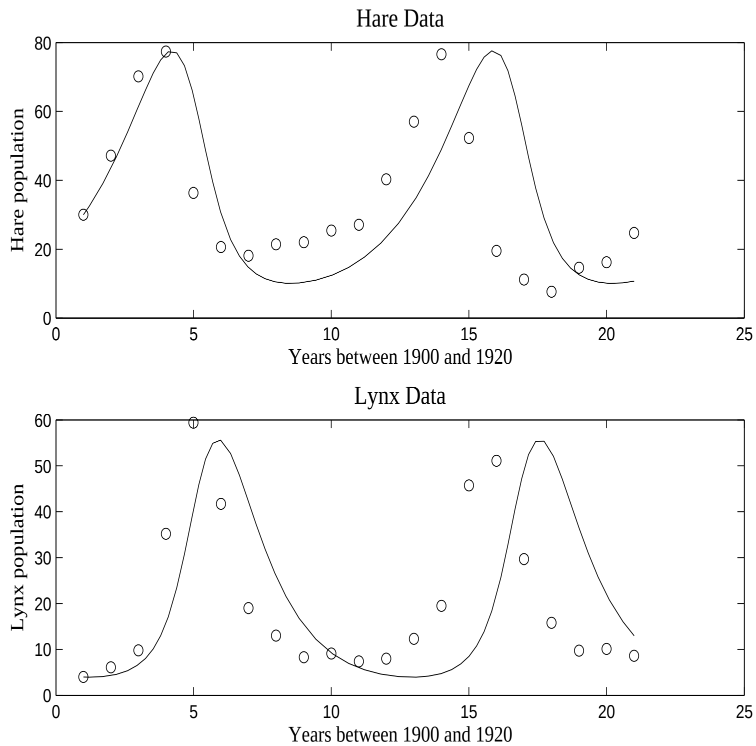

Figure 2.10: Model and data for Lynx–Hare example.

Quantity	Dimensions
Length	L
Time	T
Mass	M
Velocity	LT^{-1}
Acceleration	LT^{-2}
Force	MLT^{-2}
Energy	ML^2T^{-2}
Momentum	MLT^{-1}
Work	ML^2T^{-2}

Quantity	Dimensions
Frequency	T^{-1}
Density	ML^{-3}
Angular momentum	ML^2T^{-1}
Viscosity	$ML^{-1}T^{-1}$
Pressure	$ML^{-1}T^{-2}$
Power	ML^2T^{-3}
Entropy	ML^2T^{-2}
Heat	ML^2T^{-2}
Momentum	MLT^{-1}

Table 3.1: Dimensions of common physical quantities.

3.1 Finding Simple Relations

Dimensional analysis can be an effective tool for determining basic relations between physical quantities.

Example 3.1. Given that the force of gravity between two objects depends on the mass of each object, $m_1$and m_2, the distance between the objects, r, and Newton's gravitational constant G, where

$$[G] = M^{-1}L^3T^{-2},$$

we can determine Newton's law of gravitation. We begin by writing $F = F(m_1, m_2, r, G)$, which is simply a convenient way of expressing that the force due to gravity depends only on these four variables. We now guess that the relation is a simple multiple of powers of these variables and write

$$F(m_1, m_2, r, G) = m_1^a m_2^b r^c G^d.$$

(If this is a bad guess, we will not be able to find values for a, b, c, and d, and we will have to use the slightly more involved analysis outlined in later sections.) Recalling that the dimensions of force are MLT^{-2}, we set the dimensions of each side equal to obtain,

$$MLT^{-2} = M^a M^b L^c M^{-d} L^{3d} T^{-2d}.$$

Equating the exponents of each of our dimensions, we have three equations for our four unknowns:

$$\begin{aligned} M: &\quad 1 = a + b - d \\ L: &\quad 1 = c + 3d \\ T: &\quad -2 = -2d. \end{aligned}$$

We see immediately that $d = 1$ and $c = -2$, though a and b remain undetermined since we have more equations than unknowns. By symmetry, however, we can argue that a and b must be the same, so that $a = b = 1$. We conclude that Newton's law of gravitation must take the form

$$F = G\frac{m_1 m_2}{r^2}.$$

△

Example 3.2. Suppose an object is fired straight upward from the earth with initial velocity v, where v is assumed small enough so that the object will remain close to the earth. Ignoring air resistance, we can use dimensional analysis to determine a general form for the time at which the object lands.

We begin by determining what quantities the final time will depend on, in this case only initial velocity and acceleration due to gravity, g. We write

$$t = t(v, g) \propto v^a g^b \Rightarrow T = L^a T^{-a} L^b T^{-2b},$$

which leads to the dimensions equations,

$$\begin{aligned} T: &\quad 1 = -a - 2b \\ L: &\quad 0 = a + b, \end{aligned}$$

from which we observe that $b = -1$ and $a = 1$. We conclude that $t \propto \frac{v}{g}$, where it's important to note that we have not found an exact form for t, only proportionality. In particular, we have $t = k\frac{v}{g}$ for some unknown dimensionless constant k. *This is as far as dimensional analysis will take us.* (We only obtained an exact form in Example 3.1 because the constant G is well known.) At this point, we should check our expression to insure it makes sense physically. According to our expression, the larger v is, the longer the object will fly, which agrees with our intuition. Also, the stronger g is, the more rapidly the object will descend.

Though in this case the constant of proportionality, k, is straightforward to determine from basic Newtonian mechanics, we typically determine proportionality constants experimentally. In this case, we would launch our object at several different initial velocities and determine k by the methods of Section 2. □

Example 3.3. Consider an object of mass m rotating with velocity v a distance r from a fixed center, in the absence of gravity or air resistance (see Figure 3.1). The *centripetal* force on the object, F_p, is the force required to keep the object from leaving the orbit. We can use dimensional analysis to determine a general form for F_p.

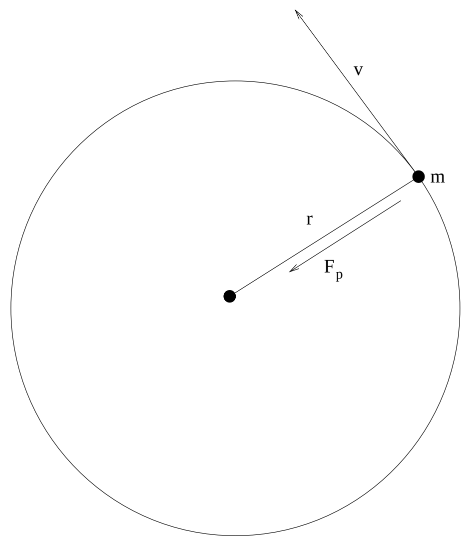

Figure 3.1: Centripetal force on a rotating object.

We begin by supposing that F_p depends only on the quantities m, r, and v, so that,

$$F_p = F_p(m, r, v) \propto m^a r^b v^c \Rightarrow MLT^{-2} = M^a L^b L^c T^{-c},$$

from which we obtain the dimensions equations,

$$\begin{aligned} M: &\quad 1 = a \\ L: &\quad 1 = b + c \\ T: &\quad -2 = -c. \end{aligned}$$

We have, then, $a = 1$, $c = 2$, and $b = -1$, so that

$$F_p = k\frac{mv^2}{r}.$$

In practice the most difficult part of applying dimensional analysis can be choosing the right quantities of dependence. In Examples 3.1 through 3.3 these quantities were given, so let's consider an example in which they are less obvious.

Example 3.4. Determine a general form for the radius created on a moon or planet by the impact of a meteorite.

We begin by simply listing the quantities we suspect might be important: mass of the meteorite, m, density of the earth, ρ_e, volume of the meteorite, V_m, impact velocity of the meteorite, v, and gravitational attraction of the earth, g (which affects how far the earth is displaced). (In a more advanced model, we might also consider density of the atmosphere, heat of the meteorite, etc.) We see immediately that we're going to run into the problem of having three equations (one for each of M, L, and T) and five unknowns, m, ρ_e, V_m, v, and g, so in order to apply the method outlined in the previous examples, we will need to make some reductions. First, let's suppose we don't need to consider both the mass and volume of the meteorite and remove V_m from our list. Next, let's try to combine parameters. Noticing that m and v can be combined into kinetic energy ($\frac{1}{2}mv^2$), we can drop them and consider the new quantity of dependence E. Finally, we are prepared to begin our analysis. We have,

$$r = r(E, \rho_e, g) \propto E^a \rho_e^b g^c \Rightarrow L = M^a L^{2a} T^{-2a} M^b L^{-3b} L^c T^{-2c},$$

from which we obtain the dimensions equations,

$$\begin{aligned} M: &\quad 0 = a + b \\ L: &\quad 1 = 2a - 3b + c \\ T: &\quad 0 = -2a - 2c. \end{aligned}$$

Subsituting $a = -b$ into the second two equations, we find $a = \frac{1}{4}$, $b = -\frac{1}{4}$, and $c = -\frac{1}{4}$, so that

$$r = k(\frac{E}{\rho_e g})^{1/4}.$$

Again, we observe that the basic dependences make sense: higher energies create larger craters, while planets with greater density or gravitational pull receive smaller craters. (Consider, for example, craters on the moon as opposed to craters on the earth.) △

3.2 More General Dimensional Analysis

Example 3.5. Consider the following slight variation on the problem posed in Example 3.2: Suppose an object is fired straight upward from a height h above the earth, and use dimensional analysis to determine a basic form for the time at which the object strikes the earth. The only real difference here is that t now depends on h as well as v and g. Proceeding as before, we have

$$t = t(h, v, g) \propto h^a v^b g^c \Rightarrow T = L^a L^b T^{-b} L^c T^{-2c},$$

from which we obtain the dimensions equations,

$$\begin{aligned} T: &\quad 1 = -b - 2c \\ L: &\quad 0 = a + b + c. \end{aligned}$$

Since mass M does not appear in any of our quantities of dependence (and according to Galileo it shouldn't), we have two equations and three unknowns. We overcame a similar problem in Example 3.4 by dropping a quantity of dependence and by combining variables, but in general, and here in particular, we cannot reasonably do this.

