MATHEMATICAL MODELING
A Comprehensive Introduction

Gerhard Dangelmayr and Michael Kirby
Department of Mathematics
Colorado State University
Fort Collins, Colorado, 80523

PRENTICE HALL, *Upper Saddle River, New Jersey 07458*

Contents

APPENDICES

Preface

These materials are being developed with support from National Science Foundation Award no. 0126650 entitled *A Mathematical Modeling Program for Undergraduates in Science, Mathematics, Engineering and Technology.*

The objective of this project is the development of innovative educational materials that incorporate a novel educational approach and perspective to enhance the teaching and learning of mathematics for the purposes of knowledge discovery. The general undergraduate educated with these materials will possess a readily applicable toolbox of mathematical ideas for quantifying real world problems as well as problem solving skills, and possibly the most importantly, the ability to interpret results and further understanding.

Our pedagogical perspective consists of the observation that mathematical modeling is often taught backwards. An application of interest is presented and then appropriate mathematical tools are subsequently invoked. The beginner is left with the obvious concern. How does one decide which method to use on a new problem? Our proposed solution to this dilemma is to teach mathematics first and then show why a given mathematical methodology can be applied to the modeling problem. We will be successful if the student completes their modeling course based on these materials with a good sense of what makes various mathematical methods inherently different. Furthermore, students that are aware of the fundamental distinguishing characteristics of the array of methodologies should now be equipped to address this question of central importance in modeling, i.e., which method when!

This text is the first of two planned works to establish "proof of concept" of a new approach to teaching mathematical modeling. The scope of the text is the basic theory of modeling from a mathematical perspective. A second applications focussed text will build on the basic material of the first volume.

It is typical that students in a mathematical modeling class come from a wide variety of disciplines. In addition, their preparation and mathematical sophistication can vary as widely as their areas of interest. This heterogeneity makes the teaching and learning of mathematical modeling a significant challenge. One of the main student prototypes is a intelligent although possibly mathematically naive student that must learn mathematically modeling to make progress in an area of research. If a course or textbook does not provide the necessary information for these good students to bridge educational gaps students everyone suffers. Indeed, most textbooks fail to be accessible to such audiences.

With enhancing accessibility as our motivation, we propose to implement a simple pedagogical device to facilitate the use of the text by students of widely varying backgrounds. This device consists of graded levels of presentation denoted by (E) for elementary, (I) for intermediate and (A) for advanced.

- (E) Mathematical beginners will find much of interest in the elementary sections as well as foundation material for further study. The diligent student can use this self-contained treatment to pave the way to reading of more advanced sections. The basic properties of mathematical techniques will be presented with an emphasis on how methods lead to specific applications.

- (I) Intermediate material builds on the elementary material and extends the students expertise. Often intermediate material will involve computer experiments to stimulate more theoretical discussions in the advanced material. A good understanding of intermediate material should permit a student to develop new applications of central mathematical ideas.

- (A) Advanced material will provide mathematically mature students with a solid theoretical foundation for the subject. Mastery of this subject matter should provide the student with the insight required to further develop mathematical models.

If a section is labeled as (E) then all its subsections are at the same level. If it is not labelled, then each individual subsection will be labelled for level of difficulty.

These texts will be pilot tested at Colorado State University during the course of development and will incorporate a fundamentally new approach to modeling through general mathematical principles rather than ad hoc lists of methods and techniques. These methods will be demonstrated within the context of on-going state-of-the-art interdisciplinary research projects. (Such an approach will have the added advantage of broadening students perspectives and appreciation for the nature of basic university research.) The basic aim of the materials is to present an innovative approach to inform and educate students about the power and importance of basic mathematics and mathematical modeling in the process of knowledge discovery.

Michael Kirby
Gerhard Dangelmayr

CHAPTER 1

Mathematical Modeling

Mathematical modeling is becoming an increasingly important subject as computers expand our ability to translate mathematical equations and formulations into concrete conclusions concerning the world, both natural and artificial, that we live in.

1.1 EXAMPLES OF MODELING

Here we do a quick tour of several examples of the mathematical process. We present the models as finished results as opposed to attempting to develop the models.

1.1.1 Modeling with Difference Equations

Consider the situation in which a variable changes in discrete time steps. If the current value of the variable is a_n then the predicted value of the variable will be a_{n+1}. A mathematical model for the evolution of the (still unspecified) quantity a_n could take the form

$$a_{n+1} = \alpha a_n + \beta$$

In words, the new value is a scalar multiple of the old value offset by some constant β. This model is common, e.g., it is used for modeling bank loans. One might amend the model to make the dependence depend on more terms and to include the possibility that every iteration the offset can change, thus,

$$a_{n+1} = \alpha_1 a_n + \alpha_2 a_n^2 + \beta_n$$

This could correpsond to, for example, a population model where the the migration levels change every time step. In some instances, it is clear that information required to predict a new value goes back further than the current value, e.g.,

$$a_{n+1} = a_n + a_{n-1}$$

Note now that two initial values are required to evolve this model. Finally, it may be that the form of the difference equations are unknown and the model must be written

$$a_{n+1} = f(a_n, a_{n-1}, a_{n-M-1})$$

Determining the nature of f and the step M is at the heart of model formulation with difference equations. Often observed data can be employed to assist in this effort.

1.1.2 Modeling with Ordinary Differential Equations

Although modeling with ordinary differential equations shares many of the ideas of modeling with the difference equations discussed above, there are many fundamen-

tal differences. At the center of these differences is the assumption that time is a continuous variable.

One of the simplest differential equations is also an extremely important model, i.e.,

$$\frac{dx}{dt} = \alpha x$$

In words, the rate of change of the quantity x depends on the amount of the quantity. If $\alpha > 0$ then we have exponential growth. If $\alpha < 0$ the situation is exponetial decay. Of course additional terms can be added that fundamentally alter the evolution of $x(t)$. For example

$$\frac{dx}{dt} = \alpha_1 x + \alpha_2 x^2$$

The model formulation again requires the development of the appropriate right-hand side.

In the above model the value x on the right hand side is implicitly assumed to be evaluated at the time t. It may be that there is evidence that the instantaneous rate of change at time t is actually a function of a previous time, i.e.,

$$\frac{dx}{dt} = f(x(t)) + g(x(t-\tau))$$

This is referred to as a delay differential equation.

1.1.3 Modeling with Partial Differential Equation

In the previous sections on modeling the behaviour of a variable as a function of time we assumed that there was only one independent variable. Many situations arise in practice where the number of independent variables is larger than two. For spatio-temporal models we might have time and space (hence the name!), e.g.,

$$\frac{\partial f}{\partial t} = \alpha \frac{\partial^2 f}{\partial x^2}$$

or

$$\frac{\partial^2 f}{\partial t^2} = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

1.1.4 Optimization

In many modeling problems the goal is to compute the "best" solution. This may correspond to maximizing profit in a company, or minimizing loss in a conflict. It is no surprise that optimization techniques take a central seat in the mathematical modeling literature.

Now one may allow $x \in \mathbb{R}^n$ and require that

$$x^* = \arg\min f(x)$$

The quantity $f(x)$ is referred to as the *objective function* while the vector x consists of decision variables. Because x sits in $\mathbb{R}^n$ the problem is referred to as unconstrained.

Alternatively, one might require that the solution x have all positive components. If we refer to this set as S then the optimization problem is constrained

$$x^* = \arg\min_{x \in S} f(x)$$

If the objective function as well as the equations that define the constraint set are linear, than the optimization problem is called a *linear programming problem*. Otherwise, the problem is referred to as a nonlinear programming problem. As we shall see, solution methods for linear and nonlinear programming problems are very different.

1.1.5 Modeling with Simulations

Many problems may afford a mathematical formulation yet be analytically intractable. In these situations a computer can implement the mathematics literally and repetitively often times to extreme advantage.

Simulating Games.

- What is the probability that you can win a game of solitaire?
- What is the best strategy for playing blackjack?
- Given a baseball team consisting of certain players, in what order should they hit?

On the other hand, computer simulations can be employed to model evolution equations. Applications in the realm of fluid dynamics and weather prediction are well established. A striking new example of such simulation modeling is attempting to model electrical activity in the brain.

1.1.6 Function Fitting: Data Modeling

Often data is available from a process to assist in the modeling. How can functions be computed that reflect the relationships between variables in the data. Produce a model

$$y = f(x; w)$$

and using the set of input output pairs compute the parameters w. In some cases the form of f may be guessed. In other cases a model free approach can be used.

1.2 THE MODELING PROCESS

The goal in all modeling problems is *added value*. Something novel must be learned from the modeling process or one has completed an exercise in futility, or mathematical wheel spinning, depending on your perspective. There are many obvious questions the answers to which have inherent added value. For example:

- Should a stock be bought or sold?
- Is the earth becoming warmer?

- Does creating a law have a positive or negative societal effect?
- What is the most valuable property in monopoly?

Clearly this is a very small start to an extremely long list.

1.2.1 An Algorithm for Modeling?

The modeling process has a sequence of common steps that serve as an abstraction for the modeler:

- Identify the problem and questions.
- Identify the relevant variables in a problem.
- Simplify until tractable.
- Relate these variables mathematically.
- Solve.
- Does the solution provide added value?
- Tweak model and compare solutions.

1.3 THE DELICATE SCIENCE OF ERRORS

If one had either infinite time or infinite computing power error analysis would presumably be a derelict activity: all models would be absolutely accurate. Obviously, in reality, this is not the case and a well-accepted *modus operandi* in modeling is committing admissible errors. Of course, in practice, the science is more *ad hoc*. If terms in an equation introduce computational difficulties the immediate question arises as to what would happen if those terms are ignored? In theory we would rather keep them but in practice we can't afford to. Thus the delicate science of modeling concerns retaining just enough features to make the model useful but not so many as too make it more expensive to compute than necessary to get out the desirable information.

We illustrate this concept by examining the seemingly innocuous junior high school problem

$$\epsilon x^2 + x + 1 = 0$$

Of course we can solve this problem exactly using the quadratic formula[1]

$$x = -\frac{1}{2\epsilon} \pm \frac{\sqrt{1-4\epsilon}}{2\epsilon} \tag{1.1}$$

For a moment, let us assume that the quadratic term were actually an unknown term, e.g.,

$$\epsilon f(x) + x + 1 = 0$$

[1] If you don't recall this, then the famous Science Fiction write Robert Heilein suggested you not be allowed to vote.

and that actually computing f might be rather expensive. We might argue that if ϵ were very small that this term could safely be ignored. Now let us return to the simple case of $f(x) = x^2$. If ϵ is taken as zero then clearly it follows that

$$x = -1$$

is the unique solution. However, we know from our quadratic equation however that if $\epsilon = 0.0000001$ (any non-zero number would do), then there are two solutions rather than one. So we have actually lost a potentially important solution by ignoring what appeared to be a small quantity. In addition, we may also have introduced inaccuracies into the obtained solution and this issue must be explored.

In essence we are concerned with how quickly the solution changes about the point $\epsilon = 0$. A quick graph of Equation (1.1) reveals that the solution changes rather quickly.

To see how this solution changes as a function of ϵ consider the series expansion

$$x = a_0 + a_1\epsilon + a_2\epsilon^2 + a_3\epsilon^3 + \dots$$

Substituting this expansion into the original quadratic results in the new equation

$$a_0 + 1 + (a_0^2 + a_1)\epsilon + (2a_0a_1 + a_2)\epsilon^2 + \cdots = 0$$

Setting the coefficients of the different powers of ϵ to zero gives the series solution for x as

$$x = -1 - \epsilon - 2\epsilon^2 + \dots \tag{1.2}$$

So if $\epsilon \approx 0.01$ we can conclude the error is on the order of 1% and the error will grow quickly with ϵ.

This problem is explored further in the exercises and function iteration is introduced to track down the 2nd solution in the quadratic equation. For further discussion of these ideas see [4].

1.4 PURPOSE OF THIS COURSE

The primary goal of this course is to assist the student to develop the skills necessary to effectively employ the ideas of mathematics to solve problems. At the simplest level we seek to promote an understanding of why mathematics is useful as a language for characterizing the interaction and relationships among quantifiable concepts, or in mathematical terms, variables. Throughout the text we emphasize the notion of *added value* and why it is the driving force behind modeling. For a given mathematical model to be deemed a success something must be learned that was not obvious without the modeling procedure. Very often added value comes in the form of a prediction. In the absence of added value the modeling procedure becomes an exercise not unrelated to digging a ditch simply to fill it back up again.

The emphasis in this course is on learning why certain mathematical concepts are useful for modeling. We proceed *from mathematics to models* rather than the popular reverse approach and downplay interdisciplinary expertise required in many specific contexts. We firmly believe that by focusing on mathematical concepts the ability to transfer knowledge from one setting to another will be significantly enhanced. Hence, we emphasize the efficacy of certain mathematics for constructing models.

PROBLEMS

1.1. Name three problems that might be modeled mathematically. Why do you think mathematics may provide a key to each solution. What is the added value in each case?

1.2. Consider the differential equation

$$\frac{dx}{dt} = x$$

Translate this model to a difference equation. Compare the solutions and discuss.

1.3. Consider the equation

$$x^2 + \epsilon x - 1 = 0$$

for small ϵ. How does ignoring the middle term ϵx change your solution? Is this a serious omission?

1.4. Using a Taylor series expansion express the solution to the quadratic equation in Equation 1.1 as a series. Include terms up to cubic order.

1.5. Find the cubic term in the expansion in Equation (1.2).

1.6. One approach to determining zeros of a general function, i.e., computing roots to $f(x) = 0$, is to rewrite the problem as $f(x) = x - g(x)$ and to employ the iteration $x_{n+1} = g(x_n)$.

(a) If we take

$$g(x) = -\frac{1}{x}$$

show that the iteration can be written

$$x_{n+1} = -\frac{1}{\epsilon}(1 + \frac{1}{x_n})$$

(b) Let $x_0 = -1/\epsilon$ and compute x_1. By considering the Taylor series of the solution of the quadratic equation argue that this is a two term approximation to the *missing* solution.

(c) Compute x_2.

1.7. **(a)** Substitute $x = y/\varepsilon$ into the equation

$$\varepsilon x^2 + x + 1 = 0 \tag{1.3}$$

and multiply the resulting equation for y by ε. Show that this leads to

$$y^2 + y + \varepsilon = 0. \tag{1.4}$$

When $\varepsilon = 0$, the equation (1.4) has two solutions: $y = 0$ and $y = -1$. This suggests that (1.4) allows us to compute both solutions of (1.3) through a perturbation analysis.

(b) Reproduce the solution to (1.3) given by Eq. (1.2) by computing a solution of the form

$$y = b_1\varepsilon + b_2\varepsilon^2 + b_3\varepsilon^3 + \dots$$

for (1.4).

(c) Proceed similarly using the form

$$y = -1 + c_1\varepsilon + c_2\varepsilon^2 + \dots$$

to find the expansion of the missing solution to (1.3).

CHAPTER 2

Qualitative Modeling with Functions

It is often surprising that very simple mathematical modeling ideas can produce results with added value. Indeed, the solutions may be elegant and provide quality of understanding that obviates further exploration by more technical or complex means. In this chapter we explore a few simple approaches to qualitatively modeling phenomena with well-behaved functions.

2.1 MODELING SPECIES PROPAGATION

This problem concerns the factors that influence the number of species existing on an island. The discussion is adapted from [1].

One might speculate that factors affecting the number of species could include

- Distance of the island from the mainland
- Size of the island

Of course limiting ourselves to these influences has the dual effect of making a tractable model that needs to be recognized as omitting many possible factors.

The number of species may increase due to new species discovering the island as a suitable habitat. We will refer to this as the *migration rate*. Alternatively, species may become extinct due to competition. We will refer to this as the *extinction rate*. This discussion will be simplified by employing an aggregate total for the number of species and not attempting to distinguish the nature of each species, i.e., birds versus plants.

Now we propose some basic modeling assumptions that appear reasonable.

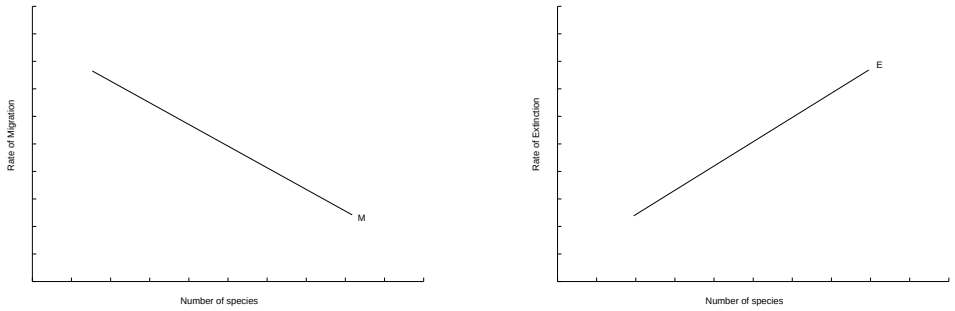
> *The migration rate of new species decreases as the number of species on the island increases.*

The argument for this is straight forward. The more species on an island the smaller the number of new species there is to migrate. See Figure 2.1 (a) for a qualitative picture.

> *The extinction rate of species increases as the number of species on the island increases.*

Clearly the more species there are the more possibilities there are for species to die out. See Figure 2.1 (b) for a qualitative picture.

If we plot the extinction rate and the migration rate on a single plot we identify the point of intersection as an equilibrium, i.e., the migration is exactly offset by the extinction and the number of species on the island is a constant. We



(a) Migration curve.

(b) Extinction curve

FIGURE 2.1: Qualitative form of the migration and extinction curves.

will assume in this discussion that we are considering islands for which the number of species is roughly constant over time, i.e., they are in a state of equilibrium.

Now we consider whether this simple model provides any added value. In particular, can it be used to address our questions posed at the outset.

First, what is the effect of the distance of the island from the mainland on the number of bird species? One can characterize this effect by a shift in the migration curve. The further the island is away from the mainland, the less likely a species is to successfully migrate. Thus the migration curve is shifted down for *far* islands and shifted up for *near* islands. Presumably, this distance of the island from the mainland has no impact on the extinction curve. Thus, by examining the shift in the equilibrium, we may conclude that the number of species on an island decreases as the island's distance from the mainland increases. See Figure 2.2.

Note in this model we assume that the time-scales are small enough that new species are not developed via evolution. While this may seem reasonable there is evidence that in some extreme climates, such as those found in the Galapagos Islands, variation may occur over shorter periods. There have been 140 different species of birds

2.2 SUPPLY AND DEMAND

In this section we sketch a well-known concept in economics, i.e., supply and demand. We shall see that relatively simple laws, when taken together, afford interesting insight into the relationship between producers and consumers. Furthermore, we may use this framework to make predictions such as

- What is the impact of a tax on the sale price?
- What is the impact an increase in employees wages on sales price? Can the owner of the business pass this increase on to the consumer?

Law of Supply: An increase in the price of a commodity will result in an increase of the amount supplied.

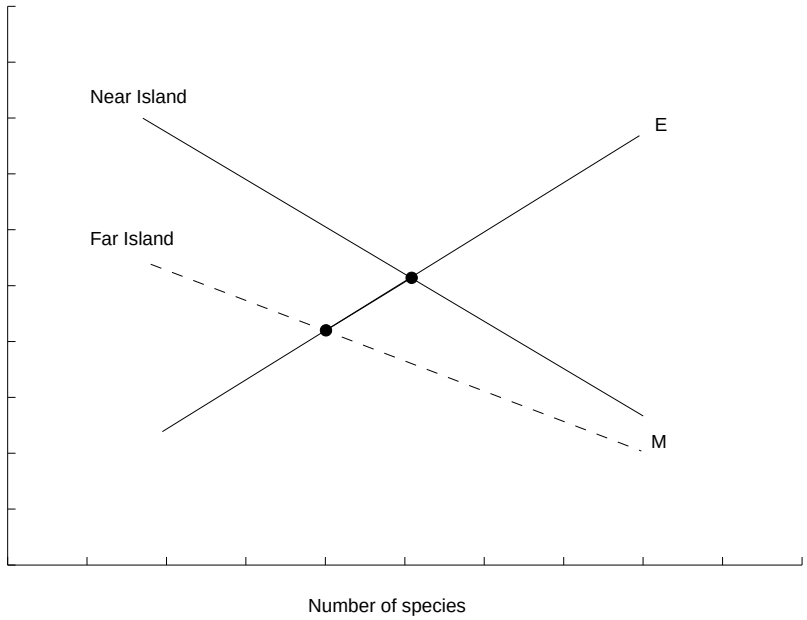


FIGURE 2.2: The effect of distance of the island from the mainland is to shift the migration curve. Consequently the equilibrium solution dictates a smaller number of species will be supported for islands that are farther away from the mainland.

Law of Demand: If the price of a commodity increases, then the quantity demanded will decrease.

Thus, we may model the supply curve qualitatively by a monotonically increasing function. For simplicity we may assume a straight line with positive slope. Analogously, we may model the demand curve qualitatively by a monotonically decreasing function, which again we will take as a straight line.

A flat demand curve may be interpreted as consumers being very sensitive to the price of a commodity. If the price goes up just a little, then the quantity in demand goes down significantly. Steep and flat supply and demand curves all have similar qualitative interpretations (see the problems).

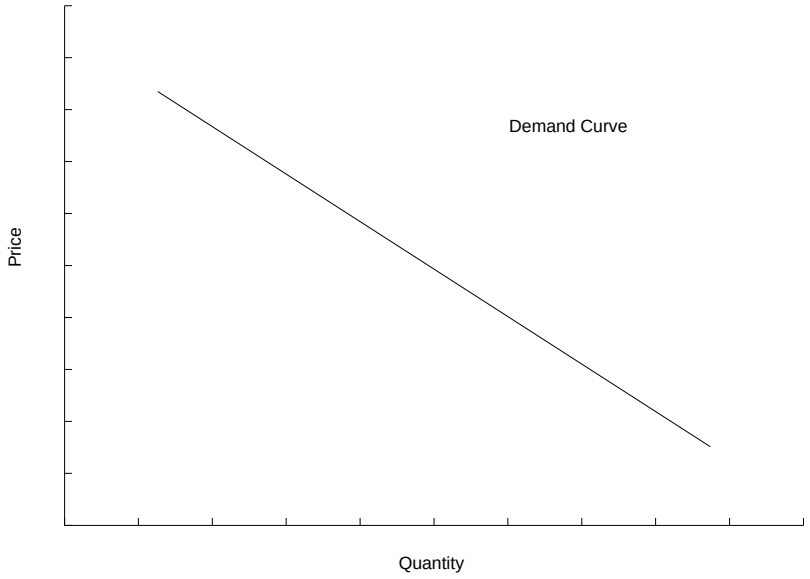
2.2.1 Market Equilibrium

Given a supply curve and a demand curve we may plot them on the same axis and note their point of intersection (q_*, p_*). This point is special for the following reason:

- The seller is willing to supply q_* at the price p_*
- The demand is at the price p_* is q_*

So both the supplier(s) and the purchaser(s) are happy economically speaking.

(a) Supply curve



(b) Demand curve

FIGURE 2.3: (a) Qualitative form of supply and demand curves.

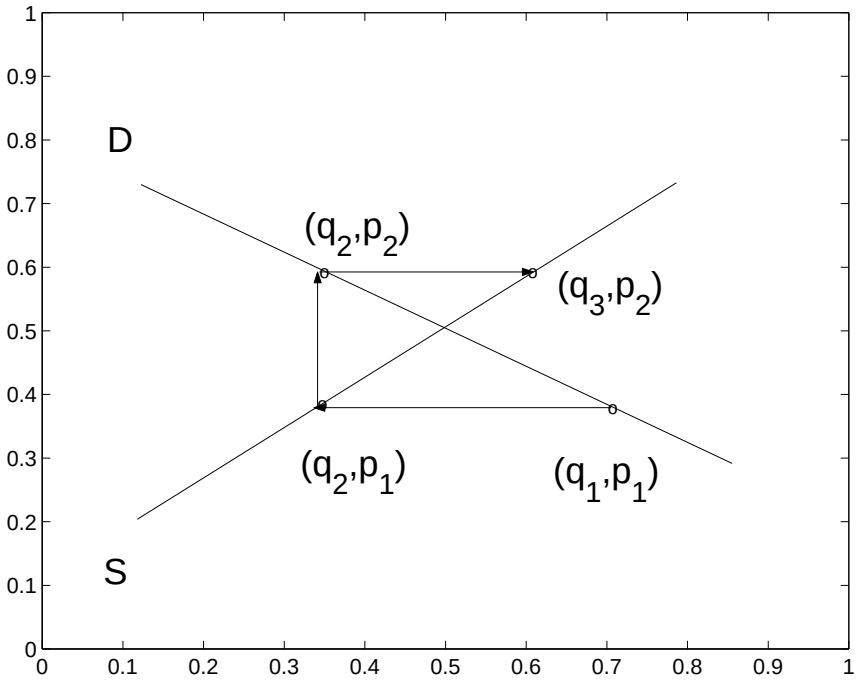


FIGURE 2.4: The cobweb model illustrating a sequence of market adjustments.

2.2.2 Market Adjustment

Of course, in general markets do not exist in the perfect economic utopia described above. We may model the market adjustment as a sequence of points on the demand and supply curves.

Based on market research it is estimated that consumers will demand a quantity q_1 at a price p_1. The supply and demand curves will permit a prediction of how the market will evolve. For simplicity, we will assume that the initial point (q_1, p_1) is on the demand curve to the right of the equilibrium point.

At the price p_1 the supplier looks to his supply curve and proposes to sell a reduced quantity q_2. Thus we move from right to left horizontally. Note that moving vertically to the supply curve would not make sense as this would correspond to offering the quantity q_1 at an increased price. These goods will not sell at this price.

From the point (q_1, p_2) the consumer will respond to the new reduced quantity q_2 by being willing to pay more. This corresponds to moving vertically upward to the new point (q_2, p_2) on the supply curve.

Now the supplier adjusts to the higher price being paid in the market place by increasing the quantity produced to q_3. This process then continues, in theory, until an equilibrium is reached. It is possible that this will never happen, at least not without a basic adjustment to the shape of either the supply of demand curves, for example through cost cutting methods such as improved efficiency, or layoffs.

2.2.3 Taxation

The effect of a new tax on a product is to shift the demand curve down because consumers will not be willing to pay as much for the product (before the tax). Note that this leads to a new equilibrium point which reduces the price paid to the seller per item and reduces the quantity supplied by the producer. Thus one may conclude from this picture that the effect of a tax on alcohol is to reduce consumption as well as profit for the supplier. See Figure 2.5.

2.3 MODELING WITH PROPORTION AND SCALE

In the previous sections we have considered how simple functions may be employed to qualitatively model various situations and produce added value. Now we turn to considerations that assist in determining the nature of these functional dependencies in more complex terms.

2.3.1 Proportion

If a quantity y is *proportional* to a quantity x then we write

$$y \propto x$$

by which is meant

$$y = kx$$

for some constant of proportionality k.

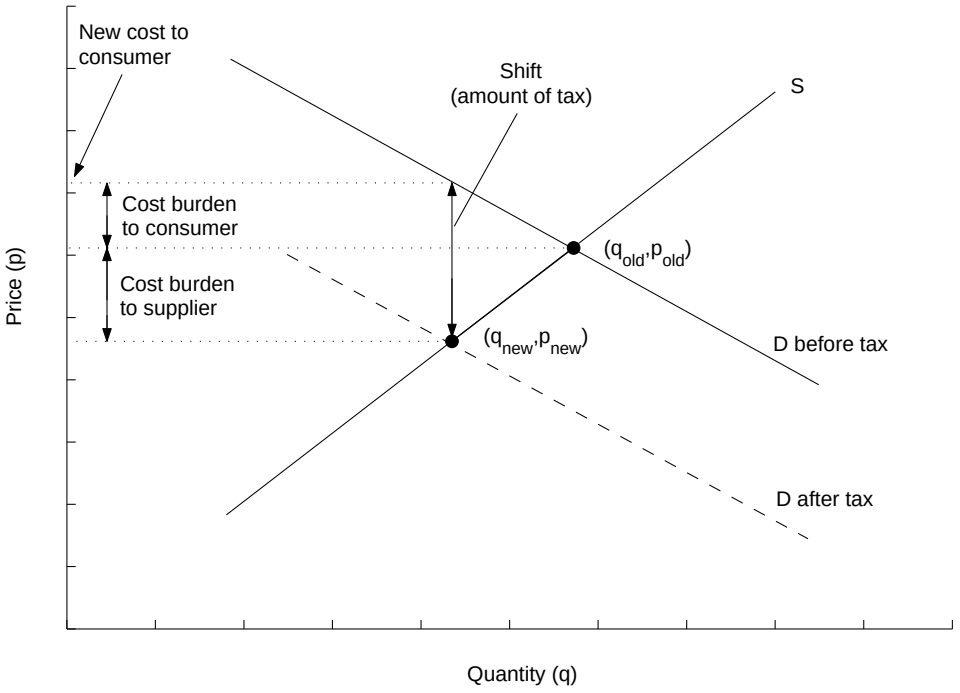


FIGURE 2.5: A tax corresponds to a downwards shift in the demand curve.

EXAMPLE 2.1

In 1678 Robert Hooke proposed that the restoring force F of a spring is proportional to its elongation e, i.e.,

$$F \propto e$$

or,

$$F = ke$$

where k is the *stiffness* of the spring.

Note that the property of proportionality is symmetric, i.e.,

$$y \propto x \rightarrow x \propto y \tag{2.1}$$

and transitive, i.e.,

$$y \propto x \quad \text{and} \quad z \propto y \rightarrow z \propto x \tag{2.2}$$

EXAMPLE 2.2

If $y = kx + b$ where k, b are constants, then

$$y \not\propto x$$

but

$$y - b \propto x$$

Inverse proportion. If $y \propto 1/x$ then y is said to be *inversely* proportional to x.

EXAMPLE 2.3

If y varies inversely as the square-root of x then

$$y = \frac{k}{\sqrt{x}}$$

Joint Variation. The volume of a cylinder is given by

$$V = \pi r^2 h$$

where r is the radius and h is the height. The volume is said to vary *jointly* with r^2 and h, i.e.,

$$V \propto r^2 \quad \text{and} \quad V \propto h$$

EXAMPLE 2.4

The volume of a given mass of gas is proportional to the temperature and inversely proportional to the pressure, i.e., $V \propto T$ and $V \propto 1/P$, or,

$$V = k\frac{T}{P}$$

EXAMPLE 2.5

Frictional drag due to the atmosphere is jointly proportional to the surface area S and the velocity v of the object.