Before introducing our more general method of dimensional analysis, let's see what's happening behind the scenes in Example 3.5. According to Newton's second law of motion, the height of our object at time t is given by

$$y(t) = -gt^2/2 + vt + h.$$

In order to find the time at which our object strikes the earth, we need only solve $y(t) = 0$, which gives

$$t = \frac{-v \pm \sqrt{v^2 + 2gh}}{-g}. \tag{3.2}$$

We have the right quantities of dependence; it's our assumption that t is a simple product of powers that breaks down.

Returning to the problem posed in Example 3.5, let's take a slightly different tack. Instead of beginning with the expression $t = t(h, v, g)$, we will begin now searching for *dimensionless products*,

$$\pi = \pi(h, v, g, t),$$

where t has now become a quantity of dependence and π is dimensionless. (The designation of dimensionless products by π is standard, if perhaps unfortunate.) We have, then,

$$\pi = \pi(h, v, g, t) \propto h^a v^b g^c t^d \Rightarrow 1 = L^a L^b T^{-b} L^c T^{-2c} T^d,$$

from which we obtain the dimensions equations

$$\begin{aligned} T: &\quad 0 = -b - 2c + d \\ L: &\quad 0 = a + b + c. \end{aligned}$$

Since we have two equations and four unknowns, two of our unknowns will remain undetermined and can be chosen. For example, we might choose $d = 1$ and $c = 0$, which determines $b = 1$ and $a = -1$. Our first dimensionless product becomes $\pi_1 = \frac{vt}{h}$. Alternatively, we can choose $d = 0$ and $c = 1$, which determines $b = -2$ and $a = 1$, making our second dimensionless product $\pi_2 = \frac{hg}{v^2}$. Finally, we will take $a = 1$ and $b = 1$, which determines $c = -2$ and $d = -3$, providing a third dimensionless product $\pi_3 = \frac{hv}{g^2 t^3}$. Notice, however, that π_3 is nothing more than π_1^{-3} multiplied by π_2^{-2} ($\pi_3 = \pi_1^{-3}\pi_2^{-2} = \frac{h^3}{v^3 t^3} \cdot \frac{v^4}{h^2 g^2} = \frac{hv}{g^2 t^3}$) and in this sense doesn't give us any new information. In fact, *any* other dimensionless product can be written as some multiplcation of powers of π_1 and π_2, making them a *complete set* of dimensionless products. We will prove this last assertion below, but for now let's accept it and observe what will turn out to be the critical point in the new method: our defining equation for t (nondimensionalized by dividing by h),

$$-gt^2/(2h) + vt/h + 1 = 0,$$

can be rewritten entirely in terms of π_1 and π_2, as

$$-\pi_1^2 \pi_2/2 + \pi_1 + 1 = 0.$$

Solving for π_1, we find

$$\pi_1 = \frac{-1 \pm \sqrt{1 + 2\pi_2}}{-\pi_2},$$

from which we conclude

$$t = \frac{h}{v} \cdot \frac{-1 \pm \sqrt{1 + 2\frac{hg}{v^2}}}{-\frac{hg}{v^2}},$$

which corresponds exactly with (3.2). *Notice that it was critical that our dependent variable t appeared in only one of π_1 and π_2.* Otherwise, we would not have been able to solve for it. △

Our general method of dimensional analysis hinges on the observation that we can always proceed as above. To this end, we have the following theorem.

Theorem 3.1. (Buckingham's Theorem) If the dimensions for an equation are consistent and $\{\pi_1, \pi_2, ..., \pi_n\}$ form a complete set of dimensionless products for the equation, then there exists come function f so that the equation can be written in the form

$$f(\pi_1, \pi_2, ..., \pi_n) = 0.$$

Before proving Theorem 3.1, we will consider two further examples that illustrate its application.

Example 3.6. Determine a general form for the terminal velocity of a raindrop.

As usual, we begin by making a list of the quantities we suspect the terminal velocity of a raindrop should depend on: the size of the raindrop, measured by its radius r (though it won't be precisely spherical), the density ρ and viscosity μ of air (viscosity measures resistance to motion, due in air to collisions between particles. Imagine yourself as a particle walking across campus: density is a measure of how many people are walking around and viscosity is a measure of how many collisions they're having), and of course gravity g. Our general dimensionless product takes the form

$$\pi = \pi(r, \rho, \mu, g, v) = r^a \rho^b \mu^c g^d v^e \Rightarrow 1 = L^a M^b L^{-3b} M^c L^{-c} T^{-c} L^d T^{-2d} L^e T^{-e},$$

from which we obtain the dimensions equations,

$$\begin{aligned} L: &\quad 0 = a - 3b - c + d + e \\ M: &\quad 0 = b + c \\ T: &\quad 0 = -c - 2d - e. \end{aligned}$$

For three equations and five unknowns, we should require exactly 2 dimensionless products—one that depends on the dependent variable v and one that does not. Taking $e = 1$ and $d = 0$, we find $c = -1$, $b = 1$, and $a = 1$, so that our first dimensionless product is $\pi_1 = \frac{r\rho v}{\mu}$. On the other hand, taking $e = 0$ and $d = 1$, we find $c = -2$, $b = 2$, and $a = 3$, so that our second dimensionless product is $\pi_2 = \frac{r^3\rho^2 g}{\mu^2}$. What Buckingham's theorem now tells us is that there exists some function $f(\pi_1, \pi_2)$ so that the equation we're trying to find can be written in the form

$$f(\pi_1, \pi_2) = 0. \tag{3.3}$$

According to the Implicit Function Theorem (see appendix), there exists a function $g(\pi_2)$ so that equation (3.3) can be solved by $\pi_1 = g(\pi_2)$. We have, then, finally

$$v = \frac{\mu}{r\rho} \cdot g(\frac{r^3\rho^2 g}{\mu^2}),$$

which is as far, in this case, as dimensional analysis will take us.

At first glance, it may appear that since g is entirely unknown we haven't made much progress, but observe that while we initially had a function of four variables to find ($v = v(r, \rho, \mu, g)$), we now have a function of only one variable to determine—g, with the single variable $\frac{r^3\rho^2 g}{\mu^2}$. Recalling Section 2 on parameter estimation, we see that this is a critical simplification. △

Before considering our final example, we review the steps of our general method for dimensional analysis.

1. Identify the variables of dependence.

2. Determine a complete set of dimensionless products, $\{\pi_1, \pi_2, ..., \pi_n\}$, making sure that the dependent variable appears in only one, say π_1.

3. Apply Buckingham's Theorem to obtain the existence of a (typically unknown) function f satisfying

$$f(\pi_1, \pi_2, ..., \pi_n) = 0.$$

4. Apply the Implicit Function Theorem to obtain the existence of a (typically unknown) function g satisfying

$$\pi_1 = g(\pi_2, \pi_3, ..., \pi_n).$$

5. Solve the equation from Step 4 for the dependent variable and use experimental data to determine the form for g.

Example 3.7. How long should a turkey be roasted?

The variables that cooking time should depend on are (arguably): size of the turkey, measured by length r, the initial temperature of the turkey, T_t, the temperature of the oven (assumed constant) T_o, and the *coefficient of heat conduction* for the turkey, k ($[k] = L^2T^{-1}$). Typically, temperature is measured in *Kelvins*, an SI (metric) unit which uses as its base the temperature at which all three phases of water can exist in equilibrium. For our purposes, a good measure of temperature is heat (or energy) per volume, hence T_t and T_o will both have units $ML^{-1}T^{-2}$ (see Table 3.1). Our dimensionless products take the form

$$\pi = \pi(r, T_t, T_o, k, t) = r^a T_t^b T_o^c k^d t^e \Rightarrow 1 = L^a M^b L^{-b} T^{-2b} M^c L^{-c} T^{-2c} L^{2d} T^{-d} T^e,$$

from which we obtain the dimensions equations

$$\begin{aligned} L:& \quad 0 = a - b - c + 2d \\ M:& \quad 0 = b + c \\ T:& \quad 0 = -2b - 2c - d + e. \end{aligned}$$

As in Example 3.6, we have three equations and five unknowns and require two dimensionless products. Taking $e = 1$ and $b = 0$, we find $c = 0$, $d = 1$, and $a = -2$, so that our first dimensionless product is $\pi_1 = \frac{kt}{r^2}$. On the other hand, taking $e = 0$ and $b = 1$, we find $c = -1$, $d = 0$, and $a = 0$, so that our second dimensionless product is $\pi_2 = \frac{T_t}{T_o}$. According to Buckingham's theorem, there exists a function f so that

$$f(\pi_1, \pi_2) = 0,$$

and by the Implicit Function Theorem another function g so that

$$\pi_1 = g(\pi_2).$$

We conclude that

$$t = \frac{r^2}{k} g(\frac{T_t}{T_o}).$$

△

3.3 A Proof of Buckingham's Theorem

The proof of Buckingham's Theorem depends on an application of linear algebra, so let's first consider the general technique for solving the types of systems of equations that arose in Examples 3.1 through 3.7. Recalling Example 3.5, we have the system

$$\begin{aligned} T:& \quad 0 = -b - 2c + d \\ L:& \quad 0 = a + b + c, \end{aligned}$$

which we now re-write in matrix form

$$\begin{pmatrix} 0 & -1 & -2 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{3.4}$$

We proceed now by standard Gauss–Jordon elimination. The *augmented matrix* for this system is

$$\left(\begin{array}{cccc|c} 0 & -1 & -2 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{array}\right),$$

and only three row operations are required: we swap rows, then add the new Row 2 to the new Row 1, and finally multiply Row 2 by -1, giving

$$\left(\begin{array}{cccc|c} 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 2 & -1 & 0 \end{array}\right),$$

which is typically referred to as *reduced row echelon form* (RREF). We conclude that with c and d chosen arbitrarily, we require only

$$\begin{aligned} a =& c - d \\ b =& -2c + d. \end{aligned}$$

Choosing $d = 1$ and $c = 0$ determines $a = -1$ and $b = 1$, while choosing $d = 0$ and $c = 1$ determines $a = 1$ and $b = -2$, so that we have two solutions to equation (3.4),

$$V_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \quad \text{and} \quad V_2 = \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix}.$$

Accordingly, *any* solution to equation (3.4) can be written as a *linear combination* of V_1 and V_2, $V = c_1V_1 + c_2V_2$, where c_1 and c_2 are scalars.

Recalling, now, Example 3.5, we observe that V_1 corresponds with the dimensionless product π_1, while V_2 corresponds with the dimensionless product π_2. Given any new dimensionless product π_3 (for example, $\pi_3 = \frac{hv}{g^2t^3}$), there corresponds a V_3 (e.g., $V_3 = (1, 1, -2, -3)^{\text{tr}}$) so that $V_3 = c_1V_1 + c_2V_2$ (in the example, $c_1 = -3$ and $c_2 = -2$). Consequently, π_3 can be written as $\pi_3 = \pi_1^{c_1}\pi_2^{c_2}$ $(= \pi_1^{-3}\pi_2^{-2})$, which establishes that π_1 and π_2 do indeed form a complete set of dimensionless products for the governing equation of Example 3.5. This means that every expression in the governing equation of Example 3.5 can be written as a product of powers of π_1 and π_2 and consequently that a function f must exist so that $f(\pi_1, \pi_2) = 0$.

General proof of Buckingham's Theorem. Let $x_1, x_2, ..., x_k$ denote the physical quantities under consideration and define a (one-to-one) function on their products by $h : \mathbb{R} \to \mathbb{R}^k$,

$$h(x_1^{a_1}x_2^{a_2} \cdots x_k^{a_k}) = (a_1, a_2, ..., a_k).$$

Similarly, define a second function $\phi : \mathbb{R} \to \mathbb{R}^n$ (n is the number of fundamental dimensions, usually for us $n = 3$: M, L, T)

$$\phi(x_1^{a_1}x_2^{a_2} \cdots x_k^{a_k}) = (d_1, d_2, ..., d_n),$$

where d_1, d_2, ..., d_n are the powers on the fundamental dimensions (e.g., for $n = 3$, $M^{d_1}L^{d_2}T^{d_3}$). Consider, now, the map $\phi h^{-1} : \mathbb{R}^k \to \mathbb{R}^n$. Let $\{b_1, b_2, ..., b_j\}$ be a basis for the null space of ϕh^{-1} ($\phi h^{-1}(b_m) = 0_n$ for all $m = 1, ..., j$), and extend it (if necessary) to a basis for $\mathbb{R}^k$, $\{b_1, b_2, ..., b_k\}$. Observe that the first j elements of this basis correspond with dimensionless products (they have been chosen so that the powers on the dimenions are all 0). Define $\pi_m = h^{-1}(b_m), m = 1, ..., k$, and observe that since $\{b_1, b_2, ..., b_k\}$ forms a basis for $\mathbb{R}^k$, any vector $(a_1, a_2, ..., a_k)$ can be written as a linear combination of the b_m. Notice also that $h(\pi_1^{i_1}\pi_2^{i_2} \cdots \pi_k^{i_k}) = i_1h(\pi_1) + i_2h(\pi_2) + ... + i_kh(\pi_k) = i_1b_1 + i_2b_2 + ... + i_kb_k$. In particular, the vector $(a_1, a_2, ..., a_k) = (1, 0, ..., 0) = \sum_{l=1}^k c_lb_l$, so that $x_1 = h^{-1}(\sum_{l=1}^k c_lb_l) = h^{-1}((\sum_{l=1}^k c_lh(\pi_l)) = \pi_1^{c_1}\pi_2^{c_2} \cdots \pi_k^{c_k}$, and similarly for $x_2, ..., x_k$ so that the x_m can all be written as products of powers of the π_m. Hence, any equation that can be written in terms of the x_m can be written in terms of the π_m. Finally, we must resolve the issue that for $m > j$, π_m is *not* dimensionless. For these π_m there exist changes of units (meters to feet etc.) which change π_m but do not change the $\pi_1, ..., \pi_j$. But we have assumed that our physical law is independent of units, and so it cannot depend on the π_m. □

3.4 Nondimensionalizing Equations

A final useful trick associated with dimensional analysis is the *nondimensionalization of equations*: writing equations in a form in which each summand is dimensionless. Through nondimensionalization, we can often reduce the number of parameters in an equation and consequently work with the most "bare bones" version of the model.

Example 3.8. As part of the Ballistics Project, we consider a projectile traveling under the influences of gravity and linear air resistance. According to Newton's second law of motion, a projectile traveling vertically under these forces has height $y(t)$ given by

$$y''(t) = -g - by'(t), \tag{3.5}$$

where the coefficient of air resistance, b, has units T^{-1}. Letting L and T represent dimensional constants to be chosen, we define the nondimensional variables $\tau = \frac{t}{T}$ and $Y(\tau) = \frac{y(t)}{L}$. We calculate

$$\frac{d}{d\tau}Y(\tau) = \frac{d}{d\tau}\frac{y(t)}{L} = \frac{1}{L}\frac{d}{d\tau}y(\tau T) = \frac{T}{L}y'(\tau T),$$

so that

$$\frac{d^2}{d\tau^2}Y(\tau) = \frac{T^2}{L}y''(\tau T) = \frac{T^2}{L}(-g - by'(\tau T)) = -g\frac{T^2}{L} - bTY'(\tau),$$

and we have the dimensionless equation

$$Y''(\tau) = -g\frac{T^2}{L} - bTY'(\tau).$$

Choosing finally

$$T = \frac{1}{b} \quad \text{and} \quad L = \frac{g}{b^2},$$

we obtain the reduced dimensionless equation

$$Y''(\tau) = -1 - Y'(\tau).$$

△

Example 3.9. Establish a dimensionless form for the Lotka–Volterra predator–prey model,

$$\begin{aligned}\frac{dx}{dt} &= ax - bxy \\ \frac{dy}{dt} &= -ry + cxy.\end{aligned}$$

First, for population dynamics, we require a new dimension, typically referred to as *biomass,* B ($[x] = B$, $[y] = B$). Assuming the Lotka–Volterra model is dimensionally consistent, we can now determine the dimensions of each parameter:

$$[\frac{dx}{dt}] = [a][x] - [b][x][y] \Rightarrow BT^{-1} = [a]B - [b]B^2 \Rightarrow [a] = T^{-1} \text{ and } [b] = T^{-1}B^{-1}.$$

Similarly, $[r] = T^{-1}$ and $[c] = T^{-1}B^{-1}$. Now, we search for dimensionless variables,

$$\tau = \frac{t}{T}, \quad X(\tau) = \frac{x(t)}{B_1}, \quad Y(\tau) = \frac{y(t)}{B_2},$$

with

$$\begin{aligned}\frac{d}{d\tau}X(\tau) =& \frac{T}{B_1}x'(\tau T) = \frac{T}{B_1}(ax(\tau T) - by(\tau T)x(\tau T)) \\ =& \frac{T}{B_1}(aB_1X(\tau) - bB_1B_2X(\tau)Y(\tau)) = TaX(\tau) - bB_2TX(\tau)Y(\tau),\end{aligned}$$

and similarly,

$$Y'(\tau) = -rTY(\tau) - bB_2TX(\tau)Y(\tau).$$

In this case, we have four parameters and only three scalings, so we will not be able to eliminate all the paramters we did in Example 3.9. We can, however, eliminate three. To this end, we choose $T = a^{-1}$ and $B_2 = ab^{-1}$ to eliminate both parameters from the $X(\tau)$ equation, and we choose $B_1 = ac^{-1}$ to arrive at the dimensionless system,

$$\begin{aligned} X' =& X - XY \\ Y' =& -\frac{r}{a}X + XY, \end{aligned}$$

where $k := -\frac{r}{a}$ becomes our single dimensionless parameter. △

4 Well-posedness Theory

Despite the clunky name, *well-posedness* analysis is one of the most important things for an applied mathematician to understand. In order to get an idea of the issues involved, we will consider the example of a pendulum, initially perturbed but otherwise under the influence of gravity alone.

Example 4.1. Consider the motion of a mass, m, swinging at the end of a rigid rod, as depicted in Figure 4.1.