Superposition of Proportions. Often a quantity will vary as the sum of proportions.

EXAMPLE 2.6

The stopping distance of a car when an emergency situation is encountered is the sum of the reaction time of the driver and the amount of time it takes for the breaks to dissipate the energy of the vehicle. The reaction distance is proportional to the velocity. The distance travelled once the breaks have been hit is proportional to the velocity squared. Thus,

$$\text{stopping distance} = k_1 v + k_2 v^2$$

EXAMPLE 2.7

Numerical error in the computer estimation of the center difference formula for the derivative is given by

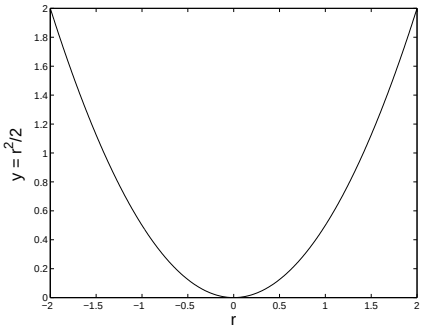
$$e(h) = \frac{c_1}{h} + c_2 h^2$$

where the first term is due to roundoff error (finite precision) and the second term is due to truncation error. The value h is the distance δx in the definition of the derivative.

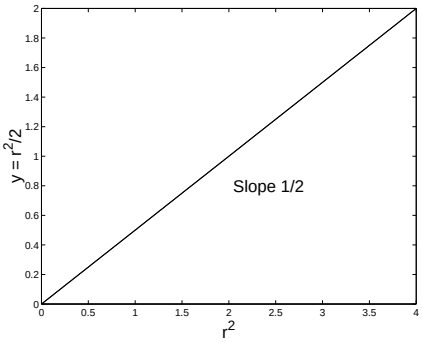
Direct Proportion. If

$$y \propto x$$

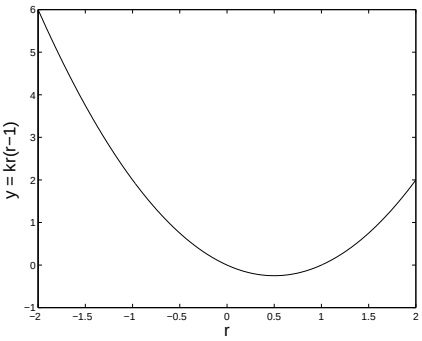
we say y varies in *direct* proportion to x. This is not true, for example, if $y \propto r^2$. On the other hand, we may construct a direct proportion via the obvious change of variable $x = r^2$. This simple trick always permits the investigation of the relationship between two variables such as this to be recast as a direct proportion.



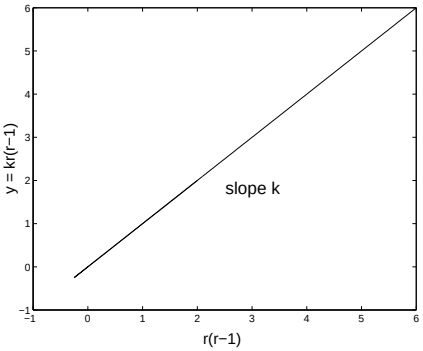
(a) Plot of y against r for $y = r^2/2$.



(b) Plot of $y = r^2/2$ against r^2.



(c) Plot of y against r for $y = kr(r+1)$.



(d) Plot of $y = kr(r+1)$ against$r(r+1)$

FIGURE 2.6: Simple examples of how a proportion may be converted to a direct proportion.

2.3.2 Scale

Now we explore how the size of an object can be represented by an appropriate length scale if we restrict our attention to replicas that are *geometrically similar*. For example, a rectangle with sides l_1 and w_1 is geometrically similar to a rectangle with sides l_2 and w_2 if

$$\frac{l_1}{l_2} = \frac{w_1}{w_2} = k \tag{2.3}$$

As the ratio $\kappa = l_1/w_1$ characterizes the geometry of the rectangle it is referred to as the *shape factor*. If two objects are geometrically similar, then it can be shown that they have the same shape factor. This follows directly from multiplying Equation (2.3) by the factor l_2/w_1, i.e.,

$$\frac{l_1}{w_1} = \frac{l_2}{w_2} = k\frac{l_2}{w_1}$$

Characteristic Length.

Characteristic length is useful concept for characterizing a family of geometrically similar objects. We demonstrate this with an example.

Consider the area of a rectangle of side l and width w where l and w may vary under the restriction that the resulting rectangle be geometrically similar to the rectangle with length l_1 and width w_1. An expression for the area of the varying triangle can be simplified as a consequence of the constraint imposed by geometric similarity. To see this

$$\begin{aligned} A &= lw \\ &= l(\frac{w_1 l}{l_1}) \\ &= \kappa l^2 \end{aligned}$$

where $\kappa = w_1/l_1$, i.e., the shape factor. See Figure 2.7 for examples of characteristic lengths for the rectangle.

EXAMPLE 2.8

Watering a farmer's rectangular field requires an amount of area proportional to the area of the field. If the characteristic length of the field is doubled, how much additional water q will be needed, assuming the new field is geometrically similar to the old field? Solution: $q \propto l^2$, i.e., $q = kl^2$. Hence

$$q_1 = kl_1^2$$

$$q_2 = kl_2^2$$

Taking the ratio produces

$$\frac{q_1}{q_2} = \frac{l_1^2}{l_2^2}$$

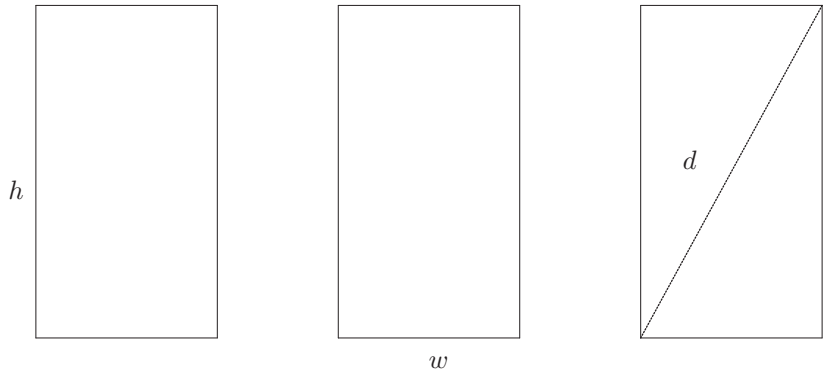


FIGURE 2.7: The height l_1, the width l_2 and the diagonal l_3 are all characteristic lengths for the rectangle.

Now if $q_2 = 100$ acre feet of water are sufficient for a field of length $l_2 = 100$, how much water will be required for a field of length $l_1 = 200$? Sol.

$$q_1 = q_2 \frac{l_1^2}{l_2^2} = 100 \frac{200^2}{100^2} = 400 \,\text{acre feet} \quad \square$$

EXAMPLE 2.9

Why are gymnasts typically short? It seems plausible that the ability A, or natural talent, of gymnast would be proportional to strength and inversely proportional to weight, i.e.,

$$A \propto \text{strength}$$

and

$$A \propto \frac{1}{\text{weight}}$$

and taken jointly

$$A \propto \frac{\text{strength}}{\text{weight}}$$

One model for strength is that the strength of a limb is proportional to the cross-sectional area of the muscle. The weight is proportional to the volume (assuming constant density of the gynmast). Now, assuming all gymnasts are geometrically similar with characteristic length l

$$\text{strength} \propto \text{muscle area} \propto l^2$$

and

$$\text{weight} \propto \text{volume} \propto l^3$$

so the ability A follows

$$A \propto \frac{l^2}{l^3} \propto \frac{1}{l}$$

So shortness equates to a talent for gymnastics. This problem was originally introduced in [2]. □

EXAMPLE 2.10

Proportions and terminal velocity. Consider a uniform density spherical object falling under the influence of gravity. The object will travel will constant (terminal) velocity if the accelerating force due to gravity $F_g = mg$ is balance exactly by the decelerating force due to atmospheric friction $F_d = kSv^2$; S is the cross-sectional surface area and v is the velocity of the falling object. Our equilibrium condition is then

$$F_g = F_d$$

Since surface area satisfies $S \propto l^2$ it follows $l \propto S^{1/2}$. Given uniform density $m \propto w \propto l^3$ so it follows $l \propto m^{1/3}$. Combining proportionalities

$$m^{1/3} \propto S^{1/2}$$

from which it follows by substitution into the force equation that

$$m \propto m^{2/3}v^2$$

or, after simplifying,

$$v \propto m^{1/6}$$

□

EXAMPLE 2.11

In this example we will attempt to model observed data displayed in Table 2.1 that relates the heart rate of mammals to there body weight. From the table we see that we would like to relate the heart rate as a function of body weight. Smaller animals have a faster heart rate than larger ones. But how do we estimate this proportionality?

We begin by assuming that all the energy E produced by the body is used to maintain heat loss to the environment. This heat loss is in turn proportional to the surface area s of the body. Thus,

$$E \propto s$$

The energy available to the body is produced by the process of respiration and is assumed to be proportional to the oxygen available which is in turn proportional

mammal	body weight (g)	pulse rate
shrew[2]	3.5	782
pipistrelle bat[1]	4	660
bat[2]	6	588
mouse[1]	25	670
hamster[2]	103	347
kitten[2]	117	300
rat[1]	200	420
rat[2]	252	352
guinea pig[1]	300	300
guinea pig[1]	437	269
rabbit[2]	1,340	251
rabbit[1]	2,000	205
oppossum[2]	2,700	187
little dog[1]	5,000	120
seal[2]	22,500	100
big dog[1]	30,000	85
goat[2]	33,000	81
sheep[1]	50,000	70
human[1]	70,000	72
swine[2]	100,000	70
horse[2]	415,000	45
horse[1]	450,000	38
ox[1]	500,000	40
elephant[1]	3,000,000	48

TABLE 2.1: Superscript 1 data source A.J. Clark; superscript 2 data source Altman and Dittmer. See also [1] and [2].

to the blood flow B through the lungs. Hence, $B \propto s$ If we denote the pulse rate as r we may assume

$$B \propto rV$$

where V is the volume of the heart.

We still need to incorporate the body weight w into this model. If we take W to be the weight of the heart assuming constant density of the heart it follows

$$W \propto V$$

Also, if the bodies are assumed to be geometrically similar then $w \propto W$ so by transitivity $w \propto V$ and hence

$$B \propto rw$$

Using the geometric similarity again we can relate the body surface area s to its weight w. From characteristic length scale arguments

$$v^{1/3} \propto s^{1/2}$$

so

$$s \propto w^{2/3}$$

from which we have $rw \propto w^{2/3}$ or

$$r = kw^{-1/3}$$

To validate this model we plot $w^{-1/3}$ versus r for the data Table 2.8. We see that for the larger animals with slower heart rates that this data appears linear and suggests this rather crude model actually is supported by the data. For much smaller animals there appear to be factors that this model is not capturing and the data falls off the line.

2.4 DIMENSIONAL ANALYSIS

In this chapter we have explored modeling with functions and proportion. In some instances, such as the mammalian heart rate, it is possible to cobble enough information together to actually extract a model; in particular, to identify the functional form for the relationship between the dependent and independent variables. Now we turn to a surprisingly powerful and simple tool known as *dimensional analysis*[1].

Dimensional analysis operates on the premise that equations contain terms that have units of measurement and that the validity of these equations, or laws, are not dependent on the system of measurement. Rather these equations relate variables that have inherent physical dimensions that are derived from the fundamental dimensions of *mass, length* and *time*. We label these dimensions generically as M, L and T, respectively.

As we shall see, dimensional analysis provides an effective tool for mathematical modeling in many situations. In particular, some benefits include

[1]This dimension should not be confused with the usual notion of geometric dimension.

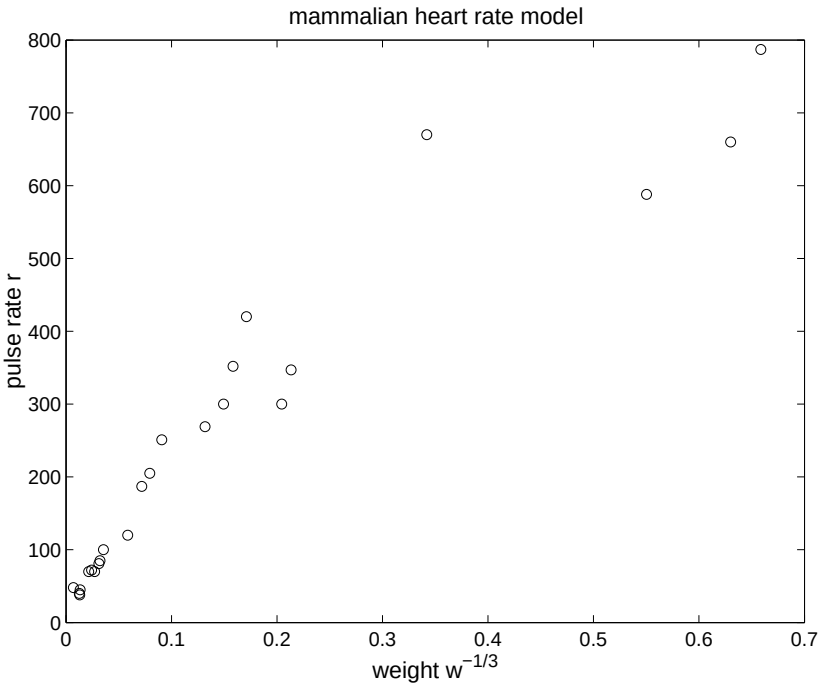


FIGURE 2.8: Testing the model produced by proportionality. For the model to fit, the data should sit on a straight line eminating from the origin.

- determination of the form of a joint proportion
- reduce number of variables in a model
- enforcement of dimensional consistency
- ability to study scaled versions of models

2.4.1 Dimensional homogeneity

An equation is said to be *dimensionally homogeneous* if all the terms in the equation have the same physical dimension.

EXAMPLE 2.12

All the laws of physics are dimensionally homogeneous. Consider Newton's law

$$F = ma$$

The units on the right side are

$$M \cdot \frac{L}{T^2}$$

so we conclude that the physical dimension of a force must be MLT^{-2}. □

EXAMPLE 2.13

The equation of motion of a linear spring with no damping is

$$m\frac{d^2x}{dt^2} + kx = 0$$

What are the units of the spring constant? Dimensionally we can recast this equation as

$$MLT^{-2} + M^aL^bT^cL = 0$$

Matching exponents for each dimension permits the calculation of a, b and c.

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

Thus we conclude that the spring constant has the dimensions MT^{-2}. □

EXAMPLE 2.14

Let v be velocity, t be time and x be distance. The model equation

$$v^2 = t^2 + \frac{x}{t}$$

is dimensionally inconsistent.

EXAMPLE 2.15

An angle may be defined by the formula

$$\theta = \frac{s}{r}$$

where the arclength s subtends the angle θ and r is the radius of the circle. Clearly this angle is dimensionless.

2.4.2 Discovering Joint Proportions

If in the formulation of a problem we are able to identify a dependent and one or more independent variables, it is often possible to identify the form of a joint proportion. The form of the proportion is actually constrained by the fact that the equations must be dimensionally consistent.

EXAMPLE 2.16 Drag Force on an Airplane

In this problem we consider the drag force F_D on an airplane. As our model we propose that this drag force (dependent variable) is proportional to the independent variables

- cross-sectional area A of airplane
- velocity v of airplane
- density ρ of the air

As a joint proportion we have

$$F_D = kA^a v^b \rho^c$$

where a, b and c are unknown exponents. As a consequence of dimensional consistency we have

$$\begin{aligned} MLT^{-2} &= (M^0L^0T^0)(L^2)^a(\frac{L}{T})^b(\frac{M}{L^3})^c \\ &= M^cL^{2a-3c+b}T^{-b} \end{aligned}$$

From the M exponent we conclude $c = 1$. From the T exponent $b = 2$ and from the L exponent it follows that $1 = 2a - 3c + b$, whence $a = 1$. Thus the only possibility for the form of this joint proportion is

$$F_D = kAv^2\rho$$

Note that if the density of were a constant it would be appropriate to simplify this dependency as

$$F = \tilde{k}Av^2$$

but now the constant $\tilde{k}$ actually has dimensions. □

2.4.3 Procedure for Nondimensionalization

Consider the nonlinear model for a pendulum

$$\frac{d^2\theta}{dt^2} = -\frac{g}{l}\sin\theta$$

Based on the terms in this model we may express the solution very generally as a relationship between these included terms, i.e.,

$$\phi(\theta, g, l, t) = 0$$

Note that the angle in this model is dimensionless but the other variables all have dimensions. We can convert this equation into a new equation where none of the terms have dimensions. This will be referred to, for obvious reasons, as a dimensional form of the model.

To accomplish this, let

$$\tau = \frac{t}{\sqrt{l/g}}.$$

The substitution of variables may be accomplished by noting that

$$\frac{d^2\theta}{dt^2} = \frac{d^2\theta}{\frac{l}{g}d\tau^2}$$

Thus, after cancelation, the dimensionless form for the nonlinear pendulum model is

$$\frac{d^2\theta}{d\tau^2} = -\sin\theta$$

Now the solution has the general form

$$f(\theta, \tau) = 0,$$

or equivalently,

$$f(\theta, \sqrt{\frac{l}{g}}t) = 0$$

This is a special case of a more general theory.

The Buckingham π-theorem. Any dimensionally homogeneous equation with physical variables $x_1, \ldots, x_m$ expressed

$$\phi(x_1, \ldots, x_m)$$

may be rewritten in terms of its associated dimensionless variables $\pi_1, \ldots, \pi_n$ as

$$f(\pi_1, \ldots, \pi_n) = 0$$

where

$$\pi_k = x_1^{a_{k1}} \ldots x_m^{a_{km}}$$

2.4.4 Modeling with Dimensional Analysis

Now we consider two examples of the application of the ideas described above concerning dimensional analysis. In each of these examples there is more than one dimensionless parameter and it is appropriate to apply the Buckingham π-theorem.

The Pendulum. In this example the goal is to understand how the period of a pendulum depends on the other parameters that describe the nature of the pendulum. The first task is to identify this set of parameters that act as the independent variables on which the period P depends.

Obvious candidates include From this list we are motivated to write

variable	symbol	dimensions
mass	m	M
length	l	L
gravity	g	LT^{-2}
angle	θ_0	$M^0L^0T^0$
period	P	T

TABLE 2.2: Parameters influencing the motion of a simple pendulum.

$$P = \phi(m, l, g, \theta_0)$$

As we shall see, attempting to establish the form of ϕ directly is unnecessarily complicated. Instead, we pursue the idea of dimensional analysis.

To begin this modeling procedure, we compute the values of a, b, c, d and e that make the quantity

$$\pi = m^a l^b g^c \theta_0^d P^e$$

a dimensionless parameter. Again, this is done by equating exponents on the fundamental dimensions

$$M^0L^0T^0 = M^aL^b(LT^{-2})^c(M^0L^0T^0)^dT^e$$

From M^0: $0 = a$.
From L^0: $0 = b + c$.
From T^0: $0 = -2c + e$.
From this we may conclude that

$$\pi = m^0 l^{-c} g^c \theta_0^d P^{2c}$$

or, after collecting terms,

$$\pi = \theta_0^d \left(\frac{gP^2}{l}\right)^c$$

where π is dimensionless for any values of d and c. Thus we have found a complete set of dimensionless parameters

$$\pi_1 = \theta_0$$

and

$$\pi_2 = \sqrt{\frac{g}{l}}P$$

Since the period P of the pendulum is based on dimensionally consistent physical laws we may apply the Buckingham π-theorem. In general,

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

which we rewrite as

$$\pi_2 = h(\pi_1)$$

which now becomes

$$P = \sqrt{\frac{l}{g}}h(\theta_0)$$

We may draw two immediate conclusions from this model.

- The period depends on the square root of the length of the pendulum.
- The period is independent of the mass

Of course we have not really shown these conclusions to be "true". But now we have something to look for that can be tested. We could test these assertions and if they contradict our model then we would conclude that we are missing an important factor that governs the period of the pendulum. Indeed, as we have neglected drag forces due to friction it seems our model will have limited validity.

The functional form of h may now be reasonable calculated as there is only one independent variable θ_0. If we select several different initial displacements $\theta_0(i)$ and measure the period for each one we have a set of domain–range values

$$h(\theta_0(i)) = P_i\sqrt{\frac{g}{l}}$$

to which a data fitting procedure may now be applied.

The damped pendulum. We assumed that there was no damping of this pendulum above due to air resistance. We can include a drag force F_D by augmenting the list of relevant parameters to

$$m, l, g, \theta_0, P, F_D$$

Now our dimensionless parameter takes the form

$$\pi = m^a l^b g^c \theta_0^d P^e F_D^f$$

Converting to dimensions

$$M^0L^0T^0 = M^aL^b(LT^{-2})^c(M^0L^0T^0)^dT^e(MLT^{-2})^f$$

As

$$0 = a + f$$

it is no longer possible to immediately conclude that $a = 0$. In fact, it is not. (See problems).

Fluid Flow. Consider the parameters governing the motion of an oil past a spherical ball bearing. Let's assume they include:

variable	symbol	dimensions
velocity	v	LT^{-1}
density	ρ	ML^{-3}
gravity	g	LT^{-2}
radius	l	L
viscosity	μ	$ML^{-1}T^{-1}$

TABLE 2.3: Parameters influencing the motion of a fluid around a submerged body.

The dimensionless combination has the form

$$\pi = v^a \rho^b l^c g^d \mu^e$$

Using the explicit form of the physical dimensions for each term we have

$$M^0 L^0 T^0 = (LT^{-1})^a (ML^{-3})^b (L)^c (LT^{-2})^d (ML^{-1}T^{-1})^e$$

Again, matching exponents

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

Sinc there are three equations and five unknowns the system is said to be undetermined. Given these numbers, we anticipate that there we can solve for three variable in terms of the other two. Of course, we can solve in terms of *any* of the two variables. For example,

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

Plugging these constraints into our expression for π gives

$$\pi = \left(\frac{v^2}{lg}\right)^{-d} \left(\frac{\rho l v}{\mu}\right)^{-e}$$

Thus, our two dimensionless parameters are the *Froude number*

$$\pi_1 = \frac{v^2}{lg}$$

and the *Reynolds number*

$$\pi_2 = \frac{v\rho l}{\mu}$$

For further discussion see Giordano, Wells and Wilde, UMAP module 526.

PROBLEMS

2.1. By drawing a new graph, show the effect of the size of the island on the

- extinction curve
- migration curve

Now predict how island size impacts the number of species on the island. Does this seem reasonable?

2.2. Give an example of a commodity that does not obey the

- law of supply
- law of demand

and justify your claim.

2.3. Translate into words the qualitative interpretation of the slope of the supply and demand curves. In particular, what is the meaning of a

- flat supply curve?
- steep supply curve?
- steep demand curve?

2.4. Consider the table of market adjustments below. Assuming the first point is on the demand curve, compute the equations of both the demand and supply curve. Using these equations, find the missing values A, B, C, D. What is the equilibrium point? Do you think the market will adjust to it?

quantity	price
3	0.7
0.14	0.7
0.14	0.986
0.1972	0.986
$A =?$	$B =?$
$C =?$	$D =?$

2.5. Using the cobweb plot show an example of a market adjustment that oscillates wildly out of control. Can you describe a qualitative feature of the supply and demand curves that will ensure convergence to an equilibrium?

2.6. Consider the effect of a price increase on airplane fuel (kerosene) on the airline industry. What effect does this have on the supply curve? Will the airline industry be able to pass this cost onto the flying public? How does your answer differ if the demand curve is flat versus steep?

2.7. Prove properties 2.1 and 2.2.

2.8. Is the temperature measured in degrees Fahrenheit proportional to the temperature measured in degrees centigrade?

2.9. Consider the Example 2.6 again. Demonstrate the proportionalities stated. For the case of the breaking distance equate the work done by the breaks to the dissipated kinetic energy of the car.

2.10. Items at the grocery store typically come in various sizes and the cost per unit is generally smaller for larger items. Model the cost per unit weight by considering the superposition of proportions due to the costs of

- production
- packaging

- shipping

the product. What predictions can you make from this model. This problem was adapted from Bender [1].

2.11. Go to your nearest supermarket and collect data on the cost of items as a function of size. Do these data behave in a fashion predicted by your model in the previous problem?

2.12. In this problem take the diagonal of a rectangle as it's length scale l. Show by direct calculation that this can be used to measure the area, i.e.,

$$A = \alpha l^2$$

Determine the constant of proportionality α in terms of the shape factor of the rectangle.

2.13. Consider a radiator designed as a spherical shell. If the characteristic length of the shell doubles (assume the larger radiator is geometrically similar to the smaller radiator) what is the effect on the amount of heat loss? What if the design of the radiator is a parallelpiped instead?

2.14. How does the argument in Example 2.10 change if the falling object is not spherical but some other irregular shape?

2.15. Extend the definition of geometric similarity for

- parallelpipeds
- irregularly shaped objects

Can you propose a computer algorithm for testing whether two objects are geomtrically similar?

2.16. Consider the force on a pendulum due to air friction modeled by

$$F_D = \kappa v^2$$

Determine the units of κ.

2.17. Newton's law of gravitation states that

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

where F is the force between two objects of masses m_1, m_2 and r is the distance between them.

(a) What is the physical dimension of G?

(b) Compute two dimensionless products π_1 and π_2 and show explicitly that they satisfy the Buckingham π-theorem.

2.18. This problem concerns the pendulum example described in subsection 2.4.4. Repeat the analysis to determine the dimensionless parameter(s) but now omit the gravity term g. Discuss.

2.19. This problem concerns the pendulum example described in subsection 2.4.4. Repeat the analysis for determining all the dimensionless parameters but now include a parameter κ associated with the drag force of the form $F_D = \kappa v$. Hint: first compute the dimensions of κ.

2.20. Convert the equation governing the distance travelled by a projectile,

$$\frac{d^2x}{dt^2} = \frac{-gR^2}{(x+R)^2},$$

to the form

$$\frac{d^2y}{d\tau^2} = \frac{-1}{(y+1)^2},$$

where y and τ are dimensionless.

2.21. Reconsider the example in subsection 2.4.4. Instead of solving for a, b, c in terms of d, e solve for c, d, e in terms of a and b. Show that now

$$\pi'_1 = \frac{v}{\sqrt{lg}}$$

and

$$\pi'_2 = \frac{\rho l^{3/2} g^{1/2}}{\mu}$$

Show also that both π'_1 and π'_2 can be written in terms of π_1 and π_2.

2.22. Consider an object with surface area A traveling with a velocity v through a medium with kinematic viscosity μ and density ρ.

(a) Assuming the effect of μ is small compute the drag force due to the density F_ρ.

(b) Assuming the effect of ρ is small compute the drag force due to the kinematic viscosity F_μ.

c) Compute the dimensionless ratio of these drag forces and discuss what predictions you can make.

2.23. Assume a drag force of the form

$$F_d = \kappa v^2$$

acts on a pendulum in addition to the gravity force. Use dimensional analysis to show that the solution of the pendulum equation can be written in the form

$$\theta = \psi(t\sqrt{l/g}, l\kappa/m).$$

2.24. How does the required power P of a helicopter engine depend on the length of the rotors l? The rotors are pushing air so presumable the density ρ as well as the weight of the helicopter $w = mg$ are variables that affect the power requirement. Draw a sketch of your result plotting P versus l. See [3] for more discussion of this problem.

Bibliography

1. Edward A. Bender. *An Introduction to Mathematical Modeling.* John Wiley & Sons, New York, 1978.
2. F.R. Giordano, M.D. Weir, and W.P. Fox. *A First Course in Mathematical Modeling.* Brooks/Cole, Pacific Grove, 1997.
3. T.W. Körner. *The Pleasures of Counting.* Cambridge University Press, Cambridge, U.K., 1996.
4. C.C. Lin and L.A. Segel. *Mathematics Applied to Deterministic Problems in the Natural Sciences.* Macmillan, New York, 1974.

CHAPTER 3

Linear Programming

Linear programming, like its nonlinear counterpart, is a method for making decisions based on solving a mathematical optimization problem. The general field of linear programming has been a major area of applied mathematical research in the last 50 years. A combination of new algorithms, e.g., the simplex method, and widely available computing power now make this an indispensable tool for the mathematical modeler.

We begin our discussion of linear programming by presenting the basic mathematical formulation and terminology in general terms. We will follow this with a number of examples of problems that may be formulated in terms of linear programs. Our goal here is to obtain an abstract understanding of what a linear program is and to develop an intuition that will assist the modeler in assessing whether linear programming is the right tool for a given problem.

Consider a linear function of the variables $(x_1, \ldots, x_n)$,

$$F(x_1, \ldots, x_n) = f_1 x_1 + f_2 x_2 + \cdots + f_n x_n$$

where the parameters f_i are known. We seek to pick the values of all the x_i, referred to as *decision variables*, so as to maximize $F(x_1, \ldots, x_n)$ which is referred to as the *objective function*. Clearly picking each $x_i = \infty$ (or even just one) would provide a maximum, albeit meaningless. The interest arises when the values of the x_i are constrained, e.g.,

$$a_{11} x_1 + \cdots + a_{1n} x_n \leq b_1$$

Based on the application many constraints are possible so we write

$$a_{i1} x_1 + \cdots + a_{in} x_n \leq b_i$$

for $i = 1, \ldots, m$. Note that these constraints are also linear in the decision variables. We may interpret this system of constraints geometrically as defining a region, i.e., a continuum of points such that all the constraints are simultaneously satisfied. This region is referred to as the *feasible set* S. So we may view the optimization problem as one to find the maximum value of the objective function over the feasible set S.

We now formulate this optimization problem in terms of vectors and matrices. Let $x = (x_1, \ldots, x_n)^T$ be the (column) vector of the unknown variables, and let $f = (f_1, \ldots, f_n)^T$ be the vector of coefficients of the objective function, $F(x) = f^T x$. We also introduce the $m \times n$ matrix A whose entries are the coefficients in the inequality constraints, $(A)_{ij} = a_{ij}$. If a and b are vectors of the same length then we write $a \geq b$ if $a_i \geq b_i$ holds for all components.