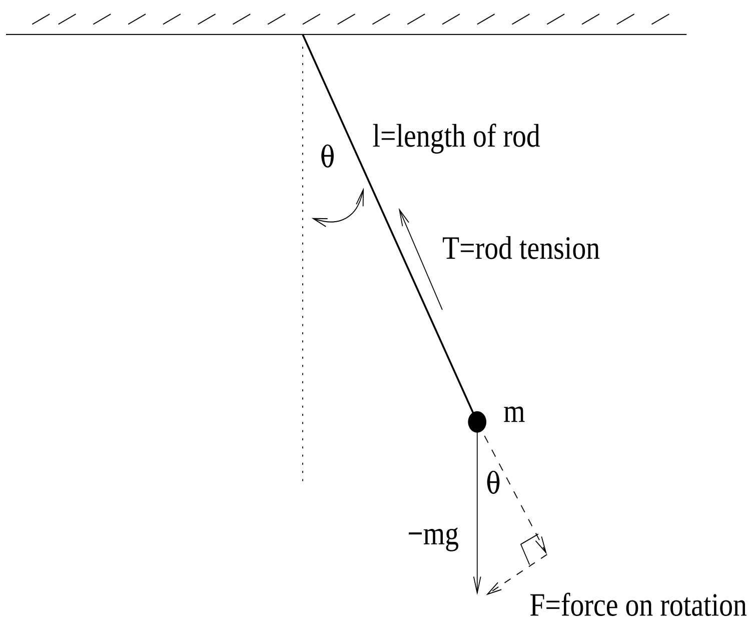

Figure 4.1: Pendulum motion under the influence of gravity alone.

The force due to gravity on m acts vertically downward, and must be decomposed into a force $-T$, which is exactly balanced by the rod, and a force F, directed tangentially to the arc of motion. Observing the right triangle, with hypotenuse of length $-mg$, we have

$$\begin{aligned} \cos\theta &= -\frac{T}{mg} \Rightarrow T = -mg\cos\theta, \\ \sin\theta &= -\frac{F}{mg} \Rightarrow F = -mg\sin\theta. \end{aligned}$$

Measuring distance as arclength, $d = l\theta$, Newton's second law of motion ($F = ma$) determines

$$\begin{aligned}\frac{d^2\theta}{dt^2} &= -\frac{g}{l}\sin\theta \\ \theta(0) = \theta_0, \quad & \frac{d}{dt}\theta(0) = \omega_0.\end{aligned} \tag{4.1}$$

In order to solve equation (4.1) with MATLAB, we must first write it as a first order system. Taking $x_1 = \theta$ and $x_2 = \frac{d\theta}{dt}$, we have

$$\begin{aligned}\frac{dx_1}{dt} =& x_2; \quad x_1(0) = \theta_0, \\ \frac{dx_2}{dt} =& -\frac{g}{l}\sin x_1; \quad x_2(0) = \omega_0.\end{aligned} \tag{4.2}$$

Taking $l = 1$, we will store this equation in the MATLAB M-file *pendode.m,*

```
function xprime = pendode(t,x);
%PENDODE: Holds ODE for pendulum equation.
g = 9.81; l = 1;
xprime = [x(2);-(g/l)*sin(x(1))];
```

and solve it with the M-file *pend.m,*

```
function f = pend(theta0,v0);
%P1: Solves and plots ODE for pendulum equation
%Inputs are initial angle and initial angular velocity
x0 = [theta0 v0];
tspan = [0 5];
[t,x] = ode45('pendode',tspan,x0);
plot(x(:,1),x(:,2));
```

Taking initial angle $\pi/4$ and initial velocity 0 with the command *pend(pi/4,0)*, leads to Figure 4.2 (I've added the labels from MATLAB's pop-up graphics window).

Notice that time has been suppressed and the two dependent variables x_1 and x_2 have been plotted in what we refer to as a *phase portrait*. Beginning at the initial point $\theta_0 = \frac{\pi}{4}$, $\omega_0 = 0$ (the right-hand tip of the football), we observe that angular velocity becomes negative (the pendulum swings to the left) and angle decreases. At the bottom of the arc, the angle is 0 but the angular velocity is at a maximum magnitude (though negatively directed), while at the left-hand tip of the football the object has stopped swinging (instantaneously), and is turning around. The remainder of the curve corresponds with the objects swinging back to its starting position. In the (assumed) absence of air resistance or other forces, the object continues to swing like this indefinitely.

Alternatively, taking initial angle 0 and initial velocity 10 with the command *pend(0,10)* leads to Figure 4.3.

Observe that in this case angular velocity is always positive, indicating that the pendulum is always swinging in the same (angular) direction: we have started it with such a large initial velocity that it's looping its axis.

Now that we have a fairly good idea of how to understand the pendulum phase diagrams, we turn to the critical case in which the pendulum starts pointed vertically upward from its axis (remember that we have assumed it is attached to a rigid rod). After changing the variable *tspan* in *pend* to $[0, 20]$ (solving now for 20 seconds), the command *pend(pi,0)* leads to Figure 4.4. In the absence of any force other than gravity, we expect our model to predict that the pendulum remains standing vertically upward. (What could possibly cause it to fall one way rather than the other?) What we find, however, is that our model predicts that it will fall to the left and then begin swinging around its axis.

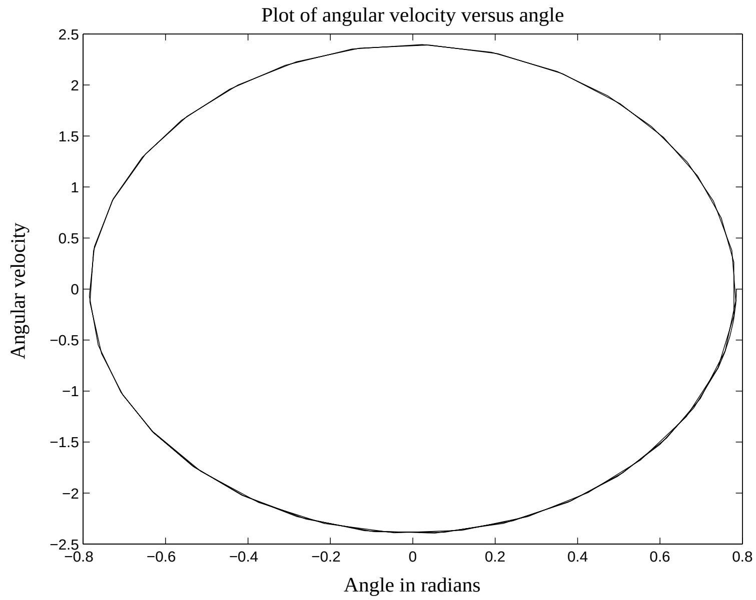

Figure 4.2: Pendulum motion for the case $\theta_0 = \frac{\pi}{4}$ and $\omega_0 = 0$.

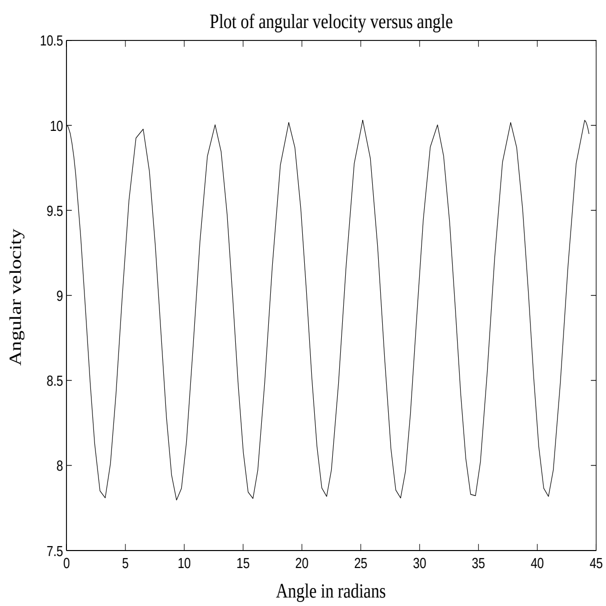

Figure 4.3: Pendulum motion for the case $\theta_0 = 0$ and $\omega_0 = 10\ \mathrm{s}^{-1}$.

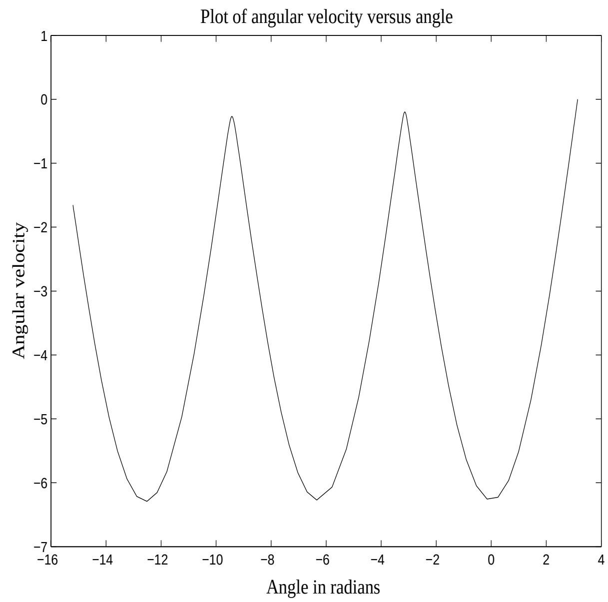

Figure 4.4: Pendulum motion for the case $\theta_0 = \pi$ and $\omega_0 = 0\ \mathrm{s}^{-1}$.