DEFINITION **1**. A linear program associated with f, A, and b is the minimum problem

$$\min_x f^T x$$

or the maximum problem

$$\max_x f^T x$$

subject to the constraint

$$Ax \leq b.$$

3.1 EXAMPLES OF LINEAR PROGRAMS

In this section we survey a variety of applications that fit exactly into the formulation of the abstract linear program.

3.1.1 Red or White?

A winemaker would like to decide how many bottles of red wine and how many bottles of white wine to produce. Given his expertise is in red wine making he can sell a bottle of red wine for \$12 while he can only sell a bottle of white wine for \$7. Clearly the winemaker would seek to maximize his profits, and, having recently completed a course in mathematical modeling, proceeds to construct the objective function

$$F(x_1, x_2) = 12x_1 + 7x_2$$

where the decision variables are the number of bottles of red wine to produce x_1 and the number of bottles of white wine to produce, i.e., x_2.

Aging wine in wooden or glass-lined vats is an integral component of the production process, but due to limited space the wine must be aged for a limited time. The wine maker has determined that red wine should be aged two years per bottle and white wine one year per bottle and his facilities allow that each batch is limited to 10,000 bottle-years (5 bottles of red and 3 bottles of white require a total of 13 bottle years ripening time). Thus the winemaker formulates a constraint

$$2x_1 + x_2 \leq 10000$$

Also the volume of grapes that may be processed is limited and it takes 3 gallons of grapes to make a bottle of red wine and two gallons of grapes to make a bottle of white wine. Furthermore, the winery can only process a total of 16,000 gallons of grapes for each batch. Thus, the winemaker produces the additional constraint

$$3x_1 + 2x_2 \leq 16000$$

Now the winemaker would like to determine how many bottles of each wine to produce as well as how much money he will expect to make. Note that we must also require that negative bottles of wine are not allowed so

$$x_1 \geq 0$$

and

$$x_2 \geq 0$$

3.1.2 How Many Fish?

A child with a new 29 gallon fish tank asks her daddy to put as many fish in the tank as possible. Sensing that too many fish is not a good thing, the dad asks the pet shop owner how many fish can go into a tank. The answer was more complex than anticipated. "You can put one inch of fish in per gallon of water." The little girl then added that she wanted only the big orange fish (Gouramis) and the small stripy fish (Zebra Danios).

As the child seeks to maximize the total number of fish her objective function is

$$F(x_1, x_2) = x_1 + x_2$$

where x_1 is the number of Gouramis and x_2 is the number of Zebra Danios.

Additionally, a full grown Gourami is two inches long while a Danio is just one inch long. The constraint of not exceeding 29 inches of total fish length can now be written

$$2x_1 + x_2 \leq 29$$

Danios are very active fish and actually require twice as much food as Gouramis. Each Danio eats 4 grams/day of fish flakes while the slower Gourami eats 2 grams/day. The dad decides that he would prefer not to go broke buying fish food and thus wants to limit the tank to 50 grams/day. Thus, we have the constraint

$$2x_1 + 4x_2 \leq 50$$

The pet shop owner adds, by the way, that Danios need to live in schools of at least 5 fish or they don't do well. Thus

$$x_2 \geq 5$$

Additionally, the little girl stipulates that she must have at least two Gouramis as they are known to kiss (hence the term Kissing Gouramis) so we add

$$x_1 \geq 2$$

How many Gouramis and Danios can the little girl have in her tank?

3.2 GEOMETRIC SOLUTION OF A 2D LINEAR PROGRAM

Let us now solve the winemaker's linear programming problem using graphical techniques. Recalling the problem:

- Objective function: $F(x_1, x_2) = 12x_1 + 7x_2$
- Constraint 1: $3x_1 + 2x_2 \leq 16000$
- Constraint 2: $2x_1 + x_2 \leq 10000$
- Constraint 3: $x_1 \geq 0$
- Constraint 4: $x_2 \geq 0$

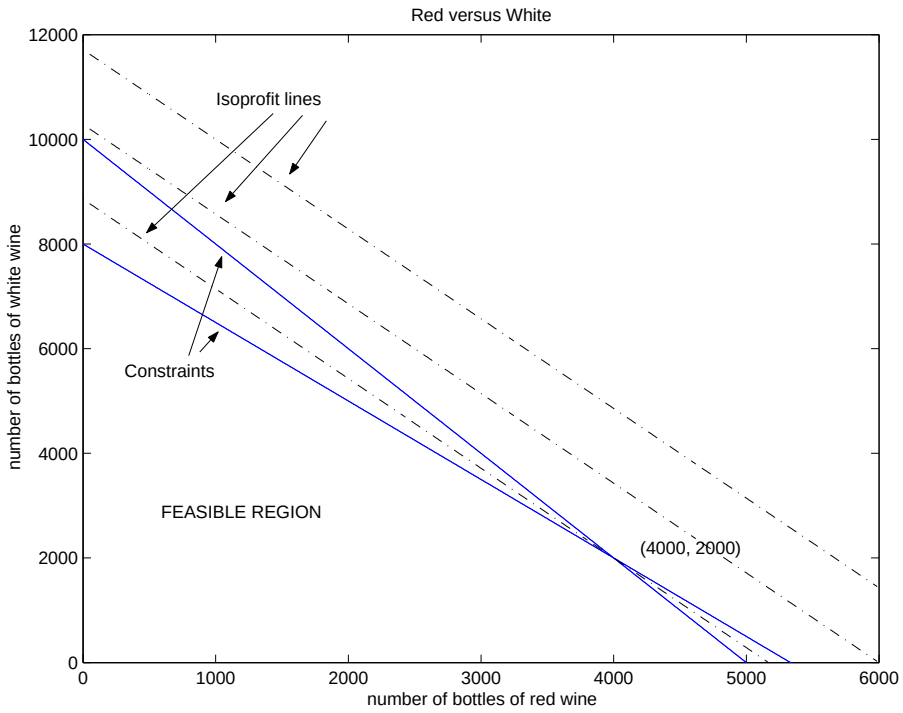


FIGURE 3.1: Geometric picture of the linear programming problem.

First, let us identify the feasible set. Again, this is the intersection of all the regions defined by the constraints. (Note that this set is independent of the objective function.) The boundary of the first constraint is defined by the equality

$$x_2 = 8000 - \frac{3}{2}x_1$$

The constraint may be viewed as a half-plane with this line dividing the region of allowed points from the unallowed points. It is easy to identify which region is the allowed region by considering a single point. For example, is the origin a point that satisfies the first constraint? Since the answer is clearly yes we know that the set of points that satisfies constraint 1 consists of the halfplane defined by $x_2 = 8000 - \frac{3}{2}x_1$ that contains origin.

Similarly, the second constraint defines a halfplane of points containing the origin and bounded by the line

$$x_2 = 10000 - 2x_1$$

The intersection of constraints 3 and 4 is the first quadrant of the x_1x_2 plane.

The intersection of all of these constraints as shown in Figure 3.1 constitutes the feasible set. Now we must pick the point in the feasible set that maximizes the objective function.

We can define an *isoprofit line* to be

$$12x_1 + 7x_2 = c$$

For all points on this line the profit is the same. We can see that as c decreases the line shifts towards the origin. So the goal is to pick the isoprofit line with the largest value of c such that x_1, x_2 is a point in the feasible set. Graphically we see that the first point the descending isoprofit line will touch is the vertex of the intersection of constraints 1 and 2. This is easily calculated to be $(4000, 2000)$.

Thus, the solution to the winemaker's linear programming problem is that he should produce 4000 bottles of red and 2000 bottles of white and that this will lead to a maximum profit of $62,000.

3.3 SENSITIVITY ANALYSIS

Often the coefficients in a linear programming model are known only approximately. Thus, it is interesting to know what the impact of modifying the terms present in the model. How is the objective function impacted? How does the optimal solution change? These questions are the subject of *sensitivity analysis*.

3.3.1 Price Sensitivity

First we examine how changing the price of a bottle of white wine impacts the optimal solution. Letting the price of the white wine be a variable w we now have the linear program

- Objective function: $F(x_1, x_2) = 12x_1 + wx_2$
- Constraint 1: $3x_1 + 2x_2 \leq 16000$

- Constraint 2: $2x_1 + x_2 \leq 10000$
- Constraint 3: $x_1 \geq 0$
- Constraint 4: $x_2 \geq 0$

From our graphical solution we know that any isoprofit line with slope between -2 and -3/2 will produce the same optimal solution of $(4000, 2000)$. Since the slope of the isoprofit line is $-12/w$ this condition is

$$-2 < \frac{12}{w} < -\frac{3}{2}$$

from which we conclude that the price of the white wine may vary as

$$6 < w < 8$$

with the solution unchanged as $(4000, 2000)$. Further examination produces Table 3.1. The double arrows here mean that any point on the isoprofit curve containing these points produces the same profit.

cost of white wine	optimal solution
$6 < w < 8$	(4000, 2000)
$w = 6$	$(4000, 2000) \leftrightarrow (5000, 0)$
$w = 8$	$(4000, 2000) \leftrightarrow (0, 8000)$
$w < 6$	(5000, 0)
$w > 8$	(0, 8000)

TABLE 3.1: The effect of pricing the white wine on the optimal solution.

3.3.2 Resource Sensitivity

Now we let the number of gallons of grapes, α, and the number of bottle-years storage capacity, β, be variable. Now the linear program becomes

- Objective function: $F(x_1, x_2) = 12x_1 + 7x_2$
- Constraint 1: $3x_1 + 2x_2 \leq \alpha$
- Constraint 2: $2x_1 + x_2 \leq \beta$
- Constraint 3: $x_1 \geq 0$
- Constraint 4: $x_2 \geq 0$

The relative values of α and β determine the geometry of the solution. For example, if $\alpha/2 > \beta$ then constraint 1 becomes irrelevant. When the intersection of constraints 1 and 2 determines the optimal solution it is readily shown that

$$x_1 = -\alpha + 2\beta$$

and

$$x_2 = 2\alpha - 3\beta$$

Hence the optimal solution to the objective function can be expressed as

$$f(x_1, x_2) = 2\alpha + 3\beta$$

Consequently, if α is increased by one unit then $f(x_1, x_2)$ is increased by 2, while if β is increased by one unit then $f(x_1, x_2)$ is increased by 3. So if a winemaker considers expanding his winery he realizes that the cost of increasing grape processing must be less than \$2 and the expense of increasing wine storage must be less than \$3. Otherwise expansion will lose money.

3.3.3 Constraint Coefficient Sensitivity

Now we consider the problem of adjusting one of the coefficients in one of the constraint equations. In particular consider the amount of time γ we age a bottle of red wine to be allowed to vary.

- Objective function: $F(x_1, x_2) = 12x_1 + 7x_2$
- Constraint 1: $3x_1 + 2x_2 \leq 16000$
- Constraint 2: $\gamma x_1 + x_2 \leq 10000$
- Constraint 3: $x_1 \geq 0$
- Constraint 4: $x_2 \geq 0$

To simplify the discussion, let's examine the effect of reducing the amount of time we age the red wine from 2 years to 1.95 years. The solution to the resulting linear program suggests that now 4444 bottles of red can be sold while 1333 bottles of white can be sold for a total profit of \$62,659, increasing the income by almost \$700. Of course, for this to be advisable it must be true that all the bottles of this "younger" red wine can still be sold at the same price, i.e., the taste has not suffered enough to reduce its popularity.

3.4 LINEAR PROGRAMS WITH EQUALITY CONSTRAINTS

In the examples treated so far the constraints defining the feasible sets have been inequalities. However, in practice it is often the case that further constraints in the form of equalities have to be met.

DEFINITION **2**. Let f be a column vector of length n, b a column vector of length m, and b_{eq} a column vector of length k. Let further A and A_{eq} be $m \times n$ and $k \times n$ matrices, respectively. A linear program associated with f, A, b, A_{eq} and b_{eq} is the minimum problem

$$\min_x f^T x \tag{3.1}$$

or the maximum problem

$$\max_x f^T x \tag{3.2}$$

subject to the constraints

$$\begin{array}{rcl} Ax & \leq & b \\ A_{eq}x & = & b_{eq}. \end{array} \tag{3.3}$$

3.4.1 A Task Scheduling Problem

A steel manufacturer produces four different sizes S_i, $1 \leq i \leq 4$ (small, medium, large, and extra large), of beams. These beams can be produced on any one of three machines M_j, $1 \leq j \leq 3$. Machine M_j produces l_{ij} feet of the beams of size S_i per hour. Each machine can be used up to 50 hours per week and the hourly operating cost of machine M_j is $\$c_j$. The manufacturer has to produce k_i feet of beams of size S_i per week. We assume that l_{ij}, c_j and k_i are given numbers.

Clearly the manufacturer wants to minimize the total operating costs. To formulate this minimization problem as a linear program, let x_{ij} be the number of hours per week machine M_j produces the beams of size S_i. The total operating costs are

$$F(x) = \sum_{j=1}^{3}\sum_{i=1}^{4} c_j x_{ij} = \left\{ \begin{array}{r} c_1(x_{11}+x_{21}+x_{31}+x_{41}) \\ +c_2(x_{12}+x_{22}+x_{32}+x_{42}) \\ +c_3(x_{13}+x_{23}+x_{33}+x_{43}) \end{array} \right. \tag{3.4}$$

and this function has to be minimized subject to the following constraints:

- Each machine can operate at most 50 hours per week. Thus the variables x_{ij} have to satisfy the inequalities

$$x_{1j} + x_{2j} + x_{3j} + x_{4j} \leq 50 \quad (1 \leq j \leq 3). \tag{3.5}$$

- Since x_{ij} cannot be negative we have to introduce twelve further inequality constraints

$$-x_{ij} \leq 0 \quad (1 \leq i \leq 4, \;\; 1 \leq j \leq 3). \tag{3.6}$$

- The number of feet of the beams of size S_i produced per week by machine M_j is $l_{ij}x_{ij}$. Thus the total number of feet of this size produced in a week is $\sum_j l_{ij}x_{ij}$, and this must be equal to

$$l_{i1}x_{i1} + l_{i2}x_{i2} + l_{i3}x_{i3} = k_i \quad (1 \leq i \leq 4). \tag{3.7}$$

We now have a linear program with fifteen inequality constraints and four equality constraints.

To match the steel manufacturer problem to Definition 2, we write the twelve variables in a column vector,

$$x = [x_{11}, x_{21}, x_{31}, x_{41}, x_{12}, x_{22}, x_{32}, x_{42}, x_{13}, x_{23}, x_{33}, x_{43}]^T.$$

The inequality constraints (3.5) and (3.6) have to be written in matrix vector form as $Ax \leq b$. Let us denote by A_1 and b_1 the 3×12–matrix and the column vector

of length 3, respectively, such that the inequalities (3.5) take the form $A_1 x \leq b_1$, i.e.

$$A_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}, \quad b_1 = \begin{bmatrix} 50 \\ 50 \\ 50 \end{bmatrix}.$$

The inequalities (3.6) can be written as $A_2 x \leq b_2$, where $A_2 = -I$ with I the 12×12–identity matrix, and b_2 the column vector of length twelve whose entries are all zero. Thus the diagonal entries of A_2 are -1 and the other entries are zero.

The full 15×12–matrix A is then obtained by appending the twelve rows of A_2 below the three rows of A_1 and similarly for b,

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \quad b = \begin{bmatrix} 50 \\ 50 \\ 50 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Likewise, setting

$$A_{eq} = \begin{bmatrix} l_{11} & 0 & 0 & 0 & l_{12} & 0 & 0 & 0 & l_{13} & 0 & 0 & 0 \\ 0 & l_{21} & 0 & 0 & 0 & l_{22} & 0 & 0 & 0 & l_{23} & 0 & 0 \\ 0 & 0 & l_{31} & 0 & 0 & 0 & l_{32} & 0 & 0 & 0 & l_{33} & 0 \\ 0 & 0 & 0 & l_{41} & 0 & 0 & 0 & l_{42} & 0 & 0 & 0 & l_{43} \end{bmatrix}, \; b_{eq} = \begin{bmatrix} k_1 \\ k_1 \\ k_3 \\ k_4 \end{bmatrix},$$

the equality constraints (3.7) can be written in the form $A_{eq} x = b_{eq}$.

3.4.2 Transportation Problems

Transportation problems are typical applications of linear programming. Assume a company has storage depots at m different locations $A_1, \ldots, A_m$ in which k different products $P_1, \ldots, P_k$ are stored. Let M_{ij} be the total amount of product P_j stored in depot A_i. The company has customers $C_1, \ldots, C_r$ in r different cities and has to deliver the amount N_{lj} of product P_j to customer C_l. We assume fixed transportation costs T_{ilj} per unit amount of product P_j if transported to customer C_l from storage deposit A_i.

Let x_{ilj} be the amount of product P_j delivered to customer C_l from deposit

A_i. The problem is to minimize the total transportation costs

$$\sum_{i=1}^{m}\sum_{l=1}^{r}\sum_{j=1}^{k} T_{ilj}x_{ilj} = \min$$

subject to the constraints

$$\begin{aligned} x_{ilj} &\geq 0 \qquad \text{for} \quad 1 \leq l \leq r, \quad 1 \leq j \leq k, \quad 1 \leq i \leq m && (3.8)\\ \sum_{l=1}^{r} x_{ilj} &\leq M_{ij} \qquad \text{for} \quad 1 \leq i \leq m, \quad 1 \leq j \leq k && (3.9)\\ \sum_{i=1}^{m} x_{ilj} &= N_{lj} \qquad \text{for} \quad 1 \leq l \leq r, \quad 1 \leq j \leq k. && (3.10) \end{aligned}$$

This is clearly a linear programming problem with inequality constraints (3.8) and (3.9) and equality constraints (3.10). If m, k, r and the numbers T_{ilj}, M_{ij}, N_{lj} are given, the vectors and matrices f, A, b, A_{eq}, b_{eq} can be constructed similarly as in Subsection 3.4.1.

3.5 A TARGETING PROBLEM

Consider the following problem of launching a rocket to a fixed altitude h in a given time T, while expending a minimum amount of fuel. Let $a(t)$ be the acceleration exerted, $y(t)$ the rocket altitude, and $v(t)$ the rocket velocity at time t. The problem can be formulated as follows.

$$\begin{aligned} &\text{Minimize} && \textstyle\int_0^T |a(t)|dt \\ &\text{Subject to} && \tfrac{dv(t)}{dt} = a(t) - g, \quad \tfrac{dx(t)}{dt} = v(t) \\ & && y(T) = h \\ & && y(t) \geq 0 \quad (0 \leq t \leq T) \\ & && y(0) = 0, \quad v(0) = 0 \\ & && |a(t)| \leq a_0 \quad (0 \leq t \leq T), \end{aligned} \tag{3.11}$$

where a_0 is the maximal acceleration that can applied due to power limitations. Clearly in order that the rocket can leave the ground a_0 must be greater than the earth acceleration g.

Note that the maximum altitude h_{max} to which the rocket can be launched is reached if $a(t) = a_0$ for $0 \leq t \leq T$. If $h > h_{max}$ then (3.11) has no solution. By integrating the equations for $dv(t)/dt$ and $dy(t)/dt$ in (3.11) we find

$$h_{max} = (a_0 - g)T^2/2.$$

3.5.1 Discretization and Solution of the Equations of Motion

Equation (3.11) belongs to the class of *continuous optimization problems* which does not fit a priori into the class of linear programming problems. In order to make the problem amenable to linear programming, we discretize time and assume that

$$a(t) = a_i = const \quad \text{for} \quad t_{i-1} < t < t_i, \tag{3.12}$$

where

$$t_i = i\tau \quad \tau = T/n,$$

and n is a positive integer. The discretized problem is described by n variables $(a_1, \ldots, a_n)$ which have to be determined.

Within each of the n sub-intervals into which the interval $0 \le t \le T$ is divided, the rocket encounters a constant acceleration,

$$\frac{dv(t)}{dt} = a_i - g, \quad \frac{dx(t)}{dt} = v(t) \quad \text{if} \quad t_{i-1} \le t \le t_i. \tag{3.13}$$

After integration these equations lead to the well known linear and quadratic time dependence of velocity and altitude in each sub-interval,

$$v(t) = (a_i - g)(t - t_{i-1}) + v(t_{i-1}) \tag{3.14}$$

$$y(t) = \frac{1}{2}(a_i - g)(t - t_{i-1})^2 + v(t_{i-1})(t - t_{i-1}) + y(t_{i-1}). \tag{3.15}$$

We now set

$$v_i = v(t_i), \quad y_i = y(t_i) \quad (1 \le i \le n),$$

and evaluate the equations (3.14) and (3.14) at $t = t_i$ to obtain

$$\begin{aligned} v_i &= (a_i - g)\tau + v_{i-1} \\ y_i &= \tfrac{1}{2}(a_i - g)\tau^2 + v_{i-1}\tau + y_{i-1}. \end{aligned} \tag{3.16}$$

Equation (3.16) is a linear system of first order difference equations for the (v_i, y_i). The initial values are $(v_0, y_0) = (0, 0)$. Methods for solving difference equations are discussed in Chapter 6, and we will show there that the solution of (3.16) is given by

$$v_i = \tau\Big(\sum_{j=1}^{i} a_j - ig\Big) \tag{3.17}$$

$$y_i = \tau^2\Big(\sum_{j=1}^{i}(\frac{1}{2} + i - j)a_j - \frac{i^2 g}{2}\Big). \tag{3.18}$$

These equations form the solution of the discretized equations of motion for any given set of acceleration values $(a_1, \ldots, a_n)$.

3.5.2 Formulation as Linear Program

Now we formulate the discretized optimization problem as linear programming problem with inequality and equality constraints. The equations of motion

$$\frac{dv(t)}{dt} = a(t) - g, \quad \frac{dx(t)}{dt} = v(t), \quad y(0) = 0, \;\; v(0) = 0 \tag{3.19}$$

have been solved already, so we only need to consider the equality and inequality constraints

$$y(T) = h, \quad |a(t)| \le a_0, \quad y(t) \ge 0 \quad (0 < t < T).$$

From equation (3.18) we infer that the discretized forms of the equality and inequality constraints for $y(t)$ (note that $y(T) = y_n$) can be written as

$$\sum_{j=1}^{n}(\frac{1}{2}+n-j)a_j \quad = \quad \frac{n^2 g}{2}+\frac{h}{\tau^2} \tag{3.20}$$

$$\sum_{j=1}^{i}(\frac{1}{2}+i-j)a_j \quad \geq \quad \frac{i^2 g}{2}, \quad (1 \leq i \leq n-1), \tag{3.21}$$

and the constraint for $a(t)$ becomes

$$|a_i| \leq a_0 \quad (1 \leq i \leq n). \tag{3.22}$$

The objective function which has to be minimized in the discretized problem is

$$\sum_{i=1}^{n}|a_i| = \min, \tag{3.23}$$

and the minimization is subject to the constraints (3.20)– (3.22).

Note that (3.22) and (3.23) involve the absolute values of the variables a_i and hence are not described by linear functions. For inequalities this is not a problem, however there is no way to rewrite the objective function (3.23) as a linear function $\sum_i f_i a_i$. To solve this problem we treat the absolute values as extra variables. Our minimization problem then depends on $2n$ unknown variables which we write again in a column vector

$$x = [x_1, \ldots, x_n, x_{n+1}, \ldots, x_{2n}]^T,$$

where

$$x_i = a_i, \quad x_{i+n} = |a_i| \quad (1 \leq i \leq n).$$

The objective function is now a linear function of x,

$$F(x) = \sum_{i=n+1}^{2n} x_i = \min\,. \tag{3.24}$$

In order that the conditions $x_{i+n} = |x_i|$ are met we have to introduce additional constraints. Since $a_i \leq |a_i|$ and $-a_i \leq |a_i|$ we impose

$$\left.\begin{array}{rcl} x_i & \leq & x_{i+n} \\ -x_i & \leq & x_{i+n} \end{array}\right\} \quad \text{for} \quad 1 \leq i \leq n. \tag{3.25}$$

Clearly the inequalities (3.25) are not equivalent to the condition $x_{i+n} = |x_i|$. However it can be shown that the solution of any linear programming problem is located on the boundary of the feasible set, and for our problem this necessarily implies that for each i one of the two inequalities in (3.25) turns into an equality if x is an optimal solution.

The inequality and equality constraints (3.20)–(3.22) are now rewritten in terms of the x_i as

$$x_{n+i} \leq a_0 \quad \text{for} \quad 1 \leq i \leq n \tag{3.26}$$

$$-\sum_{j=1}^{i}\left(\frac{1}{2}+i-j\right)x_j \leq -\frac{i^2 g}{2} \quad \text{for} \quad 1 \leq i \leq n-1 \tag{3.27}$$

$$\sum_{j=1}^{n}\left(\frac{1}{2}+n-j\right)x_j = \frac{n^2 g}{2}+\frac{h}{\tau^2}. \tag{3.28}$$

The discretized optimization problem is now to minimize the objective function $F(x)$ in (3.24) subject to the inequality constraints (3.25)–(3.27) and the equality constraint (3.28). The objective function as well the inequality and equality constraints are formulated in terms of linear functions of x and so match the abstract problem (3.1) subject to the constraints (3.3) of Definition 2. The matrix A is a $(4n-1)\times 2n$ matrix and A_{eq} is a $1\times 2n$ matrix, i.e. a row vector.

Numerical Solutions. In Figure 3.2 we show the optimal acceleration function $a(t)$ obtained by numerical solution of (3.24)–(3.28) together with $v(t)$ and $y(t)$ for $n = 5$ (Figure 3.2 (a)) and $n = 25$ (Figure 3.2 (b)), and $h = 300$ and $h = 700$. The other paramters are fixed at $g = 32$, $T = 10$, and $a_0 = 48$. The optimal solution has been computed using the *linprog* command of Matlab. The procedure for generating the plots in Figure 3.2 can be summarized as follows.

- Generate the matrices and vectors f, A, b, A_{eq}, and b_{eq} of the linear program according to equations (3.24)–(3.28).
- Apply a numerical solver to find the solution vector x. The first n components of x are the optimal acceleration values $(a_1,\ldots,a_n)$.
- Apply equations (3.17) and (3.18) to compute the velocity and altitude vectors $(v_1,\ldots,v_n)$ and $(y_1,\ldots,y_n)$.
- Use equations (3.12), (3.14) and (3.15) to compute the piecewise constant, piecewise linear and piecewise quadratic functions $a(t)$, $v(t)$ and $y(t)$.
- Plot $a(t)$, $v(t)$, $y(t)$.

As can be seen in Figure 3.2 the optimal acceleration and altitude functions show some distinct features. The acceleration function $a(t)$ starts with a_0 and stays there over a certain number of sub–intervals, then it decreases in the next sub–interval, and after that $a(t)$ is zero. The altitude function $y(t)$ is monotonically increasing and reaches the target altitude from below for larger values of h, whereas for smaller values of h it passes through a maximum and reaches the target altitude from above. In Section 3.6 we will see that the optimal solution to the discretized targeting problem is the best approximation of a known analytical solution to (3.11).

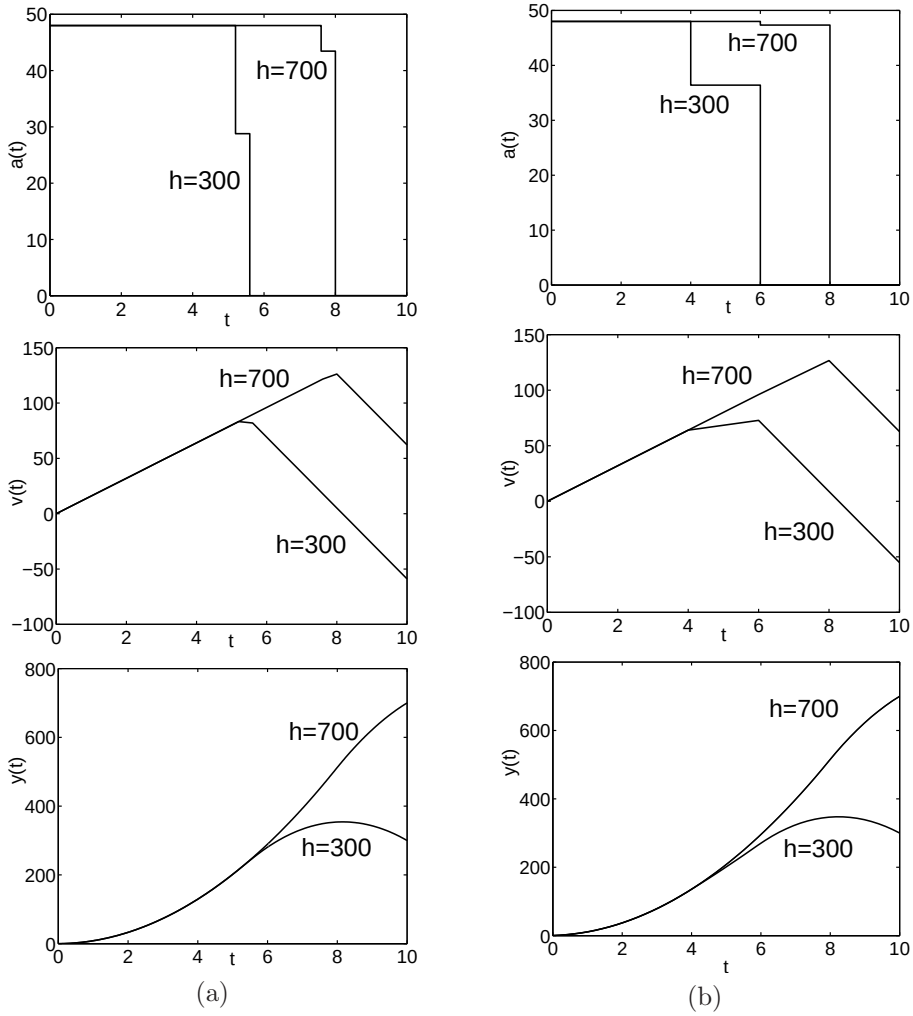


FIGURE 3.2: Graphs of $a(t), v(t), y(t)$ obtained by numerical solution of the linear program (3.24)–(3.28) for $g = 32$, $T = 10$, $a_0 = 48$, and two heights $h = 300$ and $h = 700$. (a): $n = 25$, (b): $n = 5$.