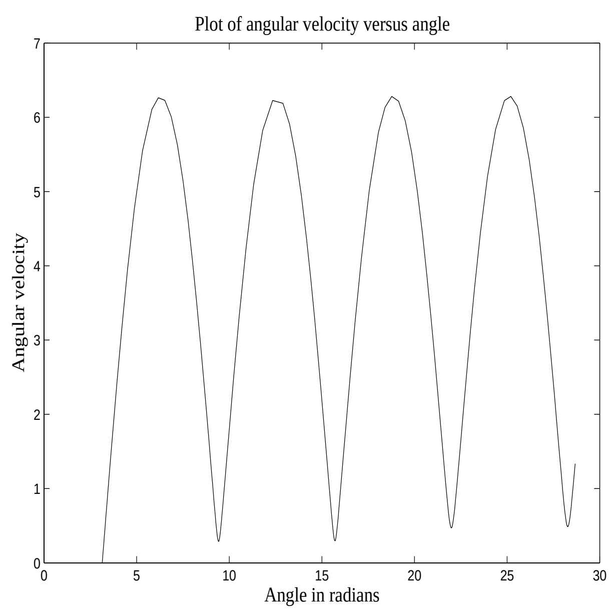

Figure 4.5: Pendulum motion for the case $\theta_0 = \pi + 10^{-12}$ and $\omega_0 = 0\ \mathrm{s}^{-1}$.

Consider finally a change in this last initial data of one over one trillion ($10^{-12} = .000000000001$). The MATLAB command *pend(pi+.000000000001,0)* produces Figure 4.5. We see that with a change in initial data as small as 10^{-12} radians, the change in behavior is enormous: the pendulum spins in the opposite direction. We conclude that our model, at least as it is solved on MATLAB, fails at the initial data point $(\pi, 0)$. In particular, we say that our model is not well-posed at this point. △

In general, for well-posedness, we will require three things of a model:

1. (Existence) There exists a solution to the model.
2. (Uniqueness) The solution is unique.
3. (Stability) The solution does not change dramatically if we only change the initial data a little.

In the next three sections, we will consider each of these in turn, beginning with stability and working our way back to the most abstract theory, existence.

4.1 Stability Theory

The difficulty we ran into in Example 4.1 is with stability. Near the initial data point $(\pi, 0)$, small changes in initial data lead to dramatic changes in pendulum behavior.

Example 4.1 continued. For systems of two first-order differential equations such as (4.2), we can study phase diagrams through the useful trick of dividing one equation by the other. We write,

$$\frac{dx_2}{dx_1} = \frac{\frac{dx_2}{dt}}{\frac{dx_1}{dt}} = \frac{-\frac{g}{l}\sin x_1}{x_2},$$

(the *phase-plane equation*) which can readily be solved by the method of separation of variables for solution

$$\frac{x_2^2}{2} = \frac{g}{l}\cos x_1 + C. \tag{4.3}$$

At $t = 0$, $x_1(0) = \theta_0$ and $x_2(0) = \omega_0$, fixing C. We will create a phase plane diagram with the M-file *penphase.m*.

```
function f = penphase(theta,w0);
%PENPHASE: Plots phase diagram for
%pendulum equation with initial angle theta
%and initial angular velocity w0.
g = 9.81; l = 1.0;
C = w0^2/2 - (g/l)*cos(theta);
if C > g/l
x = linspace(-pi,pi,50);
else
maxtheta = acos(-C*l/g); %Maximum value of theta
x = linspace(-maxtheta,maxtheta,50);
end
up = sqrt(2*g/l*cos(x)+2*C);
down = -sqrt(2*g/l*cos(x)+2*C);
plot(x,up);
hold on
plot(x,down);
```

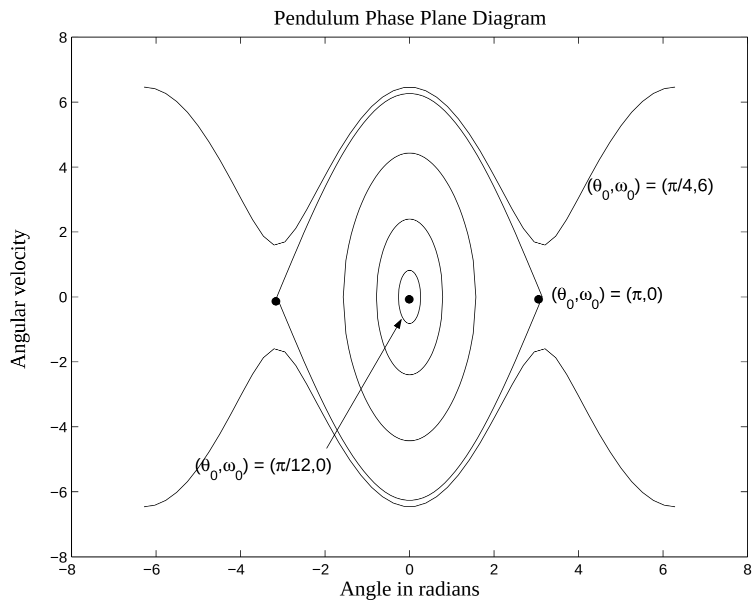

Figure 4.6: Phase plane diagram for a simple pendulum (Example 4.1 continued).

Typing in sequence *penphase(pi/4,6), penphase(pi/12,0), penphase(pi/4,0), penphase(pi/2,0), penphase(pi,0), penphase(pi/4,6)*, we create the phase plane diagram given in Figure 4.6.

The point $(\theta_0, \omega_0) = (0,0)$ corresponds with the pendulum's hanging straight down, while the points $(\theta_0, \omega_0) = (\pi, 0)$ and $(\theta_0, \omega_0) = (-\pi, 0)$ both correspond with the pendulum's standing straight up above its axis. Notice that at each of these *critical* or *equilibrium* points our model analytically predicts that the pendulum will not move. For example, at $(\theta_0, \omega_0) = (0,0)$ we find from (4.2) that $\frac{dx_1}{dt} = \frac{dx_2}{dt} = 0$: the angle and angular velocity are both zero, so the pendulum remains at rest.

Definition. (Equilibrium point) For a system of ordinary differential equations

$$\frac{dx}{dt} = f(t,x), \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix},$$

we refer to any point (t_0, x_0) so that $f(t_0, x_0) = 0$ as an *equilibrium point.*

Typically, equilibrium points govern long time behavior of physical models, so we are interested in stability near them. Perturbing the initial point $(0,0)$ a little (pushing the pendulum slightly to the right or left), we observe that the pendulum's behavior changes only slightly: if we push it one millimeter to the right, it will swing back and forth with maximum displacement one millimeter. On the other hand, as we have seen, if we perturb the initial point $(\pi, 0)$ the pendulum's behavior changes dramatically. We say that $(0,0)$ is a stable equilibrium point and that $(\pi, 0)$ and $(-\pi, 0)$ are both unstable equilibrium points.

In general, we can study stability without solving equations quite as complicated as (4.3). Suppose we want to analyze stability at the point $(0,0)$. We first recall the Taylor expansion of $\sin x$ near $x = 0$,

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + ...$$

For x near 0, higher powers of x are dominated by x, and we can take the approximation, $\sin x \cong x$, which leads to the *linearized* equations,

$$
\begin{aligned}
\frac{dx_1}{dt} &= x_2 \\
\frac{dx_2}{dt} &= -\frac{g}{l}x_1.
\end{aligned}
\tag{4.4}
$$

(That is, the right-hand sides of (4.4) are both linear, which will always be the case when we take the linear terms from a Taylor expansion about an equilibrium point.) Developing the phase plane equation as before, we now have

$$\frac{dx_2}{dx_1} = \frac{\frac{dx_2}{dt}}{\frac{dx_1}{dt}} = \frac{-\frac{g}{l}x_1}{x_2},$$

with solution

$$\frac{x_2^2}{2} + \frac{g}{l}\cdot\frac{x_1^2}{2} = C,$$

which corresponds with ellipses centered at $(0,0)$ with radial axis lengths $\sqrt{2C}$ and $\sqrt{2lC/g}$ (see Figure 4.7). Typically such solutions are referred to as *integral curves.* Returning to equations (4.4), we add direction along the ellipses by observing from the first equation that for $x_2 > 0$, $x_1$is increasing, and for $x_2 < 0$, x_1 is decreasing. The directed sections of integral curves along which the object moves are called *trajectories.* Our stability conclusion is exactly the same as we drew from the more complicated Figure 4.6. In particular, in the case that we have closed loops about an equilibrium point, we say the point is *orbitally stable.*

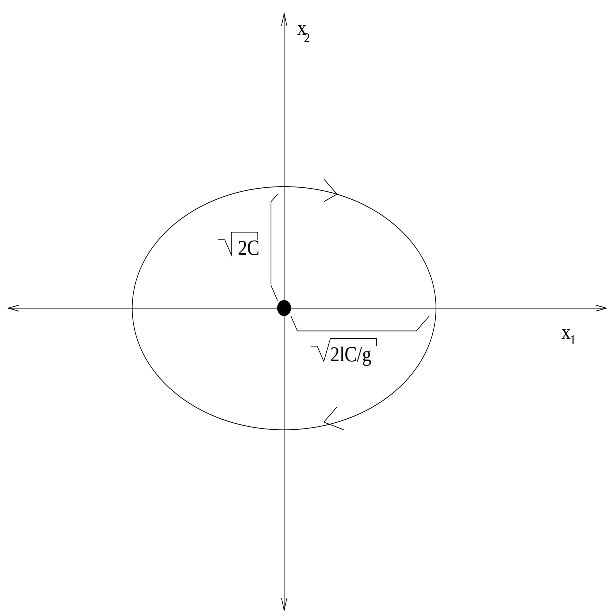

Figure 4.7: Phase plane diagram near the equilibrium point $(0, 0)$.

For the point $(\pi, 0)$ we first make the change of variables,

$$
\begin{aligned}
x_1 &= \pi + y_1 \\
x_2 &= 0 + y_2,
\end{aligned}
$$

and observe that in the variables y_1 and y_2 the equilibrium point is again at $(0, 0)$. In these variables, our system becomes,

$$\begin{aligned}\frac{dy_1}{dt} =& y_2 \\ \frac{dy_2}{dt} =& -\frac{g}{l}\sin(\pi + y_1).\end{aligned}$$

Recalling the Taylor expansion of $\sin y_1$ at the point π,

$$\sin(\pi + y_1) = \sin\pi + (\cos\pi)y_1 - \frac{\sin\pi}{2}y_1^2 + ...,$$

we arrive at the new linearized equation,

$$\begin{aligned}\frac{dy_1}{dt} =& y_2 \\ \frac{dy_2}{dt} =& \frac{g}{l}y_1.\end{aligned}$$

Proceeding exactly as above we again write the phase plane equation,

$$\frac{dy_2}{dy_1} = \frac{\frac{dy_2}{dt}}{\frac{dy_1}{dt}} = \frac{\frac{g}{l}y_1}{y_2},$$

which can be solved by the method of separation of varibles for implicit solution,

$$-\frac{y_2^2}{2} + \frac{g}{l}\frac{y_1^2}{2} = C,$$

which corresponds with hyperbolas (see Figure 4.8). Observe that in this case all trajectories move first toward the equilibrium point and then away. We refer to such an equilibrium point as an *unstable saddle.* △

Example 4.2. As a second example of stability analysis, we will consider the Lotka–Volterra predator–prey equations (2.5),

$$\begin{aligned}\frac{dx}{dt} =& ax - bxy \\ \frac{dy}{dt} =& -ry + cxy.\end{aligned}$$

First, we find all equilibrium points by solving the system of algebraic equations,

$$\begin{aligned}ax - bxy = 0 \\ -ry + cxy = 0.\end{aligned}$$

We find two solutions, $(x_1, y_1) = (0, 0)$ and $(x_2, y_2) = (\frac{r}{c}, \frac{a}{b})$. The first of these corresponds with an absence of both predator and prey, and of course nothing happens (in the short term). The second is more interesting, a point at which the predator population and the prey population live together without either one changing. If this second point is unstable then any small fluctuation in either species will destroy the equilibrium and one of the populations will change dramatically. If this second point is stable then small fluctuations in species population will not destroy the equilibrium, and we would expect to observe such equilibria in nature. In this way, *stability typically determines physically viable behavior.*

In order to study the stability of this second point, we first linearize our equations by making the substitutions

$$\begin{aligned}x =& \frac{r}{c} + z_1 \\ y =& \frac{a}{b} + z_2.\end{aligned}$$

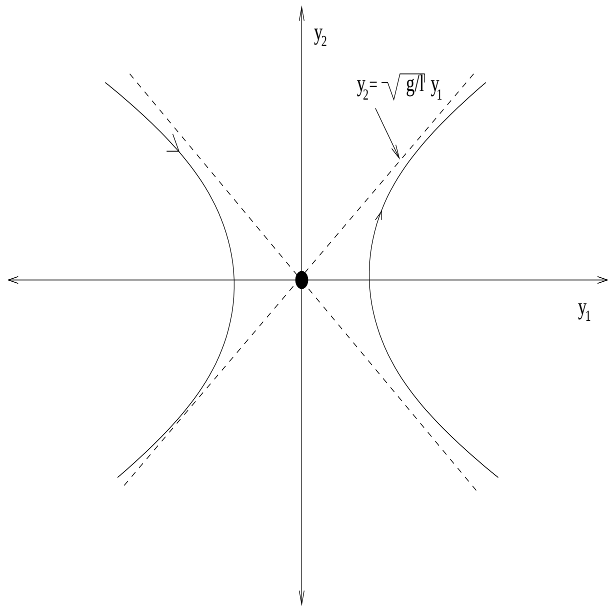

Figure 4.8: Phase plane diagram near the equilibrium point $(\pi, 0)$.

Substituting x and y directly into equation (2.5) we find

$$\begin{aligned}\frac{dz_1}{dt} =& a(\frac{r}{c} + z_1) - b(\frac{r}{c} + z_1)(\frac{a}{b} + z_2) = -\frac{br}{c} z_2 - bz_1 z_2 \\ \frac{dz_2}{dt} =& -r(\frac{a}{b} + z_2) + c(\frac{r}{c} + z_1)(\frac{a}{b} + z_2) = \frac{ca}{b} z_1 + cz_1 z_2.\end{aligned}$$

(Observe that in the case of polynomials a Taylor expansion emerges from the algebra, saving us a step.) Dropping the nonlinear terms, we arrive at our linear equations,

$$\begin{aligned}\frac{dz_1}{dt} =& -\frac{br}{c} z_2 \\ \frac{dz_2}{dt} =& \frac{ca}{b} z_1.\end{aligned}$$

Proceeding as in the previous case, we solve the phase plane equation,

$$\frac{dz_2}{dz_1} = \frac{\frac{ca}{b} z_1}{-\frac{br}{c} z_2},$$

for implicit solutions,

$$\frac{ca}{b}\frac{z_1^2}{2} + \frac{br}{c}\frac{z_2^2}{2} = C,$$

which correspond with ellipses and conseqently orbital stability. Just as in the case of the pendulum equation, these orbits correspond with periodic behavior (see Figure 2.10). △

4.2 Uniqueness Theory

Example 4.3. Consider the following variant on the problem posed in Example 3.5: Ignoring air resistance, determine an exact form for the time at which an object lauched vertically from a height h with velocity v

strikes the earth. As observed in Example 3.5, Newton's second law of motion determines an equation for the height $y(t)$ of the object,

$$y(t) = -g\frac{t^2}{2} + vt + h.$$

Setting $y(t) = 0$, we find,

$$-gt^2/(2h) + vt/h + 1 = 0,$$

with solution,

$$t = \frac{-v \pm \sqrt{v^2 + 2gh}}{-g}.$$

While we know that there is only one time at which the object can strike the ground, our model gives us two different times. This is a problem of uniqueness. (In this case, the resolution is straightforward: taking $-$ makes $t > 0$ and corresponds with the time we are looking for; taking $+$ makes $t < 0$ and corresponds with the object's trajectory being traced backward in time along its parabolic arc to the ground.) △

Though the question of uniqueness arises in every type of equation—algebraic, differential, integral, integrodifferential, stochastic, etc.—we will only develop a (relatively) full theory in the case of ordinary differential equations.

Example 4.4. Consider the ordinary differential equation,

$$\frac{dy}{dt} = y^{2/3}; \quad y(0) = 0.$$

Solving by separation of variables, we find $y(t) = t^3/27$, which we compare with the following MATLAB script:

```
>>[t,y]=ode23(inline('y^(2/3)','t','y'),[0 ,.5],0)
t =
0
0.0500
0.1000
0.1500
0.2000
0.2500
0.3000
0.3500
0.4000
0.4500
0.5000
y =
0
0
0
0
0
0
0
0
0
0
0
```

According to MATLAB, the solution is $y(t) = 0$ for all t, and indeed it is straightforward to check that this is indeed a solution to the equation. In practice, this is the fundamental issue with uniqueness: If our model does not have a unique solution, we don't know whether or not the solution MATLAB (or alternative software) gives us is the one that corresponds with the phenomenon we're modeling. $\triangle$

Two critical questions are apparent: 1. When can we insure that this problem won't arise (that solutions are unique)? and 2. In the case of nonuniqueness, can we develop a theory that selects the correct solution? The second of these questions can only be answered in the context of the phenomenon we're modeling. For example, in Example 4.3, we selected $t > 0$ because we were trying to predict a future time, and only one solution satisfied $t > 0$. As we observed, however, the other solution answered a different question that might have been posed: how long ago would the object have had to leave the ground to get to height h? Fortunately, for the first of our two questions—at least in the case of ODE—we have a definitive general theorem.

Theorem 4.1. (ODE Uniqueness) Let $f(t, y) = (f_1(t, y), f_2(t, y), ..., f_n(t, y))^{\text{tr}}$ be a vector function whose components are each continuous in both t and y in some neighborhood $a \leq t \leq b$ and $a_1 \leq y_1 \leq b_1$,$a_2 \leq y_2 \leq b_2$,...,$a_n \leq y_n \leq b_n$ and whose partial derivatives $\partial_{y_l} f_k(t, y)$ are continuous in both t and y in the same neighborhoods for each $l, k = 1, ..., n$. Then given any initial point $(t_0, y_0) \in \mathbb{R} \times \mathbb{R}^n$ such that $a \leq t_0 \leq b$ and $a_k \leq y_{0_k} \leq b_k$ for all $k = 1, ...n$, any solution to

$$\frac{dy}{dt} = f(t, y); \quad y(t_0) = y_0$$

is unique.