3.5.3 Targeting Problem with Air Resistance

Air resistance is modeled by a friction force $F_d(v)$. Since linear programming requires a linear model we assume $F_d(v)/m = -kv$, where k is a friction coefficient in which the mass is absorbed. The problem (3.11) remains the same except that the equation for $dv(t)/dt$ is now replaced by

$$\frac{dv(t)}{dt} = a(t) - g - kv(t). \tag{3.29}$$

Discretization and Solution of the Equations of Motion. Equation (3.29) is a linear first order differential equation for $v(t)$. In a later chapter we will see that the solution of (3.29) in the interval $t_{i-1} \le t \le t_i$, where $a(t) = a_i = const$, is given by

$$v(t) = \frac{a_i - g}{k} + \big(v(t_{i-1}) - \frac{a_i - g}{k}\big)e^{-k(t-t_{i-1})}. \tag{3.30}$$

The altitude $y(t)$ still satisfies $dy(t)/dt = v(t)$ and so can be found by integration of (3.30),

$$y(t) = y(t_{i-1}) + \frac{a_i - g}{k}\big(t - t_{i-1}\big) + \frac{1}{k}\big(v(t_{i-1}) - \frac{a_i - g}{k}\big)\big(1 - e^{-k(t-t_{i-1})}\big). \tag{3.31}$$

Evaluating (3.30) and (3.31) at $t = t_i$ and letting again $v_i = v(t_i)$, $y_i = y(t_i)$ yields

$$\begin{array}{rcl} v_i &=& pa_i - gp + qv_{i-1} \\ y_i &=& ra_i - gr + pv_{i-1} + y_{i-1}, \end{array} \tag{3.32}$$

where we have set

$$q = e^{-k\tau}, \quad p = (1-q)/k, \quad r = (\tau - p)/k.$$

Equations (3.32) form again a linear system of first order difference equations. This system is more complicated than (3.16), but still can be solved using the methods of Chapter 6. The solution is

$$v_i = p\sum_{j=1}^{i} q^{i-j}a_j - \frac{gp(1-q^i)}{1-q} \tag{3.33}$$

$$y_i = \sum_{j=1}^{i}\Big(r + \frac{p^2(1-q^{i-j})}{1-q}\Big)a_j - \frac{gp^2(i-1-iq+q^i)}{(1-q)^2} - igr. \tag{3.34}$$

Formulation as Linear Program. The formulation of the discretized targeting problem with friction as linear program proceeds in the same way as in Subsection 3.5.2. We introduce the vector x of variables $x_i = a_i$ and $x_{i+n} = |a_i|$ $(1 \le i \le n)$, the objective function (3.24), and the inequality constraints (3.25)–(3.27). The constraints $y_i \ge 0$ for $1 \le i \le n-1$ and $y_n = h$ become

$$-\sum_{j=1}^{i}\Big(r + \frac{p^2(1-q^{i-j})}{1-q}\Big)x_j \le -\frac{gp^2(i-1-iq+q^i)}{(1-q)^2} - igr \tag{3.35}$$

$$\sum_{j=1}^{n}\Big(r + \frac{p^2(1-q^{n-j})}{1-q}\Big)x_j = \frac{gp^2(n-1-nq+q^n)}{(1-q)^2} + ngr + h. \tag{3.36}$$

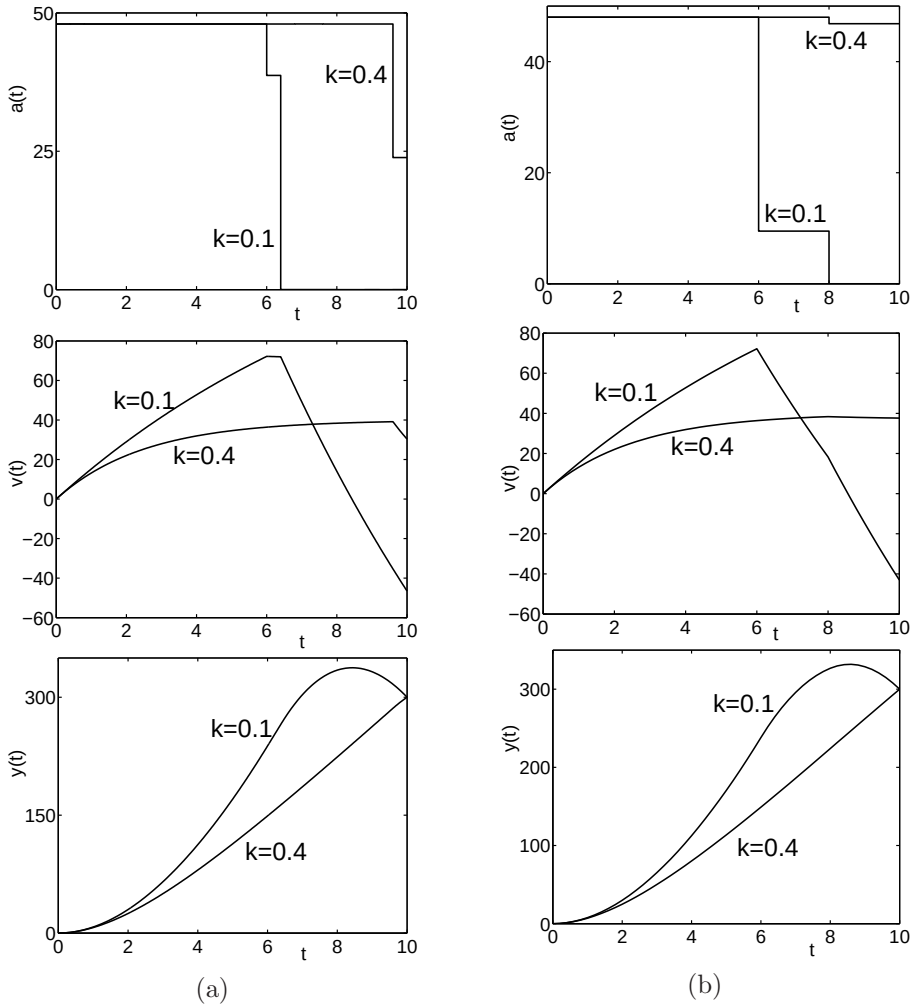


FIGURE 3.3: Graphs of $a(t), v(t), y(t)$ computed from numerical solutions of (3.24)–(3.27), (3.35)–(3.35) for $g = 32$, $T = 10$, $a_0 = 48$, $h = 300$, $k = 0.4$ and $k = 0.1$, and (a): $n = 25$, (b): $n = 5$.

The linear program for the discretized target problem with friction is now to minimize (3.24) subject to the constraints (3.25)–(3.27) and (3.35)–(3.36).

As in the case without friction the maximum possible altitude h_{max} is reached if $a_i = a_0$ for all i. From (3.31) we find

$$h_{max} = \frac{a_0 - g}{k^2}\big(kT - 1 + e^{-kT}\big),$$

and the problem has no solution if $h > h_{max}$.

Numerical Solutions. In Figure 3.3 we show the graphs of $a(t)$, $v(t)$, and $y(t)$ computed from numerical solutions of the linear program (3.24)–(3.27), (3.35)–(3.36) for $g = 32$, $T = 10$, $a_0 = 48$, $h = 300$, $k = 0.4$ and $k = 0.1$, and $n = 25$ and $n = 5$. The solutions are similar to those for the problem without friction, and clearly the greater k the greater is the fuel consumption.

3.5.4 Additional Constraints

There is no problem to impose further conditions on the optimal solution of the targeting problem (with or without friction), provided these conditions can be formulated as linear equality or inequality constraints. We describe two such conditions.

Soft Landing. Soft landing means that the target altitude is reached with velocity $v(T) = 0$. This condition can be build into the linear program by imposing the additional equality constraint

$$v_n = 0,$$

where v_n is represented in terms of the $a_j = x_j$ through equations (3.17) for $k = 0$ or (3.33) for $k > 0$. The vector b_{eq} then becomes a vector of length 2, and accordingly A_{eq} is a $2 \times 2n$–matrix.

Upper Bound for the Velocity. To avoid damage it may be necessary to restrict also the magnitude of the velocity to $|v(t)| \leq v_0$. For the discretized problem this requires that $|v_i| \leq v_0$ for $1 \leq i \leq n$. This inequality is equivalent to the two linear inequalities

$$v_i \leq v_0, \quad -v_i \leq v_0.$$

When the v_i are represented in terms of the x_i, these conditions take the form of $2n$ additional linear inequality constraints imposed on x. The vector b is then extended to a vector of length $6n-1$, and accordingly A is extended to a $(6n-1)\times 2n$–matrix.

Numerical Solutions. In Figure 3.4 the graphs of $a(t)$, $v(t)$, and $y(t)$ computed from the optimal solution of the targeting problem with the condition of soft landing are shown for $g = 32$, $T = 20$, $a_0 = 80$, $h = 400$, $k = 0.4$, and $n = 5$. Figure 3.4 (b) was obtained with the additional inequality constraint $|v(t)| \leq 30$.

We note that the targeting problem with one of the additional constraints considered in this subsection does not admit easily accessible analytical solutions. In contrast, without these additional constraints analytical solutions can be easily found as will be shown in the next section.

3.6 ANALYSIS OF THE TARGETING PROBLEM

In this section we study the targeting problem (3.11) analytically. The numerical solutions shown in Figure 3.2 suggest that the optimal acceleration function $a(t)$ is maximal in a certain initial interval $0 \leq t \leq T_1$ and zero for $T_1 < t \leq T$, where T_1 is adjusted such that the target altitude h is reached in time T. We will see that for this acceleration function the fuel consumption is indeed minimal.

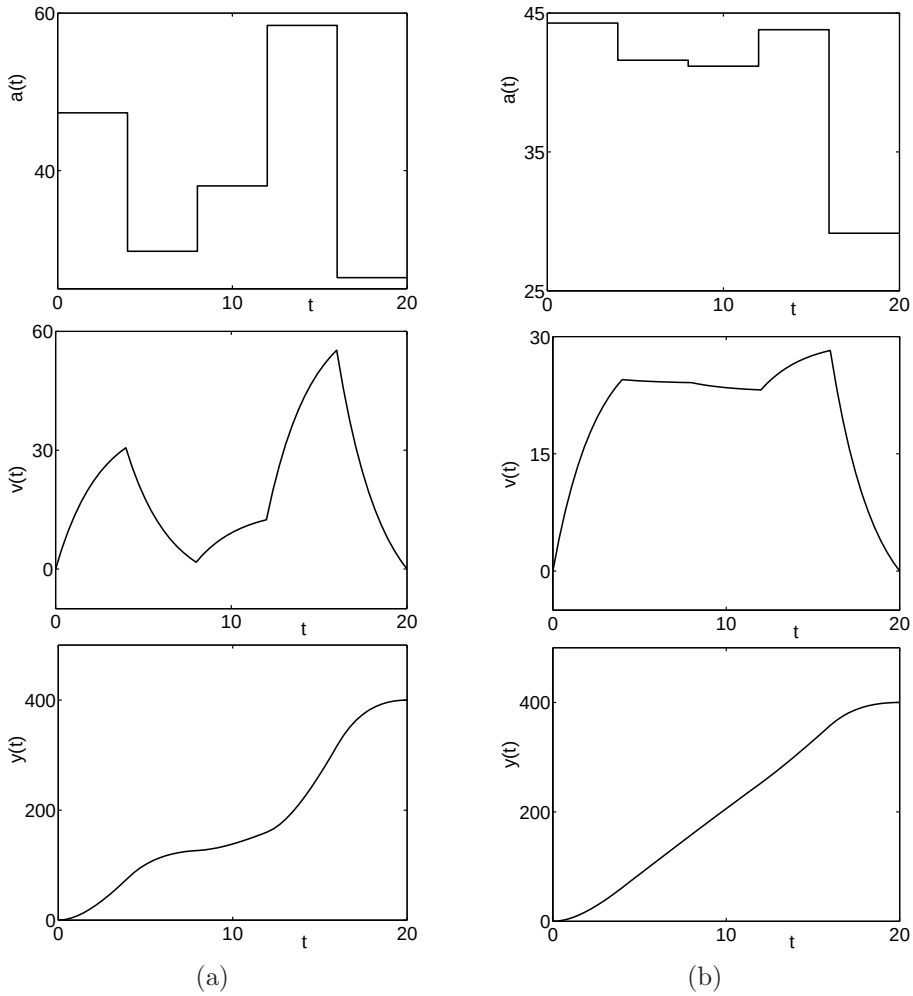


FIGURE 3.4: Graphs $a(t), v(t), y(t)$ computed from the optimal solution of the discretized targeting problem for $g = 32$, $T = 20$, $a_0 = 80$, $h = 400$, $k = 0.4$, and $n = 5$, with additional constraints (a): $v(T) = 0$, (b): $v(T) = 0$ and $|v(t)| \leq 30$.

3.6.1 Analytical Solution

Let $a(t)$ be an acceleration function of the form

$$a(t) = \begin{cases} a_0 & \text{if} \quad 0 \le t \le T_1 \\ 0 & \text{if} \quad t > T_1, \end{cases} \tag{3.37}$$

where $a_0 > g$ and $T_1 > 0$ are given numbers. For this form the solution of the equations of motion (3.19) is given by (Exercise 3.12 (a))

$$v(t) = \begin{cases} (a_0 - g)t & \text{if} \quad 0 \le t \le T_1 \\ a_0 T_1 - gt & \text{if} \quad t \ge T_1, \end{cases} \tag{3.38}$$

$$y(t) = \begin{cases} (a_0 - g)t^2/2 & \text{if} \quad 0 \le t \le T_1 \\ -a_0 T_1^2/2 + a_0 T_1 t - gt^2/2 & \text{if} \quad t \ge T_1. \end{cases} \tag{3.39}$$

Consider then the problem of launching the rocket to a prescribed altitude h in a given time T. The condition $y(T) = h$ leads to the quadratic equation

$$-\frac{1}{2}a_0 T_1^2 + a_0 T_1 T - \frac{1}{2}gT^2 = h$$

for T_1. The solution with $T_1 \le T$ is

$$T_1 = T\big(1 - \sqrt{1 - (g + 2h/T^2)/a_0}\big), \tag{3.40}$$

and in order that the expression under the square root be positive we have to require that

$$h \le (a_0 - g)T^2/2. \tag{3.41}$$

If T_1 and a_0 are related by (3.40), the fuel consumption measured by $C = \int_0^T a(t)dt = a_0 T_1$ is

$$C = a_0 T\big(1 - \sqrt{1 - (g + 2h/T^2)/a_0}\,\big). \tag{3.42}$$

The following theorem (Exercise 3.12 (c)) shows that C is the minimal fuel consumption that can be achieved if $|a(t)|$ is bounded by a_0.

> THEOREM **3**. Let $a(t)$ be an arbitrary piecewise constant acceleration function such that $y(T) = h$ for the solution of (3.19), and assume that $|a(t)| \le a_0$. Then (3.41) is satisfied, and
>
> $$\int_0^T |a(t)|dt \ge C,$$
>
> where C is given by equation (3.42).

Thus the solution of the original (not discretized) targeting problem (3.11) is given by (3.37) with T_1 and a_0 related by (3.40), provided the inequality (3.41) is satisfied. The expression $(a_0 - g)T^2/2$ on the right hand side of this inequality is the maximal altitude to which the rocket can be launched in time T if $|a(t)|$ is bounded by a_0. This altitude is reached if $T_1 = T$, i.e. for the uniform acceleration $a(t) = a_0$ for

$0 \le t \le T$. If $h > (a_0 - g)T^2/2$ then a solution to the targeting problem (3.11) does not exist.

When a numerical solver is applied to the linear program of Subsection 3.5.2, the solver seeks to find the best approximation to the analytical solution (3.37), (3.40). The best approximation is

$$a(t) = \begin{cases} a_0 & \text{if} \quad 0 \le t < m\tau \\ a_1 < a_0 & \text{if} \quad m\tau \le t \le (m+1)\tau \\ 0 & \text{if} \quad t \ge (m+1)\tau, \end{cases}$$

where m is the largest integer for which $m\tau \le T_1(a_0)$. The value of a_1 is adjusted such that the altitude h is reached from the initial data $(v(m\tau), y(m\tau))$ within time $(n-m)\tau$. In the unlikely case that T_1/τ is an integer, the discrete optimal solution coincides with the exact optimal solution.

3.6.2 Dimensionless Variables

Equation (3.40) depends on the physical variables T_1, T, h, g, a_0. We could apply dimensional analysis to reduce the number of variables, but there is a simpler way to identify the relevant dimensionless combinations. If (3.40) is divided by T, the equation can be rewritten as

$$\theta = 1 - \sqrt{1 - 1/\beta}, \tag{3.43}$$

where

$$\theta = \frac{T_1}{T} \le 1, \quad \beta = \frac{a_0}{g + 2h/T^2} \ge 1. \tag{3.44}$$

The variable θ is the ratio of T_1 and T and so $\theta \le 1$. The denominator in β is the uniform acceleration (active for $0 \le t \le T$) through which the rocket is launched to the target altitude h in time T. According to (3.41) $h + gT^2/2 \le a_0$, hence $\beta \ge 1$.

A natural dimensionless variable in terms of which the fuel consumption can be measured is the ratio

$$\gamma = C/C_0, \tag{3.45}$$

where C_0 is the fuel consumption for the uniform acceleration $g + 2h/T^2$,

$$C_0 = (g + 2h/T^2)T = a_0 T/\beta. \tag{3.46}$$

After dividing equation (3.42) by C_0 we obtain

$$\gamma = \beta - \sqrt{\beta^2 - \beta}, \tag{3.47}$$

hence the dimensionless acceleration time θ and fuel consumption γ both depend only on the single dimensionless variable β that measures a_0 in units of $g + 2h/T^2$.

When β increases from $\beta = 1$ towards ∞, γ and θ decrease monotonically from 1 to the limiting values $\gamma_\infty = 1/2$ and $\theta_\infty = 0$, respectively, see Figure 3.5. Consequently the greater β the smaller are θ and γ. In the limit $\beta \to \infty$ and hence $a_0 \to \infty$, the accelerating force becomes an impulsive force that instantaneously, in an infinitesimal time interval, brings the velocity from $v(0-) = 0$ to $v(0+) = v_0$.

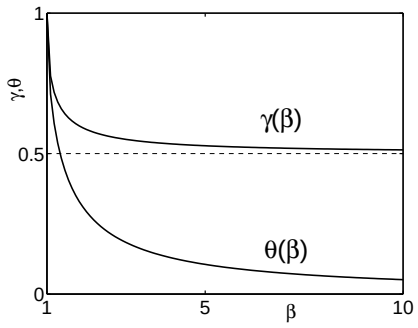


FIGURE 3.5: Graphs of θ and γ versus β, equations (3.43) and (3.47).

Then for $t > 0$ the trajectory of the rocket is $y(t) = v_0 t - gt^2/2$ and v_0 is determined by $y(T) = h$, whence

$$v_0 = \frac{1}{2}(g + 2h/T^2)T = \frac{1}{2}C_0.$$

The limiting value $\lim_{a_0 \to \infty} C(a_0) = C_0/2$ is the minimal fuel consumption if there is no constraint on $|a(t)|$.

3.6.3 Maximum Altitude

Now we address the question when the rocket reaches the target height from above or from below. The altitude function $y(t)$ has a maximum $h_m = y(T_m)$ at time $t = T_m$ determined by $v(t) = 0$,

$$T_m = \frac{a_0}{g}T_1, \quad h_m = \frac{a_0}{2}T_1^2\big(\frac{a_0}{g} - 1\big). \tag{3.48}$$

Since now h and a_0 have to be treated independently of each other, we introduce the dimensionless variables

$$\alpha = \frac{a_0}{g}, \quad \xi = \frac{2h}{gT^2}, \quad \theta_m = \frac{T_m}{T}, \tag{3.49}$$

and note that $\beta = \alpha/(1 + \xi)$. The condition that the maximum of $y(t)$ is attained in the range $0 \le t \le T$ is $\theta_m \le 1$. From (3.40) and (3.48) we find that

$$\theta_m = \alpha - \sqrt{\alpha^2 - (1 + \xi)\alpha}, \tag{3.50}$$

and this is less than one if

$$\alpha \ge \frac{1}{1 - \xi}. \tag{3.51}$$

Moreover, the condition for a solution to exist at all is $\beta \ge 1$. In terms of α and ξ this condition becomes

$$\alpha \ge 1 + \xi. \tag{3.52}$$

The boundary lines $\alpha = 1/(1 - \xi)$ and $\alpha = 1 + \xi$ separate the (α, ξ)–plane into three regions I, II, and III as shown in Figure 3.6. In regions I and II the rocket

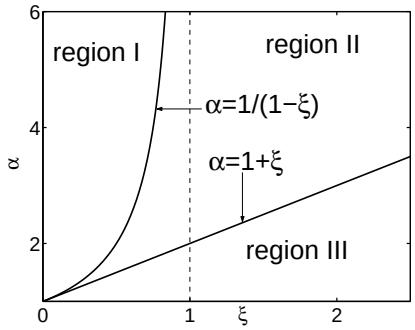


FIGURE 3.6: Regions I, II, III in the (ξ, α)–plane.

reaches h from above and below, respectively. In region III the targeting problem has no solution. We summarize this in terms of the physical variables a_0, h, T:

- If $h < gT^2(1-g/a_0)/2$ then the rocket reaches the target altitude from above.
- If $gT^2(1-g/a_0)/2 < h \leq (a_0-g)T^2/2$ then the rocket reaches the target altitude from below.
- If $h > (a_0-g)T^2/2$ then the targeting problem (3.11) has no solution.

PROBLEMS

3.1. Think of an optimization problem that can be written as a linear program with two decision variables. Specify the objective function as well as the constraints. Avoid constructing a problem the solution to which is such that one of the decision variables is zero. Can you extend your problem to more decision variables and constraints?

3.2. Solve graphically the question of how many Zebra Danios and Gouramis should be purchased for the fish tank modeled in section 3.1.2.

3.3. A new burger chain, the EcoliExpress, has has two new products: the large 1/3 pound "Big Whoopie" burger and the smaller 1/4 pound "Wimpy Whoopie" burger. It has been determined in test market trials that the Big Whoopie can be sold at a profit of 45 cents per burger and the Wimpy Whoopie at a profit of 25 cents. Furthermore, a chain knows that it can sell all its burgers if it uses 100 pounds of meat per week. In addition, the preparation time for a Big Whoopie is two minutes and for a Wimpy Whoopie is one minute and the chain has one employee working 40 hours per week preparing both types of Whoopies. Assuming the owner of the EcoliExpress wishes to maximize profits formulate a solution using linear programming. Using this model, answer the following:

(a) How many Big Whoopies and Wimpy Whoopies should be sold?

(b) Assuming the unit profit of the Big Whoopie is fixed at 45 cents, for what range of prices of the Wimpy Whoopie is the solution in a) optimal?

(c) Assuming the unit profit of the Wimpy Whoopie is fixed at 25 cents, how large does the unit profit for the Big Whoopie have to be to justify making only this type of burger?

(d) What should the cost of meat be (per pound) to justify purchasing additional quantities? Hint: the profit must increase.

In Exercises 3.4–3.8 first formulate the problem as linear program. Then use a linear program solver such as the *linprog* function of Matlab to find the optimal solution.

3.4. An agricultural mill manufactures feed for cattle, sheep and chickens. This is done by mixing the following ingredients: corn, limestone, soybeans, and fish meal. These ingredients contain the following nutrients: vitamins, protein, calcium, and crude fat. The contents of the nutrients in each kilogram of the ingredients is summarized in Table 3.4. The mill contracted to produce 10, 8, and 8

Ingredient	Vitamins	Protein	Calcium	Crude Fat
Corn	8	10	6	8
Limestone	6	5	10	6
Soybeans	10	12	6	6
Fish Meal	4	8	6	9

TABLE 3.2:

(metric) tons of cattle feed, sheep feed, and chicken feed. Because of shortages, a limited amount of the ingredients is available, namely 6 tons of corn, 10 tons of limestone, 4 tons of soybeans, and 5 tons of fish meal. The price per kilogram of these ingreients is \$0.20, \$0.12, \$0.24, and \$0.12. The minimal and maximal units of the various nutrients that are permitted is summarized in Table 3.4 for a kilogram of the cattle feed, the sheep feed, and the chicken feed. Formulate this mixed–feed problem as a linear program so that the total costs are minimized.

3.5. A tractor factory has supply depots in three cities C_1, C_2, C_3. Two traders T_1 and T_2 order 22 and 28 tractors of a certain special kind, respectively. The

Product	Vitamins		Protein		Calcium		Crude Fat	
	Min	Max	Min	Max	Min	Max	Min	Max
Cattle Feed	6	∞	6	∞	7	∞	4	8
Sheep Feed	6	∞	6	∞	6	∞	4	6
Chicken Feed	4	6	6	∞	6	∞	4	6

TABLE 3.3:

transportation costs per tractor (in dollars) from each of the three depots to the locations of the traders and the total number N of available tractors in each depot are summarized in Table 3.4. How many tractors should be delivered from each of the three cities to each of the two traders in order that the total transportation costs are minimized?

	C_1	C_2	C_3
T_1	250	80	400
T_2	300	100	200
N	15	25	25

TABLE 3.4:

3.6. Solve the scheduling problem of Subsection 3.4.1 for the following data

$$(l_{ij}) = \begin{bmatrix} 300 & 600 & 880 \\ 250 & 400 & 700 \\ 200 & 350 & 600 \\ 100 & 200 & 300 \end{bmatrix}, \quad (c_j) = \begin{bmatrix} 30 \\ 50 \\ 80 \end{bmatrix}, \quad (k_i) = \begin{bmatrix} 10000 \\ 8000 \\ 6000 \\ 6000 \end{bmatrix}.$$

3.7. A confectioner manufactures two kinds of candy bars: "ProteinPlus", that has no carbohydrates, and "SugarPlus", with no fat. ProteinPlus sells for a profit of 40 cents per bar, and SugarPlus sells for a profit of 50 cents per bar. The candy is proccessed in three main operations: blending, cooking and packaging. The following table records the average time in minutes required by each bar for each of the processing operations:

	Blending	Cooking	Packaging
ProteinPlus	1	5	3
SugarPlus	2	4	1

During each production run the blending equipment is available for a maximum of 12 machine hours, the cooking equipment is available for at most 30 machine hours, and the packaging equipment for no more than 15 hours. If this machine time can be allocated to the making of either candy type at all times that is available, the confectioner wants to know how many boxes of each type should be produced in order to realize the maximum profit.

Formulate this problem as a linear program. Sketch the feasible region and the optimal isoprofit line, and find the optimal solution.

3.8. Paul has 2200 per year to invest over the next five years. At the beginning of each year he can invest in one–, two–, and three–year deposits at interest rates of 8%, 17% (total) and 27% (total), respectively. If Paul reinvests his money available each year, how much should he invest in each of the three deposits each year so that his total cash at the end of the five years is a maximum?

The following exercises deal with the targeting problem of Sections 3.5 and 3.6.

3.9. Without using software, solve the optimization problem

$$a_1 + a_2 + a_3 = \min$$

subject to the inequality constraints

$$\begin{aligned} 32 \le a_1 \le a_0 \\ 0 \le a_2 \le a_0 \\ 0 \le a_3 \le a_0 \\ 3a_1 + a_2 \ge 128, \end{aligned}$$

and the equality constraint

$$5a_1 + 3a_2 + a_3 = 336,$$

for

(a) $a_0 = 40$,
(b) $a_0 = 64$,
(c) $a_0 = 96$.

Hint: Solve the equality constraint for a_3 and substitute this into the objective function and the inequality constraints to find a problem with only two variables a_1, a_2. Solve this two–variable problem graphically.

3.10. Consider the linear program (3.24)–(3.28) with the additional constraints $v_n = 0$ and $|v_i| \le v_0$ for $1 \le i \le n$ (see Subsection 3.5.4).

(a) Identify the vectors and matrices f, A, b, A_{eq}, b_{eq}. For example write $f_i = p_1$ for $1 \le i \le n$, $f_i = p_2$ for $n+1 \le i \le 2n$, with p_1, p_2 to be determined.

(b) Write a Matlab function that receives g, a_0, T, h, v_0 as input and generates the matrices in (a) as output.

3.11. Let $g = 32$, $T = 20$, $a_0 = 80$, $h = 100$, and $n = 25$. Use a linear program solver to find the optimal acceleration values $(a_1, \ldots, a_n)$ for the discretized targeting problem with friction constant k and the given additional constraints. If the solver fails to find a solution explain why. If it finds a solution plot the acceleration function $a(t)$, the velocity $v(t)$, and the altitude $y(t)$. Comment on these plots.

(a) $k = 0$, no additional constraint.
(b) $k = 2$, no additional constraint.
(c) $k = 0.4$, no additional constraint.
(d) $k = 0.4$, additional constraint $|v(t)| \le 30$ for $0 \le t \le T$.
(e) $k = 0.4$, additional constraint $v(T) = 0$.
(f) $k = 0.4$, additional constraints $v(T) = 0$ and $|v(t)| \le 30$ for $0 \le t \le T$.

3.12. In this exercise you work out some of the details of the analysis of Section 3.6.

(a) Verify equations (3.38) and (3.39).
(b) Verify equation (3.48).
(c) Prove Theorem 3 by induction on n.

CHAPTER 4

Modeling with Nonlinear Programming

By nonlinear programming we intend the solution of the general class of problems that can be formulated as

$$\min f(x)$$

subject to the inequality constraints

$$g_i(x) \leq 0$$

for $i = 1, \ldots, p$ and the equality constraints

$$h_i(x) = 0$$

for $i = 1, \ldots, q$. We consider here methods that search for the solution using gradient information, i.e., we assume that the function f is differentiable.

EXAMPLE 4.1

Given a fixed area of cardboard A construct a box of maximum volume. The nonlinear program for this is

$$\min xyz$$

subject to

$$2xy + 2xz + 2yz = A$$

EXAMPLE 4.2

Consider the problem of determining locations for two new high schools in a set of P subdivisions N_j. Let w_{1j} be the number of students going to school A and w_{2j} be the number of students going to school B from subdivision N_j. Assume that the student capacity of school A is c_1 and the capacity of school B is c_2 and that the total number of students in each subdivision is r_j. We would like to minimize the total distance traveled by all the students given that they may attend either school A or B. It is possible to construct a nonlinear program to determine the locations (a, b) and (c, d) of high schools A and B, respectively assuming the location of each subdivision N_i is modeled as a single point denoted (x_i, y_i).

$$\min \sum_{j=1}^{P} w_{1j} \left((a - x_j)^2 + (b - y_j)^2 \right)^{\frac{1}{2}} + w_{2j} \left((c - x_j)^2 + (d - y_j)^2 \right)^{\frac{1}{2}}$$

subject to the constraints

$$\sum_j w_{ij} \leq c_i$$

$$w_{1j} + w_{2j} = r_j$$

for $j = 1, \ldots, P$.

EXAMPLE 4.3

Neural networks have provided a new tool for approximating functions where the functional form is unknown. The approximation takes on the form

$$f(x) = \sum_j b_j \sigma(a_j x - \alpha_j) - \beta$$

and the corresponding sum of squares error term is

$$E(a_j, b_j, \alpha_j, \beta) = \sum_i \big(y_i - f(x_i)\big)^2$$

The problem of minimizing the error function is, in this instance, an unconstrained optimization problem. An efficient means for computing the gradient of E is known as the *backpropogation* algorithm.

4.1 UNCONSTRAINED OPTIMIZATION IN ONE DIMENSION

Here we begin by considering a significantly simplified (but nonetheless important) nonlinear programming problem, i.e., the domain and range of the function to be minimized are one-dimensional and there are no constraints. A necessary condition for a minimum of a function was developed in calculus and is simply

$$f'(x) = 0$$

Note that higher derivative tests can determine whether the function is a max or a min, or the value $f(x + \delta)$ may be compared to $f(x)$.

Note that if we let

$$g(x) = f'(x)$$

then we may convert the problem of finding a minimum or maximum of a function to the problem of finding a zero.

4.1.1 Bisection Algorithm

Let x^* be a root, or zero, of $g(x)$, i.e., $g(x^*) = 0$. If an initial bracket $[a, b]$ is known such that $x^* \in [a, b]$, then a simple and robust approach to determining the root is to bisect this interval into two intervals $[a, c]$ and $[c, b]$ where c is the midpoint, i.e.,

$$c = \frac{a + b}{2}$$

If

$$g(a)g(c) < 0$$

then we conclude

$$x^* \in [a, c]$$

while if

$$g(b)g(c) < 0$$

then we conclude

$$x^* \in [b, c]$$

This process may now be iterated such that the size of the bracket (as well as the actual error of the estimate) is being divided by 2 every iteration.

4.1.2 Newton's Method

Note that in the bisection method the actual value of the function $g(x)$ was only being used to determine the correct bracket for the root. Root finding is accelerated considerably by using this function information more effectively.

For example, imagine we were seeking the root of a function that was a straight line, i.e., $g(x) = ax + b$ and our initial guess for the root was x_0. If we extend this straight line from the point x_0 it is easy to determine where it crosses the axis, i.e.,

$$ax_1 + b = 0$$

so $x_1 = -b/a$. Of course, if the function were truly linear then no first guess would be required. So now consider the case that $g(x)$ is nonlinear but may be approximated locally about the point x_0 by a line. Then the point of intersection of this line with the x-axis is an estimate, or second guess, for the root x^*. The linear approximation comes from Taylor's theorem, i.e.,

$$g(x) = g(x_0) + g'(x_0)(x - x_0) + \frac{1}{2}g''(x_0)(x - x_0)^2 + \dots$$

So the linear approximation to $g(x)$ about the point x_0 can be written

$$l(x) = g(x_0) + g'(x_0)(x - x_0)$$

If we take x_1 to be the root of the linear approximation we have

$$l(x_1) = 0 = g(x_0) + g'(x_0)(x_1 - x_0)$$

Solving for x_1 gives

$$x_1 = x_0 - \frac{g(x_0)}{g'(x_0)}$$

or at the nth iteration

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}$$

The iteration above is for determining a zero of a function $g(x)$. To determine a maximum or minimum value of a function f we employ condition that $f'(x) = 0$. Now the iteration is modified as as

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

4.2 UNCONSTRAINED OPTIMIZATION IN HIGHER DIMENSIONS

Now we consider the problem of minimizing (or maximizing) a scalar function of many variables, i.e., defined on a vector field. We consider the extension of Newton's method presented in the previous section as well as a classical approach known as steepest descent.

4.2.1 Taylor Series in Higher Dimensions

Before we extend the search for extrema to higher dimensions we present Taylor series for functions of more than one domain variable. To begin, the Taylor series for a function of two variables is given by

$$\begin{aligned} g(x,y) =& g(x^{(0)},y^{(0)}) + \frac{\partial g}{\partial x}(x-x^{(0)}) + \frac{\partial g}{\partial y}(y-y^{(0)}) \\ &+ \frac{\partial^2 g}{\partial x^2}\frac{(x-x^{(0)})^2}{2} + \frac{\partial^2 g}{\partial y^2}\frac{(y-y^{(0)})^2}{2} + \frac{\partial^2 g}{\partial x \partial y}(x-x^{(0)})(y-y^{(0)}) \\ &+ \text{higher order terms} \end{aligned}$$

In n variables $x = (x_1, \ldots, x_n)^T$ the Taylor series expansion becomes

$$g(x) = g(x^{(0)}) + \nabla g(x^{(0)})(x-x^{(0)}) + \frac{1}{2}(x-x^{(0)})^T Hg(x^{(0)})(x-x^{(0)}) + \cdots$$

where the Hessian matrix is defined as

$$\big(Hg(x)\big)_{ij} = \frac{\partial^2 g(x)}{\partial x_i \partial x_j}$$

and the gradient is written as a row vector, i.e.,

$$\big(\nabla g(x)\big)_i = \frac{\partial g(x)}{\partial x_i}$$

4.2.2 Roots of a Nonlinear System

We saw that Newton's method could be used to develop an iteration for determining the zeros of a scalar function. We can extend those ideas for determining roots of the nonlinear system

$$\begin{aligned} g_1(x_1,\ldots,x_n) &= 0 \\ g_2(x_1,\ldots,x_n) &= 0 \\ &\vdots \\ g_n(x_1,\ldots,x_n) &= 0 \end{aligned}$$

or, more compactly,

$$g(x) = 0.$$

Now we apply Taylor's theorem to each component $g_i(x_1, \ldots, x_n)$ individually, i.e., retaining only the first order terms we have the linear approximation to g_i about the point $x^{(0)}$ as

$$l_i(x) = g_i(x^{(0)}) + \nabla g_i(x^{(0)})(x-x^{(0)})$$

for $i = 1, \ldots, n$. We can write these components together as a vector equation

$$l(x) = g(x^{(0)}) + Jg(x^{(0)})(x - x^{(0)})$$

where now

$$\big(Jg(x)\big)\big)_{ij} = \frac{\partial g_i(x)}{\partial x_j}$$

is the $n \times n$–matrix whose rows are the gradients of the components g_i of g. This matrix is called the Jacobian of g.

As in the scalar case we base our iteration on the assumption that

$$l(x^{(k+1)}) = 0$$

Hence,

$$g(x^{(k)}) + Jg(x^{(k)})(x^{(k+1)} - x^{(k)}) = 0$$

and given $x^{(k)}$ we may determine the next iterate $x^{(k+1)}$ by solving an $n \times n$ system of equations.

4.2.3 Newton's Method

In this chapter we are interested in minimizing functions of several variables. Analogously with the scalar variable case we may modify the above root finding method to determine maxima (or minima) of a function $f(x_1, \ldots, x_n)$. To compute an extreme point we require that $\nabla f = 0$, hence we set

$$g(x) = \big(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\big)^T.$$

Substituting

$$g_i(x) = \frac{\partial f(x)}{\partial x_i}$$

into

$$g(x^{(k)}) + Jg(x^{(k)})(x^{(k+1)} - x^{(k)}) = 0$$

produces

$$\nabla f(x^{(k)}) + Hf(x^{(k)})(x^{(k+1)} - x^{(k)}) = 0$$

where

$$\big(Hf(x)\big)_{ij} = \big(Jg(x)\big)_{ij} = \frac{\partial g_i(x)}{\partial x_j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

Again we have a linear system for $x^{(k+1)}$.

4.2.4 Steepest Descent

Another form for Taylor's formula in n-variables is given by

$$f(x + th) = f(x) + t\nabla f(x)h \; + \; \text{higher order terms}$$

where again $(\nabla f(x))_i = \partial f(x)/\partial x_i$. Now t is a scalar and $x+th$ is a ray emanating from the point x in the direction h. We can compute the derivative of the function $f(x+th)$ w.r.t. t as

$$\frac{df}{dt}(x+th) = \nabla f(x+th)h.$$

Evaluating the derivative at the point $t=0$ gives

$$\frac{df}{dt}(x+th)|_{t=0} = \nabla f(x)h$$

This quantity, known as the directional derivative of f, indicates how the function is changing at the point x in the direction h. Recall from calculus that the direction of maximum increase (decrease) of a function is in the direction of the gradient (negative gradient). This is readily seen from the formula for the directional derivative using the identity

$$\nabla f(x)h = \|\nabla f(x)\|\|h\|\cos(\theta)$$

where θ is the angle between the vectors $\nabla f(x)$ and h. Here $\|a\|$ denotes the Euclidean norm of a vector a. We can assume without loss of generality that h is of unit length, i.e., $\|h\| = 1$. So the quantity on the right is a maximum when the vectors h and $\nabla f(x)$ point in the same direction so $\theta = 0$.

This observation may be used to develop an algorithm for unconstrained function minimization. With an appropriate choice of the scalar step-size α, the iterations

$$x^{(k+1)} = x^{(k)} - \alpha\nabla f(x^{(k)}) \tag{4.1}$$

will converge (possibly slowly) to a minimum of the function $f(x)$.

4.3 CONSTRAINED OPTIMIZATION AND LAGRANGE MULTIPLIERS

Consider the constrained minimization problem

$$\min f(x)$$

subject to

$$c_i(x) = 0$$

$i = i, \ldots, p$. It can be shown that a necessary condition for a solution to this problem is provided by solving

$$\nabla f = \lambda_1\nabla c_1 + \cdots + \lambda_p\nabla c_p$$

where the λ_i are referred to as Lagrange multipliers. Consider the case of f, c being functions of two variables and consider their level curves. In Section 4.4 we will demonstrate that an extreme value of f on a single constraint c is given when the gradients of f and c are parallel. The equation above generalizes this to several constraints c_i: an extreme value is given if the gradient of f is a linear combination of the gradients of the c_i.

We demonstrate a solution via this procedure by recalling our earlier example.

EXAMPLE 4.4

Given a fixed area of cardboard A construct a box of maximum volume. The nonlinear program for this is

$$\min xyz$$

subject to

$$2xy + 2xz + 2yz = A$$

Now $f(x, y, z) = xyz$ and $c(x, y, z) = 2xy + 2yz + 2xz - A$. Substituting these functions into our condition gives

$$\nabla f = \lambda \nabla c$$

which produces the system of equations

$$\begin{aligned} yz - \lambda(2y + 2z) &= 0 \\ xz - \lambda(2x + 2z) &= 0 \\ xy - \lambda(2y + 2x) &= 0 \end{aligned}$$

These equations together with the constraints provide four equations for (x, y, z, λ). If we divide the first equation by the second we find $x = y$. Similarly, if the second equation is divided by the third we obtain $y = z$. From the constraint it follows then that $6x^2 = A$, hence the solution is

$$x = y = z = \sqrt{\frac{A}{6}}.$$

In this special case the nonlinear system could be solved by hand. Typically this is not the case and one must resort to numerical techniques such as Newton's method to solve the resulting $(n + m) \times (n + m)$ system

$$g(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m) = 0.$$

4.4 GEOMETRY OF CONSTRAINED OPTIMIZATION

4.4.1 One Equality Constraint

Consider a two variable optimization problem

$$\min f(x, y)$$

subject to

$$c(x, y) = 0.$$

Geometrically the constraint $c = 0$ defines a curve C in the (x, y)–plane, and the function $f(x, y)$ is restricted to that curve. If we could solve the constraint equation for y as $y = h(x)$, the problem would reduce to an unconstrained, single variable optimization problem

$$\min f(x, h(x)).$$

From calculus we know that a necessary condition for a minimum is

$$\frac{d}{dx}f(x,h(x)) = \frac{\partial f}{\partial x}(x,h(x)) + \frac{\partial f}{\partial y}(x,h(x))h'(x) = 0. \tag{4.2}$$

Since $c(x,h(x)) = 0$, we also have

$$\frac{d}{dx}c(x,h(x)) = \frac{\partial f}{\partial x}(x,h(x)) + \frac{\partial c}{\partial y}(x,h(x))h'(x) = 0. \tag{4.3}$$

A necessary condition for equations (4.2) and (4.3) to hold simultaneously is

$$\frac{\partial f}{\partial x}\frac{\partial c}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial c}{\partial x} = 0. \tag{4.4}$$

From elementary linear algebra we know that if an equation $ad - bc = 0$ holds then the vectors (a,b) and (c,d) are linearly dependent, i.e. collinear, and so one of them is a multiple of the other. Thus there exists a constant λ such that

$$\nabla f = \lambda \nabla c. \tag{4.5}$$

Now let's look more closely at the curve C. The tangent of the curve $y = h(x)$ at a point $(x_0, y_0) = (x_0, h(x_0))$ is given by

$$y = (x - x_0)h'(x_0) + y_0.$$

We set $x - x_0 = t$ and write this equation in vector form as

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + t \begin{bmatrix} 1 \\ h'(x_0) \end{bmatrix}.$$

The vector $T = [1, h'(x_0)]^T$ points into the direction of the tangent line and is called a tangent vector of C at (x_0, y_0). Equation (4.3) tells that T is orthogonal to $\nabla c(x_0, y_0)$. Thus at every point on C the gradient ∇c is orthogonal to the tangent of C.

For level contours $f(x,y) = f_0$ at level f_0 (an arbitrary constant) the situation is analogous, i.e., at each point on the contour the gradient ∇f is orthogonal to the tangent. Moreover, it is shown in multivariable calculus that ∇f points into the region in which f is increasing as illustrated in Figure 4.1. Note that the vector $(\partial f/\partial y, -\partial f/\partial x)$ is orthogonal to ∇f and so is a tangent vector.

At a point (x_0, y_0) on C for which (4.5) holds, the level contour of $f_0 = f(x_0, y_0)$ intersects the curve C. Since the gradients of f and c are collinear at this point, the tangents of the contour $f = f_0$ and the curve $c = 0$ coincide, hence the two curves meet tangentially. Thus the condition (4.5) means geometrically that we search for points at which a level contour and the constraint curve C have a tangential contact.

EXAMPLE 4.5

Consider the problem of finding all maxima and minima of

$$f(x,y) = x^2 - y^2$$

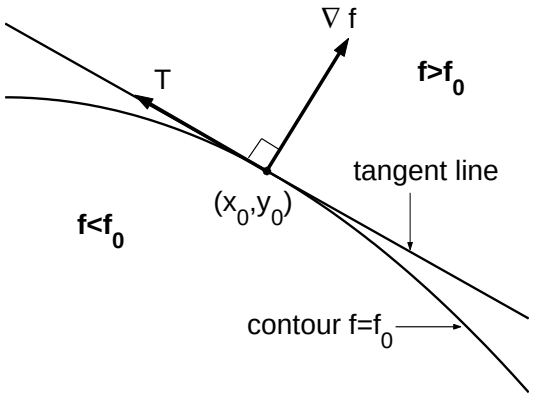


FIGURE 4.1: The gradient of f is orthogonal to the tangent of a level contour and points into the region of increasing f.

subject to

$$x^2 + y^2 = 1. \tag{4.6}$$

The equation (4.5) becomes

$$2x = 2\lambda x \tag{4.7}$$

$$2y = -2\lambda y, \tag{4.8}$$

and (4.6)–(4.8) are three equations for (x, y, λ). Equation (4.7) has the solution $x = 0$ and the solution $\lambda = 1$ if $x \neq 0$. If $x = 0$, (4.6) leads to $y = \pm 1$ giving the solution points $(0, \pm 1)$ with values $f(0, \pm 1) = -1$. If $x \neq 0$ and $\lambda = 1$, (4.8) implies $y = 0$ and so $x = \pm 1$ from (4.6). This leads to the solution points $(\pm 1, 0)$ with values $f(\pm 1, 0) = 1$. Hence the points $(0, \pm 1)$ yield minima and $(\pm 1, 0)$ yield maxima.

In Figure 4.2 (a) some level contours of f and the constraint circle (4.6) are shown. The contours $f = 1$ and $f = -1$ are the only contours that meet this circle tangentially. The points of tangency are the maximum and minimum points of f restricted to the unit circle.

A slightly more complicated objective function is

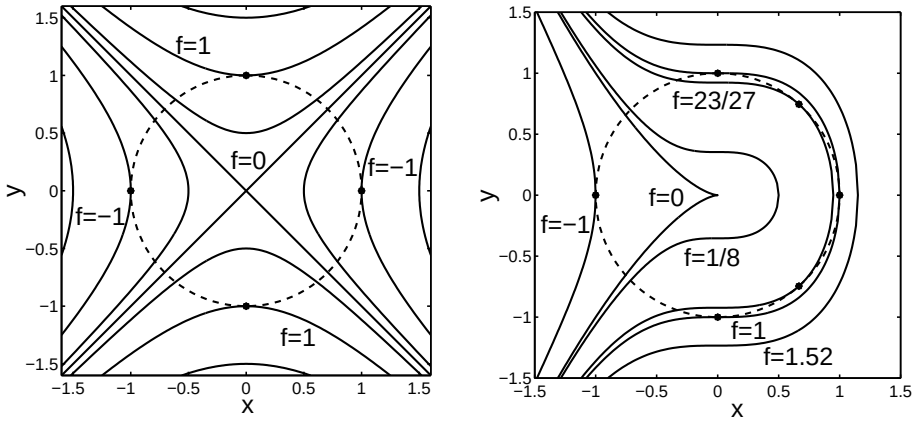
$$f(x, y) = x^3 + y^2.$$

Again we seek all maxima and minima of f subject to the constraint (4.6). The equation (4.5) now results in

$$3x^2 = 2\lambda x \tag{4.9}$$

$$2y = 2\lambda y. \tag{4.10}$$

Equation (4.9) has the solution $x = 0$ and $\lambda = 3x/2$ if $x \neq 0$. If $x = 0$ we find $y = \pm 1$ from (4.6) giving the solutions $(0, \pm 1)$ with values $f(0, \pm 1) = 1$. If $\lambda = 3x/2 \neq 0$, equation (4.10) has the solutions $y = 0$ and $\lambda = 1$ if $y \neq 0$. Now if $y = 0$ we find



(a) (b)

FIGURE 4.2: Unit circle $x^2 + y^2 = 1$ (dashed) and level contours of (a): $f(x, y) = x^2 - y^2$, (b): $f(x, y) = x^3 + y^2$. The points of tangency are the extreme points of $f(x, y)$ restricted to the unit circle.

$x = \pm 1$ from (4.6) giving the solutions $(\pm 1, 0)$ with values $f(\pm 1, 0) = \pm 1$. If $y \neq 0$ it follows that $\lambda = 1$, hence $x = 2/3$, and so $y = \pm\sqrt{5}/3$ from (4.6). The f-values of the solution points $(2/3, \pm\sqrt{5}/3)$ are both $23/27 < 1$. Thus there is a single global minimum $f = -1$ at $(-1, 0)$, and three global maxima $f = 1$ at $(0, \pm 1)$ and $(1, 0)$.

Some level contours of f and the constraint curve (4.6) are shown in Figure 4.2 (b). Note that the zero contour forms a cusp, $y = \pm(-x)^{3/2}$, $x \leq 0$. The points of tangency of a level contour and the constraint curve are again identified with extreme points. Since the points $(2/3, \pm\sqrt{5}/3)$ are located between the global maximum points they must correspond to local minima.

In three dimensions the equation $\nabla f = \lambda \nabla c$, resulting from an optimization problem with a single constraint, implies that at a solution point a level surface $f(x, y, z) = f_0$ is tangent to the constraint surface $c(x, y, z) = 0$.

EXAMPLE 4.6

Find the maxima and minima of

$$f(x, y, z) = 5x + y^2 + z$$

subject to

$$x^2 + y^2 + z^2 = 1. \tag{4.11}$$

The equation $\nabla f = \lambda \nabla c$ now leads to

$$5 = 2\lambda x \tag{4.12}$$

$$2y = 2\lambda y \tag{4.13}$$

$$1 = 2\lambda z. \tag{4.14}$$

From (4.12) and (4.14) we infer that $x = 5z$, and (4.13) has the solutions $y = 0$ and $\lambda = 1$ if $y \neq 0$. Assume first $y = 0$. The constraint (4.11) implies $x^2+z^2 = 26z^2 = 1$, hence $z = \pm 1/\sqrt{26}$, $x = \pm 5/\sqrt{26}$, and $f(\pm 5/\sqrt{26}, 0, \pm 1/\sqrt{26}) = \pm\sqrt{26}$.

Now assume $y \neq 0$, hence $\lambda = 1$, and so $x = 5/2$, $z = 1/2$. The constraint (4.11) then yields $26/4 + y^2 = 1$ which has no solution. Thus there is a unique maximum at $(5/\sqrt{26}, 0, 1/\sqrt{26})$ and a unique minimum at $(-5/\sqrt{26}, 0, -1/\sqrt{26})$.

EXAMPLE 4.7

Find the maxima and minima of

$$f(x, y, z) = 8x^2 + 4yz - 16z \tag{4.15}$$

subject to the constraint

$$4x^2 + y^2 + 4z^2 = 16. \tag{4.16}$$

Note that (4.16) defines an ellipsoid of revolution. The equation $\nabla f = \lambda \nabla c$ yields

$$16x = 8\lambda x \tag{4.17}$$
$$4z = 2\lambda y \tag{4.18}$$
$$4y - 16 = 8\lambda z. \tag{4.19}$$

From (4.18) we find $z = \lambda y/2$ and then from (4.19) $4y - 16 = 4\lambda^2 y$, i.e.

$$y = \frac{4}{1-\lambda^2}, \quad z = \frac{2\lambda}{1-\lambda^2}.$$

Equation (4.17) has the solutions $x = 0$ and $\lambda = 2$ if $x \neq 0$. Assume first $x = 0$. Substituting y, z and $x = 0$ into (4.16) yields a single equation for λ which can be manipulated to $\lambda^2(3 - \lambda^2) = 0$, i.e. $\lambda = 0$ or $\lambda^2 = 3$. Setting $\lambda = 0$ leads to $y = 4$, $z = 0$, and $f(0, 4, 0) = 0$. For $\lambda = \mp\sqrt{3}$ we find $y = 2$ and $z = \pm\sqrt{3}$, with values $f(0, -2, \pm\sqrt{3}) = \mp 24\sqrt{3}$.

If $x \neq 0$ we have $\lambda = 2$ and so $y = z = -4/3$. The missing value of x is again found from (4.16) as $x = \pm 4/3$. The values of f at these points are both $128/3$. Thus the maxima and minima of f are

$$f_{max} = f(\pm 4/3, -4/3, -4/3) = 128/3, \quad f_{min} = f(0, -2, \sqrt{3}) = -24\sqrt{3}.$$

The level surfaces for the minimum and maximum values of f and the constraint ellipsoid are shown in Figure 4.3. We see in this figure that the solution points are points of tangency of a level surface and the constraint surface.

4.4.2 Several Equality Constraints

If several constraints are present, the situation is trivial when the number of (independent) constraints equals the number of variables. In this case all constraints typically are satisfied only by a finite number of points, if any, and one merely

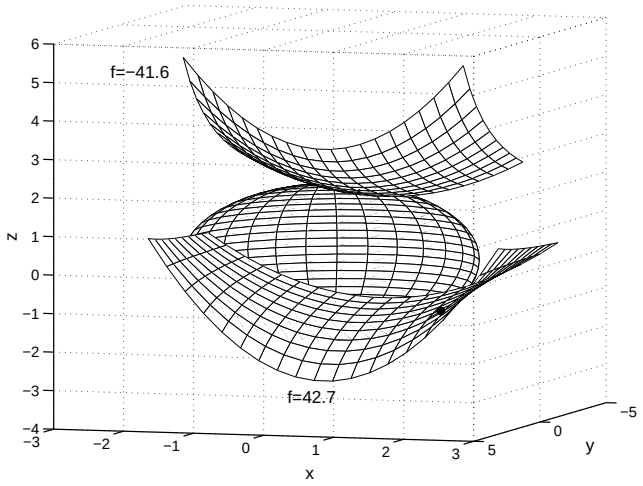


FIGURE 4.3: Level surfaces $f = f_{min} \approx -41.6$ and $f = f_{max} \approx 42.7$ for $f(x, y, z)$ defined by equation (4.15). Both level surfaces have a tangential contact with the constraint ellipsoid (4.16).

has to evaluate the objective function at these points to find the global maxima or minima. Lagrange multiplies are needed if the number of constraints is smaller than the number of variables.

Consider for simplicity the case of three variables (x, y, z) and two constraints $c_1(x, y, z) = 0$, $c_2(x, y, z) = 0$. Each of the two constraints defines a surface in three dimensional (x, y, z)–space, and both constraints together define a curve C, the intersection of the two constraint surfaces. (Two non–parallel planes in three dimensional space intersect in a straight line. Likewise, two curved surfaces typically intersect in a curve.) Now a level set $f(x, y, z) = f_0$ also defines a surface, and the condition for f to have an extreme point when restricted to C is again that a level surface and C meet tangentially at some point (x_0, y_0, z_0). This condition means that the tangent line of C at the point of contact is entirely in the tangent plane of the level surface. Since the tangent line of C is the intersection of the tangent planes of the two constraint surfaces, the tangency condition means that all three tangent planes intersect in a line. This is a special condition because in general three planes in three dimensional space have only a single point in common.

As in two dimensions, the gradient $\nabla f(x_0, y_0, z_0)$ is orthogonal to the tangent plane of the level surface $f(x, y, z) = f(x_0, y_0, z_0)$ at (x_0, y_0, z_0). The same holds for the tangent planes of the constraint surfaces $c_1 = 0$ and $c_2 = 0$. The condition that these planes intersect in a line implies that the three gradient vectors to which they are orthogonal are all located in the normal plane of that line and hence are linearly dependent as illustrated in Figure 4.4. Thus one of these gradient vectors is a linear combination of the other two, which we write as $\nabla f = \lambda_1 \nabla c_1 + \lambda_2 \nabla c_2$. For more variables and constraints the situation is similar.

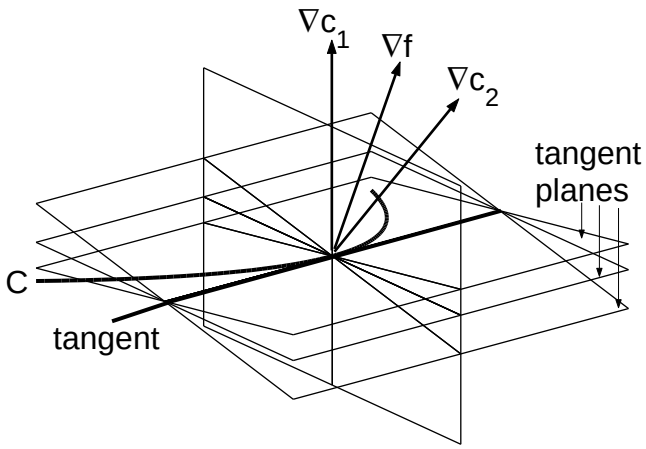


FIGURE 4.4: At a solution point of a three–variable optimization problem with two constraints the tangent plane of the level surface of f and the tangent planes of the two intersecting constraint surfaces $c_1 = 0$ and $c_2 = 0$ intersect in the tangent of the constraint curve C. As a consequence all three gradients are in the normal plane of C and so are linearly dependent.

EXAMPLE 4.8

Find the maxima and minima of

$$f(x, y, z) = x^2 + y^2 - z$$

subject to

$$\begin{aligned} x^2 + y^2 &= 1 \\ x^2 + z^2 &= 1. \end{aligned}$$

Here we can find a parametric representation of the constraint curve C. Substituting $x^2 = 1 - z^2$ from the second constraint equation into the first constraint equation yields $y^2 = z^2$, i.e. $z = \pm y$. The first constraint defines a circle which we parametrize as $x = \cos\varphi$, $y = \sin\varphi$, where $-\pi \le \varphi \le \pi$. Thus the constraints define two curves

$$C_\pm : \quad (x, y, z) = (\cos\varphi, \sin\varphi, \pm\sin\varphi).$$

Note that the two curves intersect if $z = 0$, i.e., at $\varphi = 0$ and $\varphi = \pi$.

To solve the constrained optimization problem we substitute the parametric representation of $C_\pm$ into f and set

$$f_\pm(\varphi) = 1 \mp \sin\varphi.$$

The extreme points are determined by $df_\pm/d\varphi = \mp\cos\varphi = 0$, hence $\varphi = \pm\pi/2$, with values $f_\pm(\mp\pi/2) = 2$ and $f_\pm(\pm\pi/2) = 0$. Thus there are two maxima at $(0, \pm 1, -1)$ and two minima at $(0, \pm 1, 1)$ with values 2 and 0, respectively. The intersecting

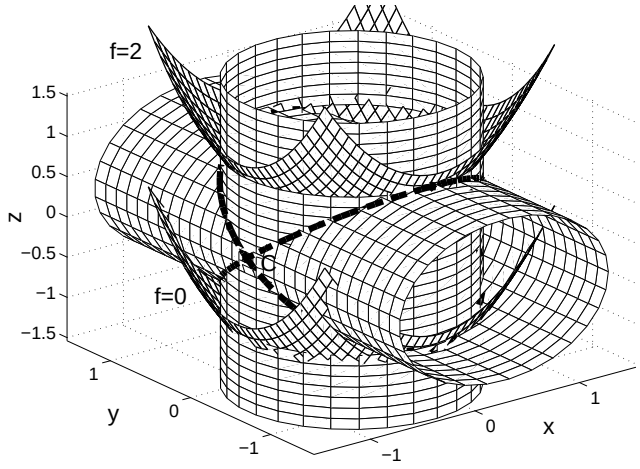


FIGURE 4.5: Intersecting constraint cylinders and level surfaces for Example 4.8.

constraint cylinders and the level surfaces for the maximum and minimum values are shown in Figure 4.5. It can be easily verified a posteriori that the gradient of f and the gradients of the two constraint functions are linearly dependent at the four extreme points.