Example 4.4 continued. Notice that our equation from Example 4.4 better not satisfy the conditions of Theorem 4.1. In this case, $f(t, y) = y^{2/3}$, which is continuous in both t (trivially) and y. Computing $\partial_y f(t, y) = \frac{2}{3} y^{-1/3}$, we see that the y-derivative of f is not continuous at the initial value $y = 0$. $\triangle$

Example 4.5. Consider again the Lotka–Volterra predator–prey model, which we can re-write in the notation of Theorem 4.1 as ($y_1 = x$, $y_2 = y$),

$$\begin{aligned}
\frac{dy_1}{dt} =& ay_1 - by_1y_2; \quad y_1(t_0) = y_{0_1} \\
\frac{dy_2}{dt} =& - ry_2 + cy_1y_2; \quad y_2(t_0) = y_{0_2}.
\end{aligned}$$

In this case, the vector $f(t, y)$ is

$$\begin{pmatrix} f_1(t, y_1, y_2) \\ f_2(t, y_1, y_2) \end{pmatrix} = \begin{pmatrix} ay_1 - by_1y_2 \\ -ry_2 + cy_1y_2 \end{pmatrix}.$$

As polynomials, $f_1, f_2, \partial_{y_1} f_1, \partial_{y_2} f_1, \partial_{y_1} f_2$, and $\partial_{y_2} f_2$ must all be continuous for all t, y_1, and y_2, so any solution we find to these equations must be unique. $\triangle$

Idea of the uniqueness proof. Before proceeding with a general proof of Theorem 4.1, we will work through the idea of the proof in the case of a concrete example. Consider the ODE

$$\frac{dy}{dt} = y^2; \quad y(0) = 1, \tag{4.5}$$

and suppose we want to establish uniqueness on the intervals $a \leq t \leq b$ and $a_1 \leq y \leq b_1$, with $0 \in [a, b]$ and $1 \in [a_1, b_1]$. We begin by supposing that $y_1(t)$ and $y_2(t)$ are both solutions to (4.5) and defining the squared difference between them as a variable,

$$E(t) := (y_1(t) - y_2(t))^2.$$

Our goal becomes to show that $E(t) \equiv 0$; that is, that $y_1(t)$ and $y_2(t)$ must necessarily be the same function. Computing directly, we have

$$\begin{aligned}\frac{dE}{dt} =&2(y_1(t) - y_2(t))(\frac{dy_1}{dt} - \frac{dy_2}{dt})\\ =&2(y_1(t) - y_2(t))(y_1(t)^2 - y_2(t)^2)\\ =&2(y_1(t) - y_2(t))(y_1(t) - y_2(t))(y_1(t) + y_2(t))\\ =&2(y_1(t) - y_2(t))^2(y_1(t) + y_2(t))\\ =&2E(t)(y_1(t) + y_2(t)).\end{aligned}$$

Since y_1 and y_2 are both assumed less than b_1, we conclude the differential inequality

$$\frac{dE}{dt} \leq 2E(t)(2b_1),$$

which upon multiplication by the (non-negative) *integrating factor* $e^{-4b_1 t}$ can be written as

$$\frac{d}{dt}[e^{-4b_1 t}E(t)] \leq 0.$$

Integrating, we have

$$\int_0^t \frac{d}{dt}[e^{-4b_1 s}E(s)]ds = e^{-4b_1 s}E(s)\Big|_0^t = e^{-4b_1 t}E(t) - E(0) \leq 0.$$

Recalling that $y_1(0) = y_2(0) = 1$, we observe that $E(0) = 0$ and consequently $E(t) \leq 0$. But $E(t) \geq 0$ by definition, so that we can concluded that $E(t) = 0$. □

Proof of Theorem 4.1. In order to restrict the tools of this proof to a theorem that should be familiar to most students, we will carry it out only in the case of a single equation. The extension to systems is almost identical, only requiring a more general form of the Mean Value Theorem.

We begin as before by letting $y_1(t)$ and $y_2(t)$ represent two solutions of the ODE

$$\frac{dy}{dt} = f(t, y); \quad y(t_0) = y_0.$$

Again, we define the squared difference between $y_1(t)$ and $y_2(t)$ as $E(t) := (y_1(t) - y_2(t))^2$. Computing directly, we have now

$$\begin{aligned}\frac{dE}{dt} =&2(y_1(t) - y_2(t))(\frac{dy_1}{dt} - \frac{dy_2}{dt})\\ =&2(y_1(t) - y_2(t))(f(t, y_1) - f(t, y_2)).\end{aligned}$$

At this point, we need to employ the Mean Value Theorem (see Appendix A), which asserts in this context that for each t there exists some number $c \in [y_1, y_2]$ so that

$$f'(c) = \frac{f(t, y_1) - f(t, y_2)}{y_1 - y_2}, \quad \text{or} \quad f(t, y_1) - f(t, y_2) = \partial_y f(t, c)(y_1 - y_2).$$

Since $\partial_y f$ is assumed continuous on the closed interval $t \in [a, b], y \in [a_1, b_1]$, the Extreme Value Theorem (see Appendix A) guarantees the existence of some constant L so that $|\partial_y f(t, y)| \leq L$ for all $t \in [a, b], y \in [a_1, b_1]$. We have, then, the so-called *Lipschitz inequality*,

$$|f(t, y_1) - f(t, y_2)| \leq L|y_1 - y_2|.$$

We conclude that

$$\frac{dE}{dt} \leq 2|y_1(t) - y_2(t)|L|y_1(t) - y_2(t)| = 2LE(t),$$

from which we concluded exactly as above that $E(t) \equiv 0$. □

4.3 Existence Theory

Existence theory is one of the most abstract topics in applied mathematics. The idea is to determine that a solution to some problem exists, even if the solution cannot be found.

Example 4.6. Prove that there exists a real solution to the algebraic equation

$$x^7 + 6x^4 + 3x + 9 = 0.$$

While actually finding a real solution to this equation is quite difficult, it's fairly easy to recognize that such a solution must exist. As x goes to $+\infty$, the left hand side becomes positive, while as x goes to $-\infty$ the left hand side becomes negative. Somewhere in between these two extremes, the left hand side must equal 0. In this way we have deduced that a solution exists without saying much of anything about the nature of the solution. (Mathematicians in general are notorious for doing just this sort of thing.) △

If we really wanted to ruin MATLAB's day, we could assign it the ODE

$$\frac{dy}{dt} = t^{-1}; \quad y(0) = 1.$$

Solving by direct integration, we see that $y(t) = \log t + C$, so that no value of C can match our initial data. (The current version of MATLAB simply crashes.) As with the case of uniqueness, we would like to insure the existence of some solution before trying to solve the equation. Fortunately, we have the following theorem, due to Picard.

Theorem 4.2. (ODE Existence)[7] Let $f(t,y) = (f_1(t,y), f_2(t,y), ..., f_n(t,y))^{\text{tr}}$ be a vector function whose components are each continuous in both t and y in some neighborhood $a \le t \le b$ and $a_1 \le y_1 \le b_1$,$a_2 \le y_2 \le b_2$,...,$a_n \le y_n \le b_n$ and whose partial derivatives $\partial_{y_l} f_k(t,y)$ are continuous in both t and y in the same neighborhoods for each $l, k = 1, ..., n$. Then given any initial point $(t_0, y_0) \in \mathbb{R} \times \mathbb{R}^n$ such that $a \le t_0 \le b$ and $a_k \le y_{0_k} \le b_k$ for all $k = 1, ...n$, there exists a solution to the ODE

$$\frac{dy}{dt} = f(t,y); \quad y(t_0) = y_0 \tag{4.6}$$

for some domain $|t - t_0| < \tau$, where $\tau > 0$ may be extremely small. Moreover, the solution y is a continuous function of the independent variable t and of the parameters t_0 and y_0.

Example 4.7. Consider the ODE

$$\frac{dy}{dt} = y^2; \quad y(0) = 1.$$

Since $f(t,y) = y^2$ is clearly continuous with continuous derivatives, Theorem 4.2 guarantees that a solution to this ODE exists. Notice particularly, however, that the interval of existence is not specified. To see exactly what this means, we solve the equation by separation of variables, to find

$$y(t) = \frac{1}{1-t},$$

from which we observe that though $f(y)$ and its derivatives are continuous for all t and y, existence is lost at $t = 1$. Referring to the statement of our theorem, we see that this statement is equivalent to saying that $\tau = 1$. Unfortunately, our general theorem does not specify τ for us a priori. △

Idea of the proof of Theorem 4.2, single equations. Consider the ODE

$$\frac{dy}{dt} = y; \quad y(0) = 1.$$

[7]The assumptions here are exactly the same as those for Theorem 4.1, so together Theorems 4.1 and 4.2 constitute a complete existence–uniqueness theory.