4.4.3 Inequality Constraints

Finally consider the case of inequality constraints for a problem with n variables. Inequality constraints define a feasible region S in n–dimensional space, and the objective function is restricted to S. Extreme points can be located in the interior of S as well as on the boundary. If there are no solutions to $\nabla f = 0$ in the interior, all extreme points are on the boundary. Assume that $c(x) \geq 0$ is one of the inequality constraints. The boundary of this constraint is the hypersurface defined by $c(x) = 0$. Finding an extreme point on this boundary amounts to solving an optimization problem with a single equality constraint (and possibly an additional set of inequality constraints). If two inequality constraints $c_1 \geq 0$, $c_2 \geq 0$ are present, the optimal solution may also be located on the intersection of the two boundary hypersurfaces $c_1 = c_2 = 0$ which leads to a problem with two equality constraints etc. The situation is naturally much more complicated than in linear programming. Linear programming problems do not have solutions in the interior of the feasible region.

EXAMPLE 4.9

Consider the problem of minimizing the objective function

$$f(x, y) = \frac{x^4}{4} - \frac{x^2}{2} + \frac{y^2}{2}.$$

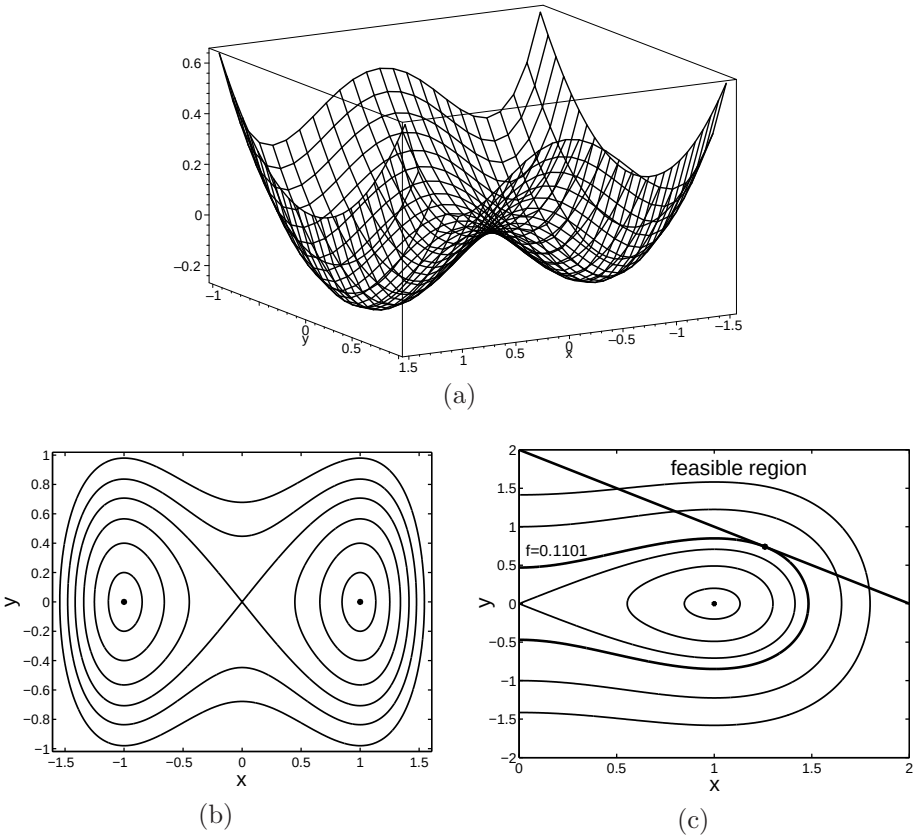


FIGURE 4.6: (a): Three dimensional plot of $f(x,y) = x^4/4 - x^2/2 + y^2/2$. (b): Level contours of f. (c): Contours of f in the right half plane and the constraint boundary $x + y = 2$.

Unconstrained optimization leads to the equations

$$\begin{aligned} \frac{\partial f}{\partial x} &= x^3 - x = 0 \quad \Rightarrow \quad x = 0 \text{ or } x = \pm -1 \\ \frac{\partial f}{\partial y} &= y = 0. \end{aligned}$$

To check the types of the extreme points $(0,0)$ and $(\pm 1, 0)$ we compute the Hessean matrices,

$$Hf(0,0) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad Hf(\pm 1, 0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

From the form of these matrices it follows that $(\pm 1, 0)$ are minimum points ($f = -1/4$), and $(0,0)$ is a saddle point ($f = 0$). A three–dimensional surface plot of f is shown in Figure 4.6 (a), and some level contours are displayed in Figure 4.6 (b).

Now consider the problem of minimizing $f(x,y)$ subject to the inequality constraint

$$c(x,y) = x + y \geq 2.$$

Since $c(\pm 1, 0) < 2$, the global minima of f are not in the feasible region, hence the optimal solution must be on the boundary. We are then led to the problem of minimizing f subject to the equality constraint

$$x + y = 2.$$

The equation (4.5) leads to

$$x^3 - x = \lambda, \quad y = \lambda \quad \Rightarrow \quad x^3 - x - y = 0.$$

Substituting $y = 2 - x$ from the constraint equation into this equation gives $x^3 - 2 = 0$, with the solution $x = 2^{1/3} = 1.2600$, and hence $y = 2 - 2^{1/3} = 0.7401$. The numerical value of f at this point is 0.11012. Note that the equation for x also follows directly from the unconstrained, single variable optimization problem associated with $f(x, 2 - x)$.

In Figure 4.6 (c) the constraint line and some level contours are shown. The solution point is again revealed as point of tangency.

4.5 MODELING EXAMPLES

EXAMPLE 4.10

A manufacturerer of colored TV's is planning the introduction of two new products: a 19–inch stereo color set with a manufacturerer's suggested retail price of \$339 per year, and a 21–inch stereo color set with a suggested retail price of \$339 per year. The cost of the company is \$195 per 19–inch set and \$225 per 21–inch set, plus additional fixed costs of \$400,000 per year. In the competitive market the number of sales will affect the sales price. It is estimated that for each type of set, the sales price drops by one cent for each additional unit sold. Furthermore, sales of the 19–set will affect sales of the 21–inch set and vice versa. It is estimated that the price for the 19–inch set will be reduced by an additional 0.3 cents for each 21–inch sold, and the price for 21–inch sets will decrease for by 0.4 cents for each 19–inch set sold. The company believes that when the number of units of each type produced is consistent with these assumptions all units will be sold. How many units of each type of set should be manufactured such the profit of the company is maximized?

The relevant variables of this problem are:

s_1: number of units of the 19–inch set produced per year,
s_2: number of units of the 21–inch set produced per year,
p_1: sales price per unit of the 19–inch set (\$),
p_2: sales price per unit of the 21–inch set (\$),
C: manufacturing costs (\$ per year),
R: revenue from sales (\$ per year),
P: profit from sales (\$ per year).

The market estimates result in the following model equations,

$$\begin{aligned} p_1 &= 339 - 0.01s_1 - 0.003s_2 \\ p_2 &= 399 - 0.04s_1 - 0.01s_2 \\ R &= s_1p_1 + s_2p_2 \\ C &= 400,000 + 195s_1 + 225s_2 \\ P &= R - C. \end{aligned}$$

The profit then becomes a nonlinear function of (s_1, s_2),

$$P(s_1, s_2) = -400,000 + 144s_1 + 174s_2 - 0.01s_1^2 - 0.007s_1s_2 - 0.01s_2^2. \qquad (4.20)$$

If the company has unlimited resources, the only constraints are $s_1, s_2 \geq 0$.

Unconstrained Optimization. We first solve the unconstrained optimization problem. If P has a maximum in the first quadrant this yields the optimal solution. The condition for an extreme point of P leads to a linear system of equations for (s_1, s_2),

$$\begin{aligned} \frac{\partial P}{\partial s_1} &= 144 - 0.02s_1 - 0.007s_2 = 0 \\ \frac{\partial P}{\partial s_2} &= 174 - 0.007s_1 - 0.02s_2 = 0. \end{aligned}$$

The solution of these equations is $s_1^* = 4735$, $s_2^* = 7043$ with profit value $P^* = P(s_1^*, s_2^*) = 553,641$. Since s_1^*, s_2^* are positive, the inequality constraints are satisfied. To determine the type of the extreme point we inspect the Hessean matrix,

$$HP(s_1^*, s_2^*) = \begin{bmatrix} -0.02 & -0.007 \\ -0.007 & -0.02 \end{bmatrix}.$$

A sufficient condition for a maximum is that $(HP)_{11} < 0$ and $\det(HP) > 0$. Both of these conditions are satisfied and so our solution point is indeed a maximum, in fact a global maximum. In Figure 4.7 (a) a three–dimensional plot of $P(s_1, s_2)$ is shown. Some level contours are displayed in Figure 4.7 (b). The level contours play here the role of isoprofit lines. Because P is a nonlinear function, the isoprofit lines form closed curves that surround the maximum at (s_1^*, s_2^*).

Constrained Optimization. Now assume the company has limited resources which restrict the number of units of each type produced per year to

$$s_1 \leq 5,000, \quad s_2 \leq 8,000, \quad s_1 + s_1 \leq 10,000.$$

The first two constraints are satisfied by (s_1^*, s_2^*), however $s_1^* + s_2^* = 11,778$. The global maximum point of P is now no longer in the feasible region, thus the optimal solution must be on the boundary. We therefore solve the constrained optimization problem

$$\max P$$

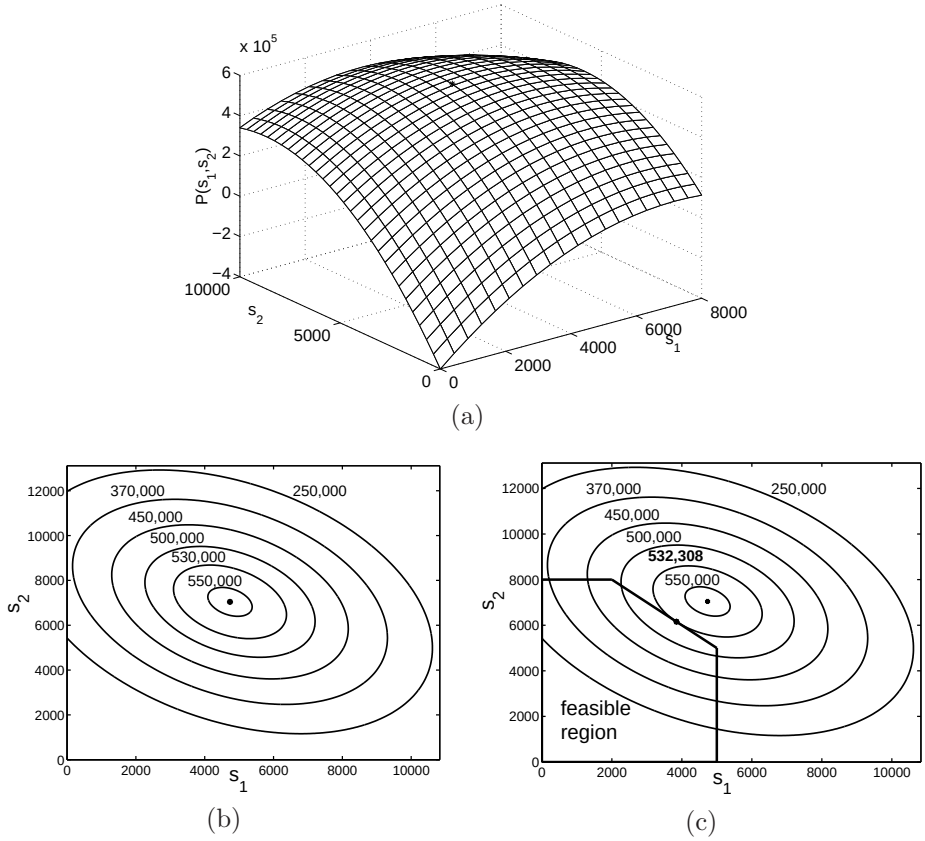


FIGURE 4.7: (a): Three dimensional plot of $P(s_1, s_2)$, equation (4.20). (b): Level contours of P. (c): Level contours of P and feasible region for the constrained optimization problem.

subject to

$$c(s_1, s_2) = s_1 + s_2 - 10,000 = 0.$$

We can either substitute s_2 or s_1 from the constraint equation into P and solve an unconstrained one–variable optimization problem, or use Lagrange multipliers. Choosing the second approach, the equation $\nabla P = \lambda \nabla c$ becomes

$$\begin{aligned} 144 - 0.02s_1 - 0.007s_2 &= \lambda \\ 174 - 0.007s_1 - 0.02s_2 &= \lambda, \end{aligned}$$

which reduces to a single equation for s_1, s_2. Together with the constraint equation we then have again a system of two linear equations,

$$\begin{aligned} -0.013s_1 + 0.013s_2 &= 30 \\ s_1 + s_2 &= 10,000. \end{aligned}$$

The solution is $s_1^* = 3846$, $s_2^* = 6154$, with profit value $P^* = 532,308$. In Figure 4.7 (c) the feasible region and some contour levels are shown. The optimal solution

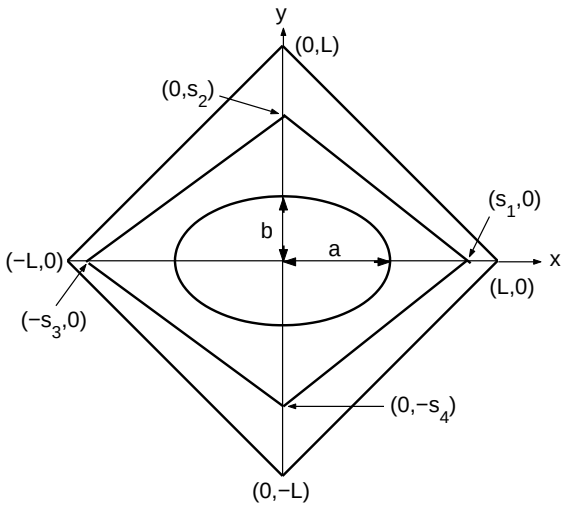


FIGURE 4.8: Geometry of the problem of Example 4.11.

is revealed as point of tangency of the isoprofit line $P = P^*$ and the constraint line. It is also clear from the figure that the solution point cannot be located on one of the two other boundary lines $s_1 = 5,000$ or $s_2 = 8,000$.

EXAMPLE 4.11

A fish farm has a fish lake on a square area. The length of the diagonal of the square is $2L$. The fish lake has the shape of an ellipse with semi–axes a and b. The center of the lake is at the center of the square and the semi–axes are on the diagonals. The owner of the fish farm has fencing material of length l where $l < 4\sqrt{2}L$. She wants to surround the lake by a fence in the form of a quadrilateral whose corner points are on the diagonals of the square. In order that the owner has enough space to work at the lake, the distance between fence and lake must not be smaller than a given distance d_m. What is the position of the corner points of the fence such that the enclosed area is maximal?

To formulate this problem as a nonlinear program, we introduce a (x, y)–coordinate whose origin is at the center of the square. The corner points of the square are $(\pm L, 0)$ and $(0, \pm L)$. The equation of the fish lake's boundary is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

The corner points of the fence's quadrilateral have coordinates $(s_1, 0)$, $(0, s_2)$, $(-s_3, 0)$, and $(0, -s_4)$ $(0 \leq s_j \leq L)$ with (s_1, s_2, s_3, s_4) to be determined, see Figure 4.8.

To invoke the distance restriction, we have to compute the minimal distance between the ellipse and the four edges of the quadrilateral. Consider the edge in

the first quadrant. The equation of this edge is $y = (s_2/s_1)(s_1 - x)$. Some thought reveals that the minimal distance between this straight line and the ellipse is given by

$$(s_1 s_2 - d(s_1, s_2))/\sqrt{s_1^2 + s_2^2}, \tag{4.21}$$

where

$$d(s_1, s_2) = \sqrt{a^2 s_2^2 + b^2 s_1^2},$$

provided $s_1 s_2 \geq d(s_1, s_2)$. Thus the minimum distance condition for this edge can be formulated as

$$s_1 s_2 - d(s_1, s_2) \geq d_m \sqrt{s_1^2 + s_2^2}.$$

The minimum distance conditions for the other three edges are obtained by replacing (s_1, s_2) in this inequality by (s_3, s_2), (s_3, s_4), and (s_1, s_4), respectively.

The area enclosed by the fence is

$$A(s_1, s_2, s_3, s_4) = \frac{1}{2}(s_1 s_2 + s_2 s_3 + s_3 s_4 + s_4 s_1).$$

Now the optimization problem can be formulated as

$$\max A(s_1, s_2, s_3, s_4)$$

subject to the inequality constraints

$$\begin{aligned}
s_1 s_2 - d(s_1, s_2) &\geq d_m \sqrt{s_1^2 + s_2^2} \\
s_3 s_2 - d(s_3, s_2) &\geq d_m \sqrt{s_3^2 + s_2^2} \\
s_3 s_4 - d(s_3, s_4) &\geq d_m \sqrt{s_3^2 + s_4^2} \\
s_1 s_4 - d(s_1, s_4) &\geq d_m \sqrt{s_1^2 + s_4^2} \\
s_j &\leq L \quad (1 \leq j \leq 4),
\end{aligned}$$

and the equality constraint

$$\sqrt{s_1^2 + s_2^2} + \sqrt{s_2^2 + s_3^2} + \sqrt{s_3^2 + s_4^2} + \sqrt{s_4^2 + s_1^2} = l.$$

Note that we don't need to impose the constraints $s_1 \geq a$ and $s_2 \geq b$. The minimum distance requirement for (s_1, s_2) implies $s_1 s_s \geq d(s_1, s_2)$ and this can be only satisfied if $s_1 \geq a$ and $s_2 \geq b$.

PROBLEMS

4.1. Extend Example 4.2 for a collection of S schools.

4.2. Show how Newton's method for root finding can be used to calculate $\sqrt{3}$. Compute numerically an iterated sequence that converges to this value. Stop the iteration if $|x_{n+1} - x_n| \leq 10^{-5}$. What is the effect of changing the initial condition?

4.3. Use Newton's method to find the positive root of

$$g(x) = x - \tanh(2x) = 0$$

up to five decimal places.

4.4. Plot $f(x) = x\sin(x)$ in $0 \leq x \leq 15$ and convince yourself that $f(x)$ has three local maxima in that range. Compute these maxima up to five decimal places using Newton's method.

4.5. Let

$$f(x, y) = x^4 + y^3 + xy^2 + x^2 - y + 1.$$

Find the quadratic approximation of $f(x, y)$ at the points
(a) $x_0 = y_0 = 0$,
(b) $x_0 = 1,\ y_0 = 0$,
(c) $x_0 = y_0 = 2$.

4.6. Compute the Jacobian of

$$g(x) = \begin{bmatrix} x_1x_2 - x_1 - 1 \\ x_1x_2x_3 - 2x_2 \\ e^{-x_1^2} - 3x_3 - 1 \end{bmatrix}$$

at $x_0 = [0, 0, 0,]^T$ and $x_0 = [1, 1, 1]^T$.

4.7. Minimize the objective function

$$f(x_1, x_2) = 7x_1^2 + 2x_1x_2 + x_2^2 + x_1^4 + x_2^4$$

using 50 iterations of
(a) Newton's method
(b) Steepest Descent
with starting value $x_0 = (3, 3)^T$. Plot the values of the iterates for each method on the same graph. You may experiment with the value of α in Equation (4.1). *Hint:* start small.

4.8. Consider the system of equations

$$\begin{array}{rcrcl} g_1(x, y) & \equiv & x^3 + y^3 - 1 & = & 0, \\ g_2(x, y) & \equiv & ye^{-x} - \sin(y) - a & = & 0. \end{array}$$

Apply Newton's method to find two different solutions for each of the values $a = 0.5$ and $a = 1$. Use at most 101 iterations and truncate the computation if

$$\varepsilon \equiv |g_1(x, y)| + |g_2(x, y)| < 10^{-10}.$$

Provide the solutions, the starting values, the numbers of iterations, and the final values of ε in your answer.

4.9. Find the minimum of the function

$$f(x,y) = 7x^2 + 2xy + y^2 + x^4 + y^4 + x - y$$

using Newton's method. Use at most 101 iterations and truncate the computation if

$$\varepsilon \equiv |\frac{\partial f(x,y)}{\partial x}| + |\frac{\partial f(x,y)}{\partial y}| < 10^{-10}.$$

Provide the solution, the starting value, the number of iterations, and the final value of ε in your answer.

4.10. Find the minimum of $f(x,y)$ given in Problem 4.9 using the steepest descent method with $\alpha = 0.04$, $\alpha = 0.06$, $\alpha = 0.08$, $\alpha = 0.1$ and $\alpha = 0.12$. Choose $(x_0, y_0) = (1,1)$ as starting value. Summarize the final values of ε as defined in Problem 4.21 and the approximate solutions for each of the five values of α in a table. What is the effect of the magnitude of α on the performance of the steepest descent method?

4.11. Assume a farmer has L feet of fencing for a rectangular area with lengths x and y. Determine these lengths such that the enclosed area is a maximum.

4.12. Consider an ellipse with semi-axes $a \geq b$. The area enclosed by the ellipse is $A = \pi ab$ and the circumference is $L = 4aE(a)$, where $e = \sqrt{1 - b^2/a^2}$ is the eccentricity and $E(e)$ is the complete elliptic integral of the second kind – a given function of e. Show that the constrained optimization problem

$$\max(\pi ab)$$

subject to

$$4aE(e) = L$$

leads to the following equation for e,

$$\frac{e}{1-e^2} = -\frac{2E'(e)}{E(e)},$$

where $E'(e) = dE(e)/de$. *Note:* It turns out that the only solution of this equation is $e = 0$, i.e. $a = b$. Thus the area of an ellipse with prescribed circumference is a maximum if the ellipse degenerates to a circle.

4.13. Find all extreme points (local maxima and minima) of

$$f(x,y) = x^3 + y^2$$

subject to

$$y^2 - x^2 = 1.$$

Make a sketch showing the constraint curve, some level curves of f, and the extreme points as points of tangencies.

4.14. Find the minimum distance of the surface

$$2x^2 + y^2 - z^2 = 1$$

to the origin.

4.15. Find the points on the unit sphere

$$x^2 + y^2 + z^2 = 1,$$

for which the function

$$f(x, y, z) = 2x^2 + y^2 - z^2 - x$$

has a global maximum and a global minimum, respectively.

4.16. A manufacturer of personal computers currently sells 10,000 units per month of a basic model. The manufacture cost per unit is \$700 and the current sales price is \$950. During the last quarter the manufacturer lowered the price by \$100 in a few test markets, and the result was a 50% increase in orders. The company has been advertising its product nationwide at a cost of \$50,0000 per month. The advertising agency claims that increasing the advertising budget by \$10,000 per month would result in a sales increase of 200 units per month. Management has agreed to consider an increase in the advertising budget to no more than \$100,000 per month.

Determine the price and the advertising budget that will maximize the profit. Make a table comparing the maximal profit and the corresponding values of the price, the advertising budget, and the number of sales to their current values, and to the optimal values that would result without advertisement.

Hint: Let N be the number of sales per month. Write $N = N_0 + \Delta N_p + \Delta N_a$, where N_0 is the current value of N, ΔN_p is the increase of N due to price reduction, and ΔN_a is the increase of N due to increasing the advertising budget. Note: If you don't find a solution in the interior of the feasible region, the optimal solution is on a boundary.

4.17. A local newspaper currently sells for \$1.50 per week and has a circulation of 80,000 subscribers. Advertising sells for 250/page, and the paper currently sells 350 pages per week (50 pages/day). The management is looking for ways to increase profit. It is estimated that an increase of 10 cents/week in the subscription price will cause a drop of 5,000 subscribers. Increasing the price of advertising by \$100/page will cause the paper to lose approximately 50 pages of advertising in a week. The loss of advertising will also affect circulations, since one of the reasons people by the newspaper is the advertisement. It is estimated that a loss of 50 pages of advertisement per week will reduce circulation by 1,000 subscribers.

(a) Find the weekly subscription price and advertisement price that will maximize the profit.

(b) Same as (a), but now with the constraint that the advertising price cannot be increased beyond \$400.

Hint: Let M be the number of advertising pages per week. Write $M = M_0 + \Delta M_a$, where M_0 is the current value of M, and ΔM_a is the change caused by increasing the advertising price. Proceed similarly for N, the number of subscribers. Here you have to consider two causes of change.

4.18. Verify the expression (4.21) in Example 4.11.

In Exercises 4.19–4.25 use an optimization software such as the *fmincon* function of Matlab to find the optimal solution.

4.19. Redo the problem of Example 4.10, but now choose as objective function the marginal profit, i.e., the ratio $(R-C)/C$ of the profit and the total manufacturing costs.

4.20. Maximize the volume xyz of a cardboard subject to the equality constraint $xy + xz + yz = 4$ and the inequality constraints

$$0 \le x \le 0.5$$
$$2 \le y \le 3$$
$$z \ge 1.$$

4.21. Find the (unconstrained) minimum of

$$f(x, y, z) = x^6 + x^2 * y^2 + y^4 + z^4 + e^{-z^2} \sin(x + y).$$

4.22. Find the minimum and maximum of

$$f(x, y) = x^3 + y^2 - xy$$

subject to

$$x^2 + 4y^2 \le 2.$$

4.23. Find the minimum of

$$f(x, y, z) = \sin(x + y) + \cos(y + z) - e^{-x^2}$$

subject to

(a)

$$x^2 + y^2 = 1, \quad z^2 \le 1, \quad x^2 \ge y^2.$$

(b) constraints as in (a) and in addition

$$x \ge 0, \quad y \le 0.$$

4.24. Solve the fencing problem of Example 4.11 for $L = 4$, $a = 1.5$, $b = 2.5$, and
(a) $l = 20$, $d_m = 0.3$,
(b) $l = 20$, $d_m = 0.4$,
(c) $l = 17$, $d_m = 0.1$.
Hint: A good starting value for s_1 is $(a + L)/2$.

4.25. Solve the school problem of Example 4.2 for five districts with coordinates

x_j	0	0	0	−100	100
y_j	0	100	−100	0	0

and
(a) $r_1 = 200$, $r_2 = 300$, $r_3 = 200$, $r_4 = 500$, $r_5 = 300$, $c_1 = 1500$, $c_2 = 1500$,
(b) $r_1 = 200$, $r_2 = 400$, $r_3 = 200$, $r_4 = 500$, $r_5 = 300$, $c_1 = 700$, $c_2 = 2000$.
Hint: A reasonable starting value for w_{ij} is $r_j/2$. For the coordinates (a, b, c, d) you may try $(0, 0, 0, 0)$, $(100, 0, -100, 0)$, or $(50, 50, -50, -50)$.

CHAPTER 5

Empirical Modeling with Data Fitting

In this chapter the model building is *empirical* in nature, i.e., the functional form of the relationship between the dependent and independent variables is found by direct examination of data related to the process.

Data fitting problems have several common elements. The *model* has the general form

$$y = f(x_1, \ldots, x_N; w_1, \ldots, w_M)$$

and the *parameters* w_i are determined empirically from the *observations*

$$\{(x^{(i)}, y^{(i)})\}_{i=1}^{P}$$

by requiring f to be such that

$$y^{(i)} = f(x_1^{(i)}, \ldots, x_N^{(i)}; w)$$

or at least that the *error function* $E(w)$ defined by

$$S(w) = \sum_i (y^{(i)} - f(x_1^{(i)}, \ldots, x_N^{(i)}; w_1, \ldots, w_M))^2$$

be *small*. This error function seeks to minimize the sum of squares of the *residuals*. As we shall see, other error functions are possible but least squares is certainly the most widely used and we will focus on this approach at the outset. Later in this chapter in Section 5.3 we will also consider the important case of uniform approximation.

EXAMPLE 5.1

A *Radial basis function* model has the form

$$f(x; w, c) = w_0 + \sum_{k=1}^{N_c} w_k \phi(\|x - c_k\|)$$

where the w_k are the weights and the c_k are the centers of the basis functions. An example of a radial basis function is

$$\phi(r) = \exp(-r^2)$$

The norm $\|\cdot\|$ is generally taken to be the Euclidean distance.

5.1 LINEAR LEAST SQUARES

In this section we begin by revisiting an example from the previous chapter followed by a general formulation of linear least squares and some simple extensions to exponential fits.

5.1.1 The Mammalian Heart Revisited

Recall from Example 2.11 in Subsection 2.3.2 that a sequence of proportionalities produced the model

$$r = kw^{-1/3}$$

where w is the body weight of a mammal and r is its heart rate. The data on Figure 2.8 corresponds to observations

$$\{(w_i^{-1/3}, r_i)\}$$

collected for various measured rates and weights. The residual error for the ith measurement is

$$\epsilon_i = r_i - kw_i^{-1/3}$$

and the total squared error is

$$E = \sum_{i=1}^{P} \epsilon_i^2$$

We rewrite this error as a function of the unknown slope parameter k as

$$E(k) = \sum_{i=1}^{P} (r_i - kw_i^{-1/3})^2$$

To minimize E as a function of k we compute the derivative of E w.r.t. k, i.e.,

$$\frac{dE}{dk} = \sum_{i=1}^{P} 2(r_i - kw_i^{-1/3}) \cdot (-w_i^{-1/3}) = 0$$

From which it follows that

$$k = \frac{\sum_{i=1}^{P} r_i w_i^{-1/3}}{\sum_{i=1}^{P} w_i^{-2/3}}$$

Thus we can obtain an estimate for the slope of the line empirically from the data.

5.1.2 General Formulation

In this section we focus our attention to one of the most widely used models

$$f(x; m, b) = mx + b$$

To clean-up the notation we now use subscripts to label points for domain data that is one dimensional; we used superscripts in the previous section when the dimension

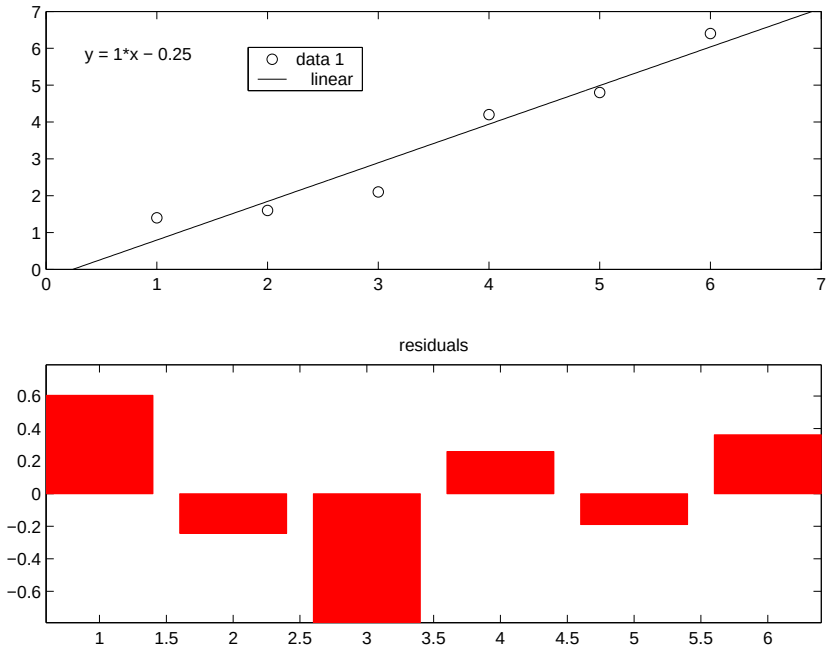


FIGURE 5.1: Linear least squares. The line $y = mx + b$ is determined such that the residuals ϵ_i^2 are minimized.

of the domain could exceed one. For a set of observations $\{(x_i, y_i)\}$, $i = 1, \ldots, P$, the total squared error is given by

$$E(m, b) = \sum_{i=1}^{P} (y_i - mx_i - b)^2 \tag{5.1}$$

Now because there are two parameters that determine the error function the necessary condition for a minimum is now

$$\frac{\partial E}{\partial m} = 0$$
$$\frac{\partial E}{\partial b} = 0$$

Solving the above equations gives the slope of the line as

$$m = \frac{(\sum y_i)(\sum x_i) - P \sum y_i x_i}{(\sum x_i)^2 - P \sum x_i^2}$$

and its intercept to be

$$b = \frac{-(\sum y_i)(\sum x_i^2) + (\sum x_i)(\sum y_i x_i)}{(\sum x_i)^2 - P \sum x_i^2}$$

Interpolation Condition. In this section we present another route to the equations for m and b produced in the previous section. Again, the input data is taken as $\{x_i\}$, the output data is $\{y_i\}$ and the model equation is $y = mx + b$. Applying the *interpolation condition* for each observation we have

$$\begin{aligned} y_1 &= mx_1 + b \\ y_2 &= mx_2 + b \\ y_3 &= mx_3 + b \\ &\vdots \\ y_P &= mx_P + b \end{aligned}$$

In terms of matrices

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_P \end{pmatrix} \begin{pmatrix} b \\ m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_P \end{pmatrix}$$

In terms of matrices we can summarize the above as

$$Xb = y$$

We can reveal the relationship between the previous approach using calculus and this approach with the interpolation condition by hitting both sides of the above matrix equation with the transpose X^T

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_P \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_P \end{pmatrix} \begin{pmatrix} b \\ m \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_P \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_P \end{pmatrix}$$

In terms of the matrices,

$$X^T X b = X^T y$$

Multiplying out produces the equations that are seen to be the same as those in the above section, i.e.,

$$\begin{pmatrix} P & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} b \\ m \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

In linear algebra these equations are referred to as the *normal* equations.

There are many algorithms in the field of numerical linear algebra developed precisely for solving the problem

$$Xb = y$$

We will consider these more in the problems.

5.1.3 Exponential Fits

We have already seen models of the form

$$y = kx^n$$

where n was given. How about if n is unknown? Can it be determined empirically from the data? Note now that the computation of the derivative of the error function w.r.t. n is now quite complicated. This problem is resolved by converting it to a linear least squares problem now in terms of logarithms. Specifically,

$$\begin{aligned} \ln y &= \ln(kx^n) \\ &= \ln k + \ln x^n \\ &= \ln k + n \ln x \end{aligned}$$

This is now seen to be a linear least squares problem

$$y' = nx' + k'$$

where we have made the substitutions $y' = \ln y, k' = \ln k$ and $x' = \ln x$. Now one can apply the standard least squares solution to determine n and k'. The value of k can be found as well by

$$k = \exp(k')$$

5.1.4 Fitting Data with Polynomials

In the previous section we consider the basic linear model $f(x; c_0, c_1) = c_0 + c_1 x$. The simplest extension to this is the second order polynomial

$$f(x; c_0, c_1, c_2) = c_0 + c_1 x + c_2 x^2$$

Note first that adding the term $c_2 x^2$ will change the least square fit values of the coefficients c_0, c_1 obtain from the linear model and hence all the coefficients c_0, c_1 and c_2 must be computed. The least squares procedure follows along lines similar to the previous section. We assume a set of observations $\{(x_i, y_i)\}$ and define the sum of the squares of the model residuals to be the error function, i.e.,

$$E(c_0, c_1, c_2) = \sum_{i=1}^{P} (y_i - c_0 - c_1 x_i - c_2 x_i^2)^2 \tag{5.2}$$

Now requiring

$$\begin{aligned} \frac{\partial E}{\partial c_0} &= 0 \\ \frac{\partial E}{\partial c_1} &= 0 \\ \frac{\partial E}{\partial c_2} &= 0 \end{aligned}$$

The resulting necessary conditions, written in terms of matrices, are then

$$\begin{pmatrix} P & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum y_i x_i^2 \end{pmatrix} \tag{5.3}$$

As for the linear model, it is possible to solve for the parameters c_0, c_1 and c_2 analytically, i.e., in closed form. It is less cumbersome to write Equation (5.3) as

$$Xc = z$$

where c is the column vector made up of the elements (c_0, c_1, c_2) and z is the column vector comprised of the elements $(\sum y_i, \sum x_i y_i, \sum y_i x_i^2)$ and X is the 3×3 matrix on the left of Equation (5.3). Now a computer package can be used to easily solve the resulting matrix equation.

Lagrange Polynomials. Consider the first degree polynomial defined by

$$P_1(x) = \frac{x - x_2}{x_1 - x_2} y_1 + \frac{x - x_1}{x_2 - x_1} y_2$$

By construction, it is clear that

$$P_1(x_1) = y_1$$

and

$$P_1(x_2) = y_2$$

So we have found the unique line passing through the points $\{(x_1, y_1), (x_2, y_2)\}$.

This procedure may be continued analogously for more points. A second degree polynomial passing through three prescribed points is given by

$$P_2(x) = \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)} y_1 + \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)} y_2 + \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)} y_3$$

By construction, it again may be verified that

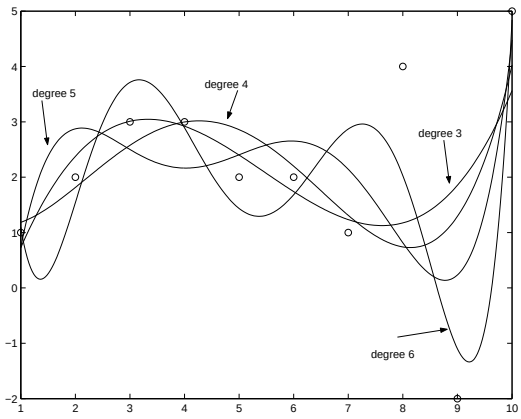
$$P_2(x_1) = y_1,$$

$$P_2(x_2) = y_2,$$

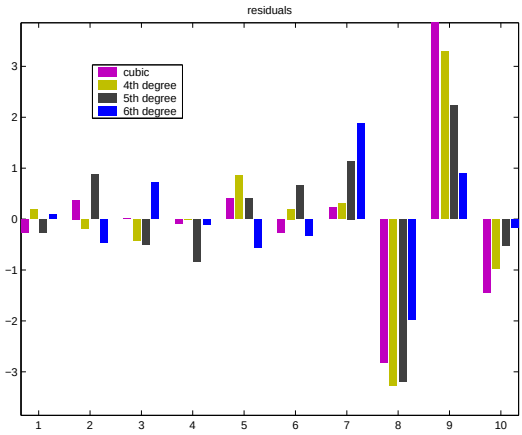
and

$$P_2(x_3) = y_3.$$

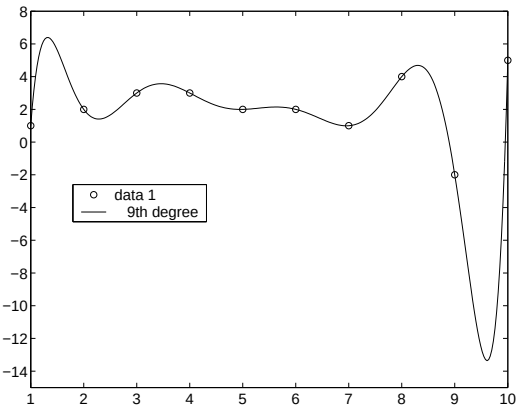
A pattern has emerged that makes it apparent that this simple procedure may be employed to fit a polynomial of degree n through a set of $n + 1$ points.



(a) Degree 3, 4, 5, 6 polynomials fit to a data set.



(b) Residual plot of degree 3, 4, 5, 6 polynomial fits.



(c) Degree 9 polynomial fit.

FIGURE 5.2: Comparative polynomial approximation to a 10 point data set.

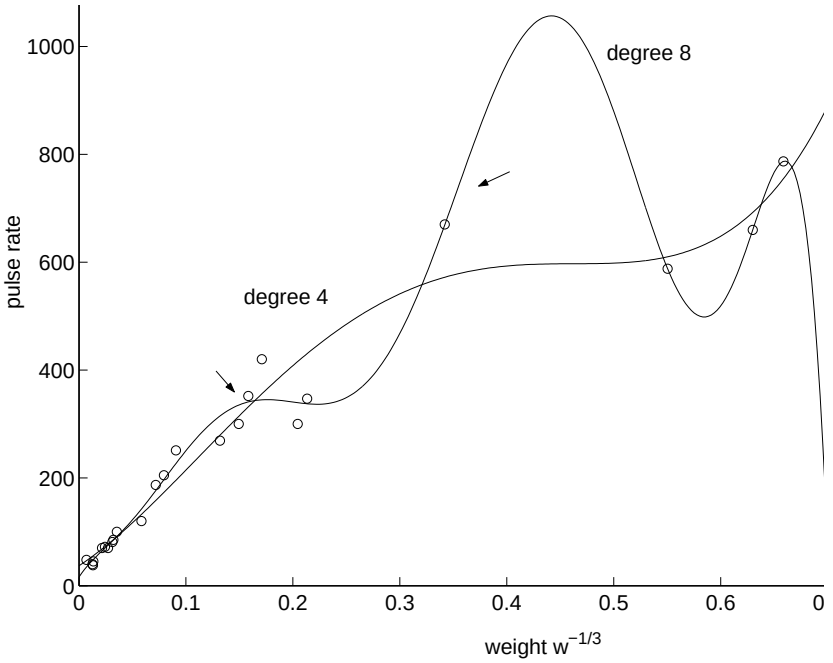


FIGURE 5.3: Degree 4 and degree 8 polynomials applied to the data $(w_i^{-1/3}, r_i)$.

5.1.5 Interpolation versus Least Squares

Dangers of Higher Order Polynomials. It is clear from the preceding section that we can always find a polynomial of degree n to exactly fit (i.e., satisfy the interpolation condition) $n+1$ points. So why not simply use high order polynomials? To answer this question consider the model

$$f(x) = c_{20}x^{20} + c_{19}x_{19} + \cdots + c_1x_1 + c_0$$

If one of the coefficients is perturbed by a small value ϵ the resulting model may predict wildly different results. For example, let

$$g(x) = (c_{20} + \epsilon)x^{20} + c_{19}x_{19} + \cdots + c_1x_1 + c_0$$

Then

$$g(x) = \epsilon x^{20} + f(x)$$

So, even if ϵ is small the difference between $f(x)$ and $g(x)$ is potentially very large. This is a manifestation of ill-conditioning of high degree polynomials, i.e., small changes in parameters may result in large changes in the function.

See Figure 5.3 for an example of the oscillations that appear with higher order polynomials.

5.2 SPLINES

One obvious procedure for reducing the need for higher order polynomials is to restrict each polynomial for the description of limited contiguous data subsets. This is the central idea behind *splines*. A spline is a piecewise defined function

$$S(x) = \begin{cases} S_1(x) & \text{if} \quad x_1 \le x < x_2, \\ S_2(x) & \text{if} \quad x_2 \le x < x_3, \\ \vdots & \vdots \\ S_j(x) & \text{if} \quad x_j \le x < x_{j+1}, \\ \vdots & \vdots \\ S_n(x) & \text{if} \quad x_n \le x < x_{n+1} \end{cases} \tag{5.4}$$

that satisfies the interpolation conditions for all the data points

$$S(x_j) = y_j = S_j(x_j) \tag{5.5}$$

Furthermore, the piecewise defined models may be joined by enforcing *auxiliary conditions* to be described. We begin with the simplest case.

5.2.1 Linear Splines

For linear splines the piecewise function takes the form

$$S_j(x) = c_0^j + c_1^j x$$

and this is valid over the interval $x_j \le x < x_{j+1}$. (See Figure 5.4.)

For $x_1 \le x < x_2$ the function $S_1(x)$ is the line passing through the points (x_1, y_1) and (x_2, y_2). The interpolation condition $S_1(x_1) = y_1$ requires

$$c_0^1 + c_1^1 x_1 = y_1.$$

The matching condition

$$S_1(x_2) = S_2(x_2) = y_2$$

requires

$$c_0^1 + c_1^1 x_2 = y_2$$

This system of two equations in two unknowns has the solutions

$$c_0^1 = \frac{x_2 y_1 - x_1 y_2}{x_2 - x_1}$$

and

$$c_1^1 = \frac{-y_1 + y_2}{x_2 - x_1}$$

Similarly for $S_2(x)$

$$S_2(x) = c_0^2 + c_1^2 x$$

The parameters c_0^2 and c_1^2 are then determined by conditions

$$\begin{cases} S_2(x_2) = y_2 \\ S_2(x_3) = S_3(y_3) = y_3 \end{cases} \tag{5.6}$$

which can be found to be

$$c_0^2 = \frac{x_3 y_2 - x_2 y_3}{x_3 - x_2}$$

and

$$c_1^2 = \frac{-y_2 + y_3}{x_3 - x_2}$$

In general, it follows

$$\begin{cases} S_j(x_j) = y_j \\ S_j(x_{j+1}) = S_{j+1}(x_{j+1}) = y_{j+1} \end{cases} \tag{5.7}$$

which can be found to be

$$c_0^j = \frac{x_{j+1} y_j - x_j y_{j+1}}{x_{j+1} - x_j}$$

and

$$c_1^j = \frac{-y_j + y_{j+1}}{x_{j+1} - x_j}$$

5.2.2 Cubic Splines

Notice that with linear splines that the function matches at interpolation points, i.e.,

$$S_j(x_{j+1}) = S_{j+1}(x_{j+1})$$

but that in general the derivative does not match, i.e.,

$$S_j^{'}(x_{j+1}) = S_{j+1}^{'}(x_{j+1})$$

This potential problem may be overcome by employing cubic splines (quadratic splines do not provide enough parameters).

For cubic splines the piecewise function takes the form

$$S_j(x) = c_0^j + c_1^j x + c_2^j x^2 + c_3^j x^3$$

and this is valid over the interval $x_j \leq x < x_{j+1}$. To simplify notation we will only consider two segments of $S(x)$

$$S(x) = \begin{cases} S_1(x) & \text{if} \quad x_1 \leq x < x_2, \\ S_2(x) & \text{if} \quad x_2 \leq x < x_3 \end{cases} \tag{5.8}$$

Now

$$S_1(x) = c_0^1 + c_1^1 x + c_2^1 x^2 + c_3^1 x^3$$

and

$$S_2(x) = c_0^2 + c_1^2 x + c_2^2 x^2 + c_3^2 x^3$$

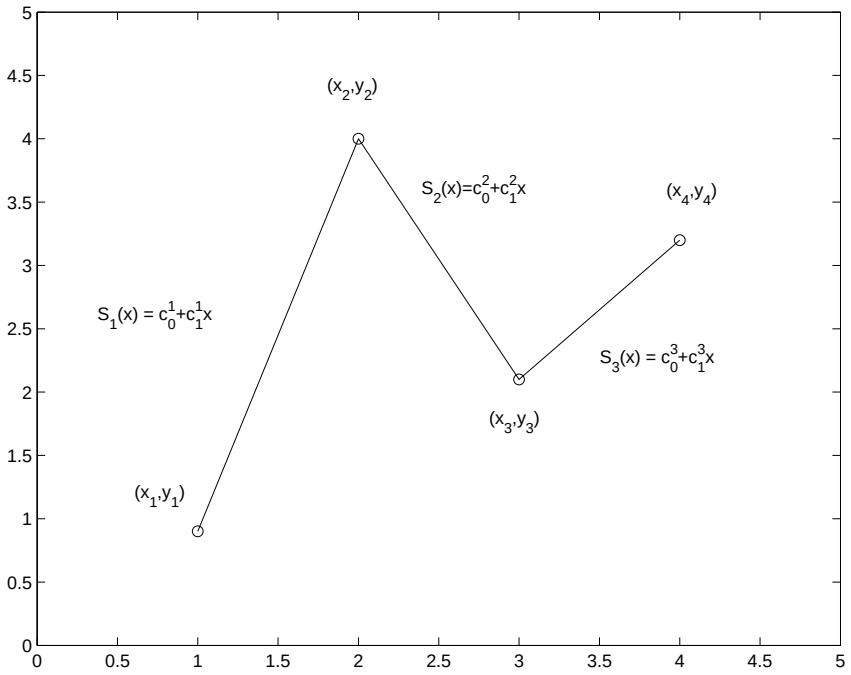


FIGURE 5.4: Linear splines.

And we observe that there are 8 parameters.

Three equations are obtained by employing the interpolation conditions

$$\begin{cases} S_1(x_1) = y_1 \\ S_2(x_2) = y_2 \\ S_2(x_3) = y_3 \end{cases} \tag{5.9}$$

The matching condition for the function value is

$$S_1(x_2) = S_2(x_2) = y_2$$

Given there are 8 parameters and only 4 equations specified we require 4 more relations.

We will require that first and second derivatives match at the interior points,

$$\begin{cases} S_1^{'}(x_2) = S_2^{'}(x_2) \\ S''_1(x_2) = S''_2(x_2) \end{cases} \tag{5.10}$$

The additional two parameters may be obtained by applying conditions on the derivatives at the endpoints x_1 and x_3. One possibility is to require

$$\begin{cases} S_1^{''}(x_1) = 0 \\ S_2^{''}(x_3) = 0 \end{cases} \tag{5.11}$$

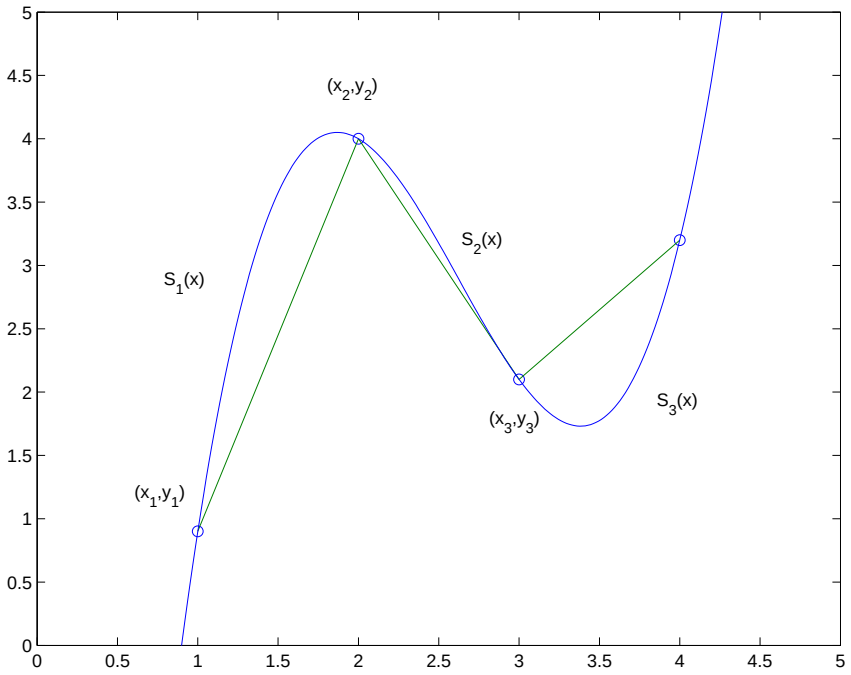


FIGURE 5.5: Cubic splines.

Since no data matching is involved these are called *natural splines*.

For a comparison of how linear and cubic splines fit a simple data set see Figure 5.5.

5.3 DATA FITTING AND THE UNIFORM APPROXIMATION

The topic of fitting a model to a data set can also be put into the framework of a linear program. Again, if we have a data set of domain (input) values $\{x_i\}$ and associated range (output) values $\{y_i\}$ we would like to determine the parameters $\{w_1, \ldots, w_k\}$ such that

$$y_i = f(x_i; w_1, \ldots, w_k)$$

An alternative to requiring that the sum of the squares of the residuals be zero is to simply minimize the maximum residual. This approach is known as the *uniform approximation*, or *Chebyshev* criterion. Since large negative residuals would make this meaningless we minimize the maximum absolute value of the residual.

To implement this idea, we first compute each residual ϵ_i as

$$\epsilon_i = y_i - f(x_i; w_1, \ldots, w_k)$$

and from all these determine the largest

$$\epsilon_{\max} = \max_i |\epsilon_i|$$

which will serve as our objective function. So the programming problem is

$$\min_i \epsilon_{\max}$$

where based on the definition of $\epsilon_{\max}$ we have the side constraints

$$|\epsilon_i| \leq \epsilon_{\max}$$

So, for a linear model, $f(x) = ax + b$ we have

$$\epsilon_i = y_i - ax_i - b$$

so the constraints become

$$|y_i - ax_i - b| \leq \epsilon_{\max}$$

or

$$-\epsilon_{\max} \leq y_i - ax_i - b \leq \epsilon_{\max}$$

Thus the linear program is to

$$\min \epsilon_{\max}$$

subject to the constraints

$$-\epsilon_{max} - y_i + ax_i + b \leq 0$$

$$-\epsilon_{\max} + y_i - ax_i - b \leq 0$$

where $i = 1, \ldots, P$.

In matrix notation we may rewrite the constraints as

$$\begin{pmatrix} -x_1 & -1 & -1 \\ x_1 & +1 & -1 \\ \vdots & \vdots & \vdots \\ -x_i & -1 & -1 \\ x_i & +1 & -1 \\ \vdots & \vdots & \vdots \\ -x_P & -1 & -1 \\ x_P & +1 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \\ \epsilon_{\max} \end{pmatrix} \leq \begin{pmatrix} -y_1 \\ +y_1 \\ \vdots \\ -y_i \\ +y_i \\ \vdots \\ -y_P \\ +y_P \end{pmatrix}$$

Now solving this linear program to implement the uniform approximation approach on the uniform noise data of Table 5.1 produces the linear model equation

$$y = 0.9937x - 0.275,$$

while the least squares error criterion produces the model

$$y = 0.8923x - 0.5466.$$

For the uniform noise data the squared error was found to be 11.543 for the uniform approximation model while for the least squares model it is 9.998. Please see the errors in Table 5.1 and the comparative plot of the two models in Figure 5.6.

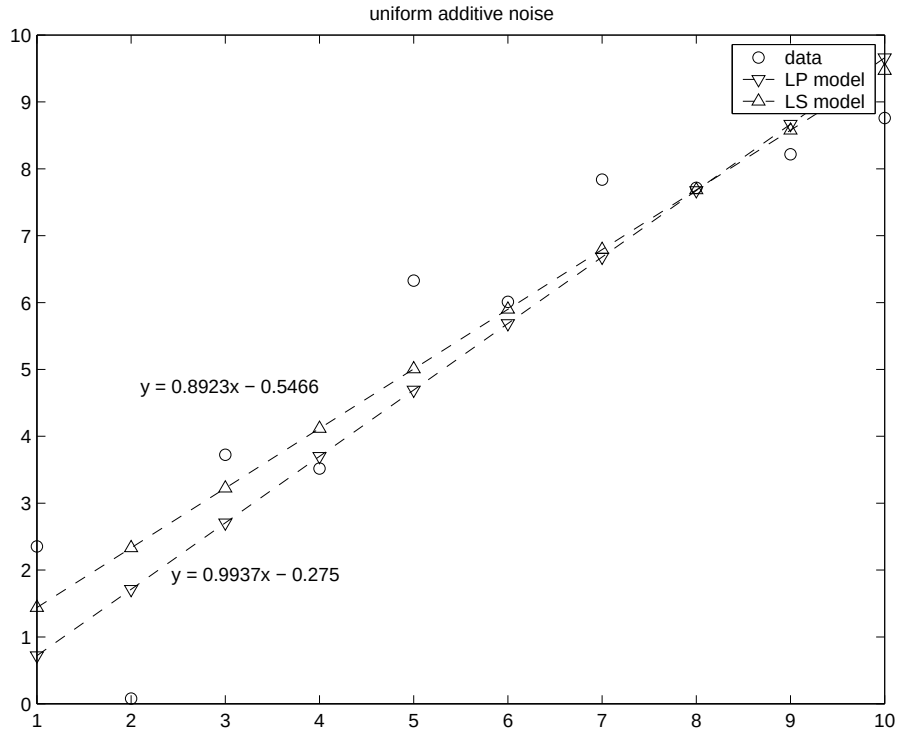


FIGURE 5.6: Least squares and uniform approximation to a linear trend with uniformly distributed additive noise.

x_i	y_i	ϵ_i uniform	ϵ_i least squares
1	2.3525	<u>1.6338</u>	0.9136
2	0.0786	<u>-1.6338</u>	<u>-2.2527</u>
3	3.7251	1.0191	0.5016
4	3.5179	-0.1818	-0.5979
5	6.3272	<u>1.6338</u>	1.3190
6	6.0113	0.3242	0.1108
7	7.8379	1.1571	1.0451
8	7.7156	0.0411	0.0305
9	8.2185	-0.4496	-0.3589
10	8.7586	-0.9032	-0.7111

TABLE 5.1: The data to be fit comprise the first two columns. The point-wise error for the uniform approximation and least squares approximation are in columns three and four, respectively. The underlined errors are those with maximum magnitude for each model. As expected, the uniform approximation has a smaller maximum error.

5.3.1 Error Model Selection?

Now we have seen that there is no unique way to compute the coefficients that fit a given model to data. The model is dependent on the way we measure the error (note that if our model is exact–e.g., the interpolation condition is satisfied–then the coefficients are unique and the error is zero). So the natural question arises: given a collection of data what is the appropriate error measure. The answer to this question lies partly in the nature of the data. If your data is very accurate, then a uniform approximation is indeed appropriate. If your data contains statistical outliers or lots of noise, a least squares approximation may be more robust.

The ultimate decision factor in what error term to use is in the added value of the model. If the predictive value of the model is superior in one error measure than another the choice is clear. Establishing superiority can often be challenging in practice.

PROBLEMS

5.1. Find the parameters a and b such that the models

$$y = ax$$

and

$$y = bx^2$$

fit the data $\{(0,0),(1,1),(2,3)\}$ according to the least squares criterion and compare the errors of the models. Without doing the calculation, can you predict what the model error would be for the model

$$y = cx + dx^2$$

Give your reasoning. *Hint:* You need not explicitly calculate c and d; considering the equations that produce them will be sufficient.

5.2. Find the line $y = b + mx$ of best fit through the data

$$\{(.1,.2),(.2,.3),(.3,.7),(.5,.2),(.75,.8)\}$$

using the least squares criterion.

5.3. Consider the model

$$f(x;c_0,c_1) = c_0x^{-3/2} + c_1x^{5/2}$$

Use the least squares approach to determine equations for c_0 and c_1 in terms of available data $(x_i,y_i)_{i=1}^P$.

5.4. Consider the mammalian pulse rate data $(r_i,w_i)_{i=1}^{24}$ provided in Table 2.1 in Subsection 2.3.2. Match the data to the models

(a) $r = b + mw^{-1/3}$

(b) $r = kw^n$

using least squares and compute the corresponding error terms given by Equation (5.1). You may use the Matlab least square codes provided, but you will first need to take appropriate transformations of the data.

5.5. Write MATLAB code to fit a second order polynomial

$$f(x;c_0,c_1,c_2) = c_0 + c_1x + c_2x^2$$

to the *linearized* mammalian heart data consisting of components $\{(w_i^{-1/3},r_i)\}$. Compute the total squared error $E(c_0,c_1,c_2)$ given by Equation (5.2) and compare with the error term $E(m,b)$ found in Problem 5.4 (a). You may modify the code provided for this problem.

5.6. Derive the matrix Equation (5.3).

5.7. Rederive the matrix Equation (5.3) using the interpolation approach of Subsection 5.1.2.

5.8. Derive the 4×4 system of equations required to fit the model

$$f(x;c_0,c_1,c_2,c_3) = c_0 + c_1x + c_2x^2 + c_3x^3$$

Put this system into matrix form and compare with the result in Equation (5.3). Can you extend this pattern to write down the matrix form of the least squares equations for the model

$$f(x) = c_0 + c_1x + c_2x^2 + \cdots + c_9x^9$$

No exact derivation is required for this last part.

5.9. Reread Section 5.1.4 and propose a formula for a polynomial of degree 3, $P_3(x)$, such that the interpolation conditions $P_3(x_1) = y_1$, $P_3(x_2) = y_2$, $P_3(x_3) = y_3$, and $P_3(x_4) = y_4$ are satisfied.

5.10. Find the linear spline through points $\{(0,0),(2,1),(3,-2),(5,2)\}$.

5.11. Apply MATLAB's cubic spline routine to the Fort Collin's daily temperature data provided on odd days in September 2002; see Table 5.2. Use this model to predict the temperature on the even days. Compare your predictions with the actual temperature values provided in Table 5.3. Plot your results.

Day	5pm temperature
1	87.0
3	83.3
5	89.7
7	82.3
9	65.6
11	59.8
13	65.5
15	77.1
17	69.1
19	63.8
21	51.1

TABLE 5.2: Fort Collins' temperatures on odd days in September, 2002. All temperatures recorded at 5pm. Use this data to build spline model.

Day	5pm temperature
2	80.3
4	87.1
6	90.1
8	79.2
10	64.4
12	60.2
14	71.2
16	73.9
18	58.4
20	73.5

TABLE 5.3: Fort Collins' temperatures on even days in September, 2002. All temperatures recorded at 5pm. Use this data to test spline model.