Our goal here is to establish that a solution exists without every actually finding the solution. (Though if we accidentally stumble across a solution on our way, that's fine too.) We begin by simply integrating both sides, to obtain the *integral equation*

$$y(t) = 1 + \int_0^t y(s)ds.$$

(Unlike in the method of separation of variables, we have integrated both sides with respect to the same variable, t.) Next, we try to find a solution by an iteration. (Technically, *Picard Iteration.*) The idea here is that we guess at a solution, say $y_{\text{guess}}(t)$ and then use our integral equation to (hopefully) improve our guess through the calculation

$$y_{\text{new guess}}(t) = 1 + \int_0^t y_{\text{old guess}}(s)ds.$$

Typically, we call our first guess $y_0(t)$ and use the initial value: here, $y_0(t) = 1$. Our second guess, $y_1(t)$, becomes

$$y_1(t) = 1 + \int_0^t y_0(s)ds = 1 + \int_0^t 1ds = 1 + t.$$

Similiarly, we compute our next guess (iteration),

$$y_2(t) = 1 + \int_0^t y_1(s)ds = 1 + \int_0^t (1+s)ds = 1 + t + \frac{t^2}{2}.$$

Proceeding similarly, we find that

$$y_n(t) = \sum_{k=0}^{n} \frac{t^k}{k!} \Rightarrow \lim_{n\to\infty} y_n(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!},$$

and our candidate for a solution becomes $y(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!}$, an infinite series amenable to such tests as the *integral test*, the *comparison test*, the *limit comparison test*, the *alternating series test*, and the *ratio test.* The last step is to use one of these tests to show that our candidate converges. We will use the ratio test, reviewed in Appendix A. Computing directly, we find

$$\lim_{k\to\infty} \frac{a_{k+1}}{a_k} = \frac{\frac{t^{k+1}}{(k+1)!}}{\frac{t^k}{k!}} = \frac{t^k t}{(k+1)k!} \cdot \frac{k!}{t^k} = \frac{t}{k+1} = 0, \text{ for all } t.$$

We conclude that $y(t)$ is indeed a solution. Observe that though we have developed a series representation for our solution, we have not found a closed form solution. (What is the closed form solution?) △

Idea of the proof of Theorem 4.2, higher order equations and systems. We consider the ODE

$$\begin{gathered} y''(t) + y(t) = 0; \\ y(0) = 0; \qquad y(1) = 1. \end{gathered} \tag{4.7}$$

In order to proceed as above and write (4.7) as an integral equation, we first write it in the notation of Theorem 4.2 by making the substitutions $y_1(t) = y(t)$ and $y_2(t) = y'(t)$:

$$\begin{aligned} \frac{dy_1}{dt} &= y_2; \quad y_1(0) = 0 \\ \frac{dy_2}{dt} &= -y_1; \quad y_2(0) = 1. \end{aligned}$$

(Notice that the assumptions of Theorem 4.2 clearly hold for this equation.) Integrating, we obtain the integral equations,

$$\begin{aligned} y_1(t) =& \int_0^t y_2(s)ds \\ y_2(t) =& 1 - \int_0^t y_1(s)ds. \end{aligned}$$

Our first three iterations become,

$$\begin{aligned} y_1(t)^{(1)} =& \int_0^t 1ds = t \\ y_2(t)^{(1)} =& 1 - \int_0^t 0ds = 1 \\ y_1(t)^{(2)} =& \int_0^t 1ds = t \\ y_2(t)^{(2)} =& 1 - \int_0^t sds = 1 - \frac{t^2}{2} \\ y_1(t)^{(3)} =& \int_0^t (1 - \frac{s^2}{2})ds = t - \frac{t^3}{3!} \\ y_2(t)^{(3)} =& 1 - \int_0^t sds = 1 - \frac{t^2}{2}. \end{aligned}$$

(By the way, I never said this was the world's most efficient algorithm.) Continuing, we find that

$$y(t) = y_1(t) = t - \frac{t^3}{3!} + \frac{t^5}{5!} - ... = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{t^{2k-1}}{(2k-1)!}.$$

Again, we can apply the ratio test to determined that this series converges (to what?). □

Proof of Theorem 4.2. As with Theorem 4.1, we will only prove Theorem 4.2 in the case of single equations. The proof in the case of systems actually looks almost identical, where each statement is replaced by a vector generalization. I should mention at the outset that this is by far the most technically difficult proof of the semester. Not only is the argument itself fairly subtle, it involves a number of theorems from advanced calculus (e.g. M409).

Integrating equation (4.6), we obtain the integral equation

$$y(t) = y_0 + \int_{t_0}^t f(s, y(s))ds.$$

Iterating exactly as in the examples above, we begin with y_0 and compute y_1, y_2,... according to

$$y_{n+1}(t) = y_0 + \int_{t_0}^t f(s, y_n(s))ds; \quad n = 0, 1, 2, ...$$

As both a useful calculation and a warmup for the argument to come, we will begin by estimating $\|y_1 - y_0\|$, where where $\| \cdot \|$ is defined similarly as in Theorem A.6 by

$$\|y(t)\| := \sup_{|t-t_0|<\tau} |y(t)|.$$

We compute

$$|y_1(t) - y_0| = |\int_{t_0}^t f(s, y_0)ds|.$$

Our theorem assumes that f is continuous on $s \in [t_0, t]$ and hence bounded, so there exists some constant M so that $\|f(s, y_0)\| \leq M$. We have, then

$$|y_1(t) - y_0| \leq \tau M.$$

Observing that the right-hand side is independent of t we can take supremum over t on both sides to obtain

$$\|y_1 - y_0\| \leq \tau M.$$

Finally, since we are at liberty to take τ as small as we like we will choose it so that $0 < \tau \leq \epsilon/M$, for some $\epsilon > 0$ to be chosen.

We now want to look at the difference between two successive iterations and make sure the difference is getting smaller—that our iteration is actually making progress. For $|t - t_0| < \tau$, we have

$$\begin{aligned} |y_{n+1}(t) - y_n(t)| =& |\int_{t_0}^{t} \Big(f(s, y_n(s)) - f(s, y_{n-1}(s))\Big) ds| \\ \leq& \int_{t_0}^{t} L|y_n(s) - y_{n-1}(s)| ds \leq L\tau \sup_{|t-t_0| \leq \tau} |y_n(t) - y_{n-1}(t)|. \end{aligned}$$

Taking supremum over both sides (and observing that t has become a dummy variable on the right-hand side), we conclude

$$\|y_{n+1} - y_n\| \leq L\tau \|y_n - y_{n-1}\|,$$

Since τ is to be taken arbitrarily small, we can choose it to be as small as we like, and take $0 < \tau \leq L/2$. In this way, we have

$$\|y_{n+1} - y_n\| \leq \frac{1}{2} \|y_n - y_{n-1}\|.$$

We see that, indeed, on such a small interval of time our iterations are getting better. In fact, by carrying this argument back to our initial data, we find

$$\begin{aligned} \|y_{n+1} - y_n\| \leq& \frac{1}{2} \|y_n - y_{n-1}\| \leq \frac{1}{2} \cdot \frac{1}{2} \|y_{n-1} - y_{n-2}\| \\ \leq& \frac{1}{2^n} \|y_1 - y_0\| \leq \frac{\epsilon}{2^n}. \end{aligned}$$

In this way, we see that for $n > m$

$$\begin{aligned} \|y_n - y_m\| =& \|\sum_{k=m}^{n-1} (y_{k+1} - y_k)\| \leq \sum_{k=m}^{n-1} \|y_{k+1} - y_k\| \\ \leq& \epsilon \sum_{k=m}^{\infty} \frac{1}{2^k} = \frac{\epsilon}{2^{m-1}}. \end{aligned}$$

We conclude that

$$\lim_{n>m\to\infty} \|y_n - y_m\| = \lim_{n>m\to\infty} \frac{\epsilon}{2^{m-1}} = 0,$$

and thus by Cauchy's Convergence Condition (Theorem A.6) $y_n(t)$ converges to some function $y(t)$, which is our solution. □

A Fundamental Theorems

One of the most useful theorems from calculus is the Implict Function Theorem, which addresses the question of existence of solutions to algebraic equations. Instead of stating its most general version here, we will state exactly the case we use.

Theorem A.1. (Implicit Function Theorem) Suppose the function $f(x_1, x_2, ..., x_n)$ is C^1 in a neighborhood of the point $(y_1, y_2, ..., y_n)$ (the function is continuous at this point, and its derivative is also continuous at this point). Suppose additionally that

$$f(y_1, y_2, ..., y_n) = 0$$

and

$$\partial_{y_1} f(y_1, y_2, ..., y_n) \neq 0.$$

Then there exists a neighborhood of $(y_2, y_3, ..., y_n)$ and a function $g(y_2, y_3, ..., y_n)$ so that

$$y_1 = g(y_2, y_3, ..., y_n).$$

Theorem A.2. (Mean Value Theorem) Suppose $f(x)$ is a differentiable function on the interval $x \in [a, b]$. There there exists some number $c \in [a, b]$ so that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Theorem A.3. (Extreme Value Theorem) Suppose $f(x)$ is a function continuous on a closed interval $x \in [a, b]$. Then $f(x)$ attains a bounded absolute maximum value $f(c)$ and a bounded absolute minimum value $f(d)$ at some numbers c and d in $[a, b]$.

Theorem A.4. (The Ratio Test) For the series $\sum_{k=1}^{\infty} a_k$, if $\lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| = L < 1$, then the series is absolutely convergent (which means that not only does the series itself converge, but a series created by taking absolute values of the summands in the series also converges). On the other hand, if $\lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| = L > 1$ or $\lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| = \infty$ the series diverges. if $\lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| = 1$, the ratio test is inconclusive.

Theorem A.5. (Cauchy's Convergence Condition) Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of points and consider the limit $\lim_{n\to\infty} a_n$. A necessary and sufficient condition that this limit be convergent is that

$$\lim_{n>m\to\infty} |a_n - a_m| = 0.$$

Theorem A.6. (Cauchy's Convergence Condition for functions, in exactly the form we require) Let the series of functions $\{y_n(t)\}_{n=1}^{\infty}$ be defined for $t \in [a, b]$, and define $\|\cdot\|$ by the relation

$$\|y(t)\| := \sup_{t\in[a,b]} |y(t)|.$$

Then if

$$\lim_{n>m\to\infty} \|y_n(t) - y_m(t)\| = 0,$$

we have

$$\lim_{n\to\infty} y_n(t)$$

converges uniformly on $t \in [a, b]$.

References

[M442.1] M442 Lecture Notes: *On the mathematics of games of chance.*

Index