5.12. Consider the data model

$$f(x) = a\sigma(bx + c)$$

where the *sigmoidal* function is given by

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

Given a data set of input output pairs (x_i, y_i) write down the least squares optimization criterion for the unknown model parameters a, b, c. Show that the

resulting system for a, b, c is nonlinear. Describe briefly how would you solve for them (without actually doing it).

5.13. Using the model in Exercise 5.12 compute a, b and c for the sample data set $\{(-.3, -.5), (-.1, -.2), (0, .1), (.2, .3), (.6, .6)\}$. Is this a good model for the data? How might you improve it?

CHAPTER 6

Modeling with Discrete Dynamical Systems

6.1 INTRODUCTION

One of the most exciting areas of modeling concerns predicting temporal evolution. The main question that is posed in this setting is how do variables of interest change over time? This type of problem is everywhere to be found, for example in areas as diverse as science, engineering and finance. Prediction means that given the values of the variables at a certain instant of time we can predict, i.e. compute their values at any future time. A system of equations that allows such a prediction is called a *Dynamical System*.

In this chapter we consider discrete dynamical systems. The mathematical assumption is that the time variable n is incremented discretely and corresponds to the integers $\{0, 1, 2, 3, 4, \dots\}$. The value of a variable x of interest is then a sequence $\{x_0, x_1, x_2, x_3, x_4, \dots\}$. Now the problem of modeling is to determine an equation of the form

$$x_{n+1} = x_n + \Delta x_n$$

and this is done by estimating how the variable x_n changes as n is incremented from time n to time $n + 1$.

We develop this topic along the following four complementary lines:

- numerical solutions,
- analytical solutions,
- qualitative behavior,
- modeling techniques.

As the terminolgy suggests, numerical approaches to difference equations will involve direct computation of these sequences via computer. In contrast, analytical solution methods seek closed form solutions; these are available only in limited circumstances.

Qualitative approaches are analytical as well as numerical approaches to determine the qualitative behaviour of the solutions in the long run. The questions addressed are: do the solutions go off to infinity, do they approach a finite value, will they oscillate or behave more complicated? Another question of interest is the sensitivity of solutions to variation of parameters. A change in the qualitative behaviour when a parameter is varied is called a bifurcation.

The topic of modeling will treat empirical and qualitative approaches for constructing difference equations. We will consider the development of models

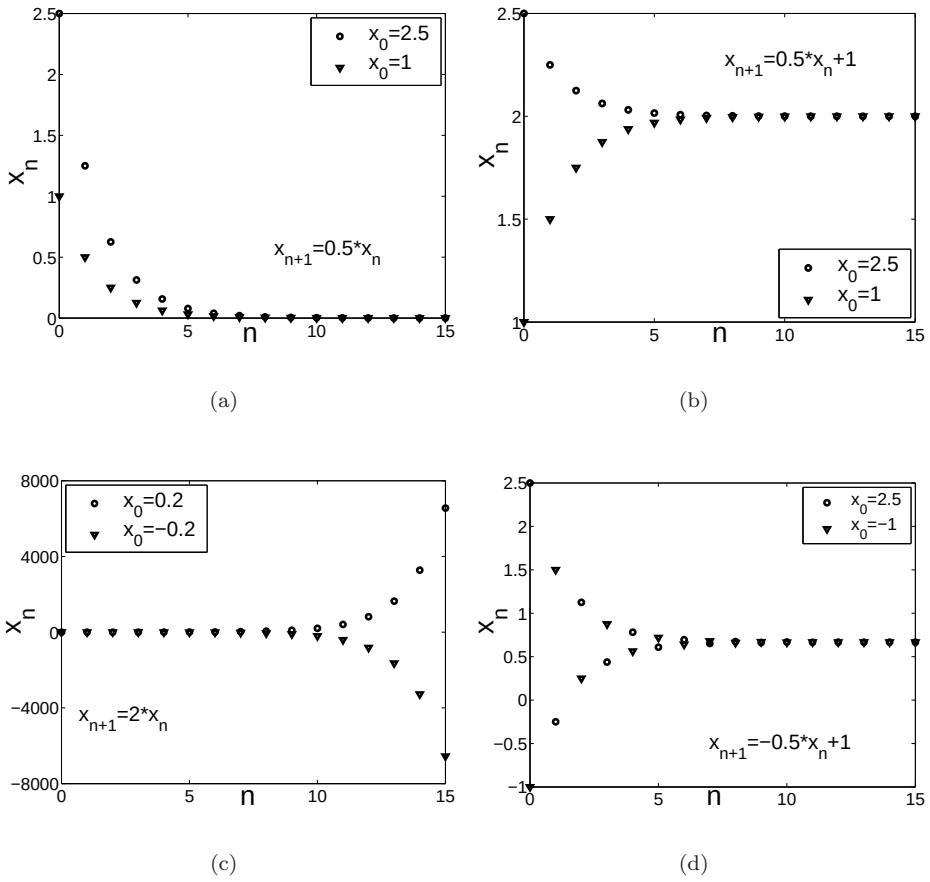


FIGURE 6.1: Comparison of the numerical solutions for some simple difference equations.

based on the qualitative approaches presented in Chapter 2 as well as the more quantitative data fitting approaches of Chapter 5.

A simple but nevertheless important difference equation is the equation

$$x_{n+1} = ax_n + b. \tag{6.1}$$

If an initial value x_0 is fixed the solution is determined for all n,

$$x_1 = ax_0 + b, \;\; x_2 = ax_1 + b, \;\; x_3 = ax_2 + b, \;\; \ldots.$$

Numerically simulated solutions of (6.1) for various values of the parameters a and b are shown in Figure 6.1. In Figure 6.1 (a) we see that the solutions decay to zero while in Figure 6.1 (b) they tend to the value 2. In Figure 6.1 (c) the initial values are close to zero. Both solutions remain close to zero for a while, but eventually they split apart and tend to $\pm\infty$. In Figure 6.1 (d) the solutions tend to $x \approx 0.7$. Here the solutions alternate between values above and below 0.7 when approaching

this value. Thus, a noticeable feature for all of these solutions is the long term behavior. Qualitatively we say the solution either blows up or approaches a finite limiting value.

EXAMPLE 6.1 Discrete Compound of Interest

Interest rates for loans or saving accounts are normally fixed on an annual basis, however the compounding scheme typically applies the interest charges monthly. Suppose you purchase something for a certain amount of $\$a_0$ and charge it to your credit card that carries an annual interest rate of $r\%$. Let a_n be the accumulated debt after n months. In Section 6.2.2 we will see that a_n satisfies the difference equation

$$a_{n+1} = (1 + \frac{r}{1200})a_n - p, \tag{6.2}$$

where p is your monthly payment. Equation (6.2) has the form of Equation (6.1). By solving this equation you can answer questions such as: when is a loan a_0 paid off given a certain monthly payment p, or what should the monthly payment be in order that the loan is paid off after a prescribed amount of time?

Equation (6.1) is called a *linear* first order difference equation. It is linear because the right hand side is a linear function of x_n. It is of first order because only one time step is involved. The simplest *nonlinear* first order difference equation is

$$x_{n+1} = ax_n + bx_n^2. \tag{6.3}$$

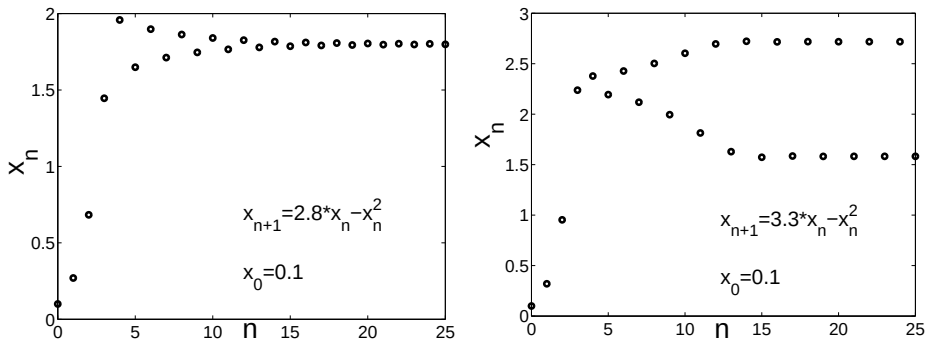
In Figure 6.2 numerical solutions of (6.3) are shown for $b = -1$ and two different values of a. In Figure 6.2 (a) we see approach to a limiting value as in Figure 6.1 (d). In contrast in Figure 6.2 (b) the solution eventually alternates between the values 1.6 and 2.7. This type of behavior cannot be found in solutions of linear equations. The solutions of nonlinear equations show a much richer variety of behaviors. Another important difference is that linear equations admit closed form solutions whereas nonlinear equations typically cannot be solved analytically.

EXAMPLE 6.2 Population Growth

Discrete dynamical systems are widely used in population modeling, in particular for species which have no overlap between successive generations and for which births occur in regular, well-defined 'breeding seasons'. Let p_n be the average population of a species between times $n\tau$ and $(n+1)\tau$. The time step τ depends on the particular species and can range from an hour to several years. For example many species of bamboo grow vegetatively for 20 years before flowering and then dying.

In population dynamics one constructs a model for the change $\Delta p_n = p_{n+1} - p_n$. The simplest model is a linear model, $\Delta p = kp_n + \beta$, where k is called the reproduction rate and β models constant immigration ($\beta > 0$) or emigration ($\beta < 0$). The difference equation that results from this model assumption,

$$p_{n+1} - p_n = kp_n + \beta,$$



(a) (b)

FIGURE 6.2: Numerical solutions for equation (6.4).

is again of the form of Equation (6.1).

Competition for resources usually leads to nonlinear difference equations. We will see that the simplest model that takes competition into account leads to the equation

$$p_{n+1} = rp_n - p_n^2, \tag{6.4}$$

which is of the form of Equation (6.3). Equation (6.4) is known in the literature as *logistic map*. Its prominent feature are very complicated, so called chaotic solutions in certain ranges of the parameter r.

The equation

$$x_{n+2} + 2x_{n+1} + 3x_n = \cos(n)$$

is an example of a linear second order difference equation. We shall see that this type of equation always can be transformed to a linear system of two first order equations. The general form of such a system is

$$\begin{aligned} x_{n+1} &= ax_n + by_n + f_n, \\ y_{n+1} &= cx_n + dy_n + g_n, \end{aligned}$$

where f_n, g_n are known sequences. If the right hand sides are replaced by nonlinear functions we have a nonlinear system, for instance

$$\begin{aligned} x_{n+1} &= ax_n - bx_n^2 - cx_n y_n, \\ y_{n+1} &= dx_n - ey_n^2 - fx_n y_n. \end{aligned}$$

This system is used in population modeling as a model for the population growth of two interacting species. The terms $-bx_n^2$ and $-ey_n^2$ model competition within each of the two species whereas the terms $-cx_n y_n$ and $-fx_n y_n$ model competition between the species.

6.2 LINEAR FIRST ORDER DIFFERENCE EQUATIONS

6.2.1 Analytical Solutions

Possibly the simplest nontrivial difference equation has the form

$$x_{n+1} = ax_n. \tag{6.5}$$

This equation has the special solution $x_n = 0$. Since it is constant it is said to be an equilibrium solution. The value of the constant, $\overline{x} = 0$, is called an equilibrium value or shortly an equilibrium. The solutions for initial values $x_0 \neq 0$ are found by implementing the iteration,

$$\begin{aligned} x_1 &= ax_0 \\ x_2 &= ax_1 = a^2x_0 \\ x_3 &= ax_2 = a^3x_0 \\ &\vdots \\ x_n &= a^nx_0. \end{aligned} \tag{6.6}$$

From (6.6) we can easily infer how the qualitative behavior of x_n depends on a: if $|a| > 1$ then x_n goes off to infinity (the equilibrium is said to be unstable), whereas if $|a| < 1$ then x_n tends to 0 (the equilibrium is said to be stable). This explains the behavior of the numerical solutions of Figures 6.1 (a) and (c). Note also that if $a > 0$ then x_n has the same sign as x_0 for all n. In contrast if $a < 0$ the solution alternates between positive and negative values.

The cases $a = 1$ and $a = -1$ are special. If $a = 1$ we have $x_n = x_0$ for all n, hence every x is an equilibrium. If $a = -1$ the solution $x_n = (-1)^n x_0$ flips back and forth between x_0 and $-x_0$.

A more general equation is the following,

$$x_{n+1} = ax_n + b. \tag{6.7}$$

An equilibrium is determined by $x_{n+1} = x_n = \overline{x}$ for all n, hence

$$\overline{x} = a\overline{x} + b \;\Rightarrow\; \overline{x} = \frac{b}{1-a},$$

where we assume $a \neq 1$. We can transform (6.7) to (6.5) by subtracting the equilibrium. Set

$$y_n = x_n - \overline{x}.$$

Then

$$\begin{aligned} y_{n+1} &= x_{n+1} - \overline{x} = ax_n + b - \overline{x} \\ &= a(y_n + \overline{x}) + b - \overline{x} = ay_n, \end{aligned}$$

and so $y_n = a^n y_0$. The solution of (6.7) is found by transforming y_n back to x_n,

$$x_n = y_n + \overline{x} = ax^n(x_0 - \overline{x}) + \overline{x} = a^n(x_0 - \frac{b}{1-a}) + \frac{b}{1-a}. \tag{6.8}$$

Again the value of $|a|$ determines whether x_n goes off to infinity or approaches $\overline{x}$, and the sign of a determines whether $x_n - \overline{x}$ alternates or has a constant sign.

EXAMPLE 6.3

The equation

$$x_{n+1} = \frac{1}{2}x_n + 1$$

is of the form of (6.7). The equilibrium is

$$\overline{x} = \frac{b}{1-a} = 2.$$

Since $a = 1/2 < 1$ and $a > 0$ the solutions approach the equilibrium 2 and the sign of $x_n - 2$ is the same for all n. This explains the behavior of the numerical solutions shown in Figure 6.1 (d).

A more general form than (6.7) is provided by the equation

$$x_{n+1} = ax_n + b_n, \tag{6.9}$$

where b_n is a given sequence. This equation is said to be nonhomogeneous due to the presence of the b_n term. If $b_n = 0$ for all n, (6.9) simplifies to (6.5) and then the equation is called homogeneous. We refer to (6.5) as the homogeneous equation associated with (6.9). In the special case in which $b_n = b = \text{const}$ we were able to transform the nonhomogeneous equation to its associated homogeneous equation, but if b_n varies with n this is no longer possible.

> DEFINITION **4**. A one parameter family of solutions of (6.9) is an expression $x_n = x_n(c)$ that depends linearly on a parameter c and satisfies (6.9) identically in n and c. A particular solution is a solution that contains no free parameters. A one parameter family of solutions is a general solution if for every particular solution p_n we can find a value $\overline{c}$ of c such that $p_n = x_n(\overline{c})$ for all n.

Consider now the *difference* $h_n = q_n - p_n$ of two particular solutions q_n and p_n of (6.9). The computation

$$\begin{aligned} h_{n+1} &= q_{n+1} - p_{n+1} = (aq_n + b_n) - (ap_n + b_n) \\ &= a(q_n - p_n) = ah_n \end{aligned}$$

shows that h_n is a solution of the homogeneous equation (6.5). Since $h_0 = q_0 - p_0$ it follows from (6.6) that $h_n = (q_0 - p_0)a^n$ and so,

$$q_n = (q_0 - p_0)a^n + p_n.$$

If we assume p_n is a known particular solution, this equation allows to find any other particular solution q_n from its initial value q_0. Thus if we write

$$x_n = ca^n + p_n, \tag{6.10}$$

and consider c as parameter, the solution q_n is simply obtained by setting $c = q_0 - p_0$. We therefore have proved the following theorem.

THEOREM **5**. Let p_n be a particular solution of the nonhomogeneous equation

$$x_{n+1} = ax_n + b_n.$$

Then the family

$$x_n = ca^n + b_n$$

is a general solution.

Note that there is no unique general solution. For instance,

$$x_n = ca^n + (p_n + 5a^n)$$

is also a general solution because $p_n + 5a^n$ is another particular solution.

EXAMPLE 6.4

Verify that $p_n = -n - 1$ is a particular solution of

$$x_{n+1} = 3x_n + 2n + 1.$$

Solution To test that an expression is a solution of a difference equation we just have to plug it into the equation and check if both sides are the same. Now the left hand side evaluates to

$$p_{n+1} = -(n+1) - 1 = -n - 2,$$

and the right hand side to

$$3p_n + 2n + 1 = 3(-n-1) + 2n + 1 = -n - 2.$$

Since these are the same we have verified that p_n is a solution. It is a particular solution because it does not depend on parameters.

EXAMPLE 6.5

Find the general solution of

$$x_{n+1} = 3x_n + 2n + 1$$

and the particular solution that satisfies $x_0 = 1$.

Solution Form Example 6.4 we know that $p_n = -n - 1$ is a particular solution. Since $a = 3$ the general solution is

$$x_n = c3^n - n - 1.$$

To find the particular solution asked for we evaluate at $n = 0$,

$$x_0 = c - 1 = 1.$$

It follows that $c = 2$, hence

$$x_n = 2\,3^n - n - 1$$

is the solution with $x_0 = 1$.

---

b_n	form of particular solution	conditions
(6.11)	$p_n = (A_0 + A_1 n + \cdots + A_M n^M) b^n$	$b \neq a$
	$p_n = n(A_0 + A_1 n + \cdots + A_M n^M) b^n$	$b = a$
(6.12)	$p_n = (A_0 + A_1 n + \cdots + A_M n^M) b^n \cos(kn)$ $+ (B_0 + B_1 n + \cdots + B_M n^M) b^n \sin(kn)$	$k \neq 0, \pi$

TABLE 6.1: Solution forms p_n for b_n given by Equations (6.11) and (6.12)

To complete the solution of the nonhomogeneous equation (6.9) we need to find a particular solution. For general terms b_n this can be a complicated task. However there is a method that applies always if b_n is a combination of powers of n (n^0, n^1, n^2 etc.), trigonometric functions of n, and powers b^n. This method is called method of undetermined coefficients.

Method of undetermined coefficients. Assume b_n has one of the following forms,

$$b_n = (c_0 + c_1 n + \ldots + c_M n^M) b^n, \tag{6.11}$$

where $c_M \neq 0$, or

$$\begin{aligned} b_n &= (c_0 + c_1 n + \cdots + c_M n^M) b^n \cos(kn) \\ &+ (d_0 + d_1 n + \ldots + d_M n^M) b^n \sin(kn), \end{aligned} \tag{6.12}$$

where at least one of c_M or d_M is nonzero. The coefficients b, k and c_j, d_j $(0 \leq j \leq M)$ are assumed to be given numbers. It can be shown that if b_n is as in (6.11) or (6.12), then there exists a unique particular solution p_n of the form as summarized in Table 6.2.1. To find the values of the coefficients A_j, B_j $(0 \leq j \leq M)$, one sets up a trial form for p_n according to the table with initially undetermined values of the coefficients, substitutes the trial form into the difference equation, and determines the values of the coefficients from the condition that p_n be a solution. If b_n is a linear combination of several terms of the form of (6.11) or (6.12), with different values of b or (b, k), each of them can be treated separately and the results are added up.

EXAMPLE 6.6

Find a particular solution of

$$x_{n+1} = 3x_n + 2n + 1.$$

Solution Here $b_n = 2n + 1$ is of the form (6.11) with $b = 1$ and $M = 1$. Thus we use $p_n = A + Bn$ as trial form and substitute this into the difference equation to obtain,

$$A + B(n+1) = 3(A + Bn) + 2n + 1,$$

or

$$(2A - B + 1) + (2B + 2)n = 0.$$

This equation holds for all n if A and B satisfy the equations $2A - B = -1$ and $2B = -2$. The solution is $A = B = -1$, hence

$$p_n = -n - 1.$$

EXAMPLE 6.7

Find a particular solution of

$$x_{n+1} = -x_n + \cos 2n.$$

Solution Substitution of the trial form $p_n = A\cos 2n + B\sin 2n$ into the difference equation yields

$$A\cos 2(n+1) + B\sin 2(n+1) = -A\cos 2n - B\sin 2n + \cos 2n.$$

We apply the formulae for $\cos(\alpha+\beta)$ and $\sin(\alpha+\beta)$ to the terms on the left hand side and then rearrange the equation as

$$[A(1+\cos 2) + B\sin 2 - 1]\cos 2n + [-A\sin 2 + B(1+\cos 2)]\sin 2n = 0.$$

This equation holds for all n if the terms in both brackets vanish. Setting these terms equal to zero gives the following system of equations for A and B,

$$\begin{aligned} A(1+\cos 2) + B\sin 2 - 1 &= 0 \\ -A\sin 2 + B(1+\cos 2) &= 0, \end{aligned}$$

with the solution

$$A = \frac{1}{2}, \quad B = \frac{\sin 2}{2(1+\cos 2)}.$$

Hence the particular solution is

$$p_n = \frac{1}{2}\cos 2n + \frac{\sin 2\cos 2n}{2(1+\cos 2)} = \frac{\cos 2n + \cos 2(n-1)}{2(1+\cos 2)}.$$

EXAMPLE 6.8

Find a particular solution of

$$x_{n+1} = x_n/2 + n(1/2)^n.$$

Solution Here $a = b = 1/2$, so the trial function is $p_n = n(An+B)(1/2)^n$. Again we substitute p_n into the difference equation,

$$[A(n+1)^2 + B(n+1)](1/2)^{n+1} = (An^2 + Bn)(1/2)^n/2 + n(1/2)^n.$$

We multiply this equation by 2^{n+1} and rearrange terms as

$$2(A-1)n + (A+B) = 0.$$

Thus $A = -B = 1$ and the particular solution is

$$p_n = (n^2 - n)(1/2)^n.$$

6.2.2 Modeling Examples

(A) Savings Accounts and Loans

Savings Accounts. Assume you open a savings account at an annual interest rate of $r\%$ and with monthly compound of interest. Let a_n be the dollar amount on the account at the end of month n after the opening date. The amount at the end of month $n+1$ is

$$a_{n+1} = a_n + i_n + p_n,$$

where p_n is the total deposit (withdrawal if $p_n < 0$) and i_n is the interest earned,

$$i_n = \left(\frac{r}{100}\frac{1}{12}\right) a_n.$$

Thus a_n satisfies the nonhomogeneous, linear first order difference equation,

$$a_{n+1} = ka_n + p_n, \tag{6.13}$$

where

$$k = 1 + \frac{r}{1200}.$$

If $p_n = p = const$ we know the solution already (Equation (6.8) with $a = k$, $b = p$, $x_n = a_n$),

$$a_n = k^n(a_0 + \frac{p}{k-1}) - \frac{p}{k-1} = k^n a_0 + \frac{(k^n - 1)p}{k-1}. \tag{6.14}$$

EXAMPLE 6.9

After graduating from High School Peter works for four years. During this time he deposits each month \$1000 on a savings account at an annual interest rate of 5% (no initial deposit). The next four years Peter spends on College. During this time he withdraws each month an amount of $\$p_w$ from his savings account so that at the end of the four years the balance is zero again. Find p_w and the total interest earned.

Solution Letting p be the the monthly deposit, the accumulated amount on Peter's savings account after the first four years is

$$a_{48} = \frac{(k^{48} - 1)p}{k - 1}.$$

After the second four years this has evolved into

$$a_{96} = k^{48}a_{48} - \frac{(k^{48} - 1)p_w}{k - 1} = \frac{k^{48} - 1}{k - 1}(k^{48}p - p_w).$$

Solving the equation $a_{96} = 0$ for p_w gives $p_w = k^{48}p$. With $p = \$1000$ and $k = 1+5/1200$ this evaluates to $p_w = \$1220.89$. The total interest earned is $48(p_w - p) = \$10,602.72$.

Loans. Equation (6.13) also holds for loans. In this case a_0 is the amount borrowed and a_n is the amount owed after n months. The term $-p_n > 0$ is the monthly payment. For constant monthly payment p the difference equation for a_n is

$$a_{n+1} = ka_n - p, \tag{6.15}$$

with the solution

$$a_n = ka_0^n - \frac{(k^n - 1)p}{k - 1}.$$

Note that (6.15) has an unstable equilibrium $\overline{a} = p/(k - 1)$. If $a_0 > \overline{a}$ the solution grows without bound when n increases. While for savings accounts this may be desirable, it is certainly not tolerable for loans.

The term of a loan is the time N (in months) when the loan is paid off. Setting $a_N = 0$ leads to a linear relation between monthly payment and initial debt,

$$p = \frac{k^N(k - 1)}{k^N - 1}a_0. \tag{6.16}$$

EXAMPLE 6.10

You decide to purchase a home with a mortgage at 6% annual interest and with a term of 30 years. For $k = 1 + 6/1200 = 1.005$ and $N = 360$ the factor

$$R = \frac{k^N(k - 1)}{k^N - 1}$$

in Equation (6.16) is $R = 0.00600$. If the house costs $a_0 = \$200,000$, the monthly payment is $p = Ra_0 = \$1,199.10$. On the other hand, if your income restricts your monthly payment to a maximum of $p_m = \$1000$, the maximal amount you can spend for the house is $p_m/R = \$166,791.61$.

If p, a_0 and k are fixed, the equation (6.16) may be considered as an equation for the term N. Writing $k^N = e^{N \ln k}$, Equation (6.16) can be rewritten as

$$e^{N \ln k} = \frac{p}{p - (k-1)a_0},$$

hence

$$N = \frac{-\ln[1 - (k-1)a_0/p]}{\ln k}. \tag{6.17}$$

Note however that the right hand side of (6.17) needs not to be an integer. Nevertheless it can be used to estimate N and then to improve p or a_0. For example, assume you need $\$200,000$ and you want your payment to be close to, but not above $\$1500$. With $r = 8\%$, $a_0 = 200,000$ and $p = 1500$, (6.17) evaluates to $N = 330.68$. If this is rounded up to $N = 331$, Equation (6.16) gives $p = \$1499.60$.

In our last example on savings accounts and loans we have to solve the nonhomogeneous equation (6.9) with nonconstant b_n.

EXAMPLE 6.11

An employee starts her position at the age of 25 with an annual salary of $\$40,000$. She deposits each month 8% of her monthly salary on a retirement savings account. The salary increases by 3% each year and the annual interest rate of her retirement savings account is 6%. What is the accumulated amount when she retires at the age of 65?

Solution Let A_m be the accumulated amount on the retirement savings account at the end of year m and let $a_{m,n}$ be the accumulated amount in month n of year $m+1$, that is,

$$a_{m,0} = A_m, \quad a_{m,12} = A_{m+1}.$$

The amount $a_{m,n}$ satisfies difference equation

$$a_{m,n+1} = k_r a_{m,n} + k_p s_m, \tag{6.18}$$

where $k_r = 1 + 6/1200 = 1.005$, $k_p = 8/1200$ and s_m is the salary in year $m+1$. The salary satisfies the homogeneous difference equation

$$s_{m+1} = k_s s_m,$$

with $k_s = 1 + 3/100 = 1.03$ and $s_0 = 40,000$, hence

$$s_m = k_s^m s_0.$$

The solution of (6.18) is

$$a_{m,n} = k_r^n a_{m,0} + \frac{(k_r^n - 1)k_s^m k_p s_0}{k_r - 1}.$$

Evaluating this at $n = 12$ yields

$$A_{m+1} = k_a A_m + f k_s^m, \tag{6.19}$$

where

$$f = \frac{(k_r^{12} - 1)k_p s_0}{k_r - 1} = \$3289.48, \quad k_a = k_r^{12} = 1.0616778.$$

It remains to solve the nonhomogenous difference equation (6.19). By using the method of undetermined coefficients a particular solution can be determined to be $p_m = fk_s^m/(k_s - k_a)$. The solution with initial value $A_0 = 0$ then is

$$A_m = \frac{(k_s^m - k_a^m)f}{k_s - k_a}.$$

For $m = 40$ this evaluates to $A_{40} = 799,106.39$. Hence the employee starts her retirement with an amount of $\$799,106.39$ on her retirement savings account.

(B) Cooling and Heating

Newton's law of cooling states that the rate of change of the temperature of an object is proportional to the difference of the temperature of the object and its surrounding. Let $\Delta T_n = T_{n+1} - T_n$ be the change in temperature of the object over a time interval τ, typically $\tau = 1$ hour. According to Newton's law of cooling we have

$$\Delta T_n \propto R_n - T_n,$$

or

$$\Delta T_n = k(R_n - T_n),$$

where R_n is the surrounding temperature. Since we know that temperature decreases if $R_n > T_n$ it follows that $k > 0$. The difference equation that arises from this model is

$$T_{n+1} = T_n + k(R_n - T_n).$$

If $R_n = R = \text{const}$ this equation is again of the form (6.7) with solution

$$T_n = (1 - k)^n(T_0 - R) + R.$$

Note that the equilibrium solution is $T_n = R$ as expected. The equilibrium is stable if $0 < k < 2$. However if $1 < k < 2$ the temperature would oscillate about the surrounding temperature which does not make sense physically, hence $0 < k < 1$.

EXAMPLE 6.12

A murder victim is discovered in an office building that is maintained at 68 degrees F. Given the medical examiner found the body temperature to be 88 degrees F at 8am and that one hour later the body temperature was 86 degrees F, at what time was the crime committed?

Solution Setting $T_0 = 98.6$ (where 0 is the time the crime was committed) and $R = 68$ we obtain

$$T_n = 68 + 30.6(1 - k)^n.$$

If we define n_1 as the time the body was observed initially by the medical examiner and the time one hour later as $n_1 + 1$ we have the equations

$$T_{n_1} = 88 = 68 + 30.6(1-k)^{n_1},$$

and

$$T_{n_1+1} = 86 = 68 + 30.6(1-k)^{n_1+1}.$$

These two equations may be solved to give $k = 1/10$ and $n_1 = 4.036$. So the crime was committed just before 4am.

6.3 LINEAR SECOND ORDER EQUATIONS

6.3.1 Homogeneous Equations

We begin by considering the second order linear homogeneous difference equation

$$x_{n+2} + \alpha x_{n+1} + \beta x_n = 0 \tag{6.20}$$

It is readily verified that this equation has solutions of the form

$$x_n = \lambda^n$$

Upon substitution into Equation (6.20) we obtain the *auxiliary* equation

$$\lambda^2 + \alpha\lambda + \beta = 0$$

This quadratic equation has solutions that break down into three cases: i) both solutions real and distinct, ii) one real double solution, and iii) a pair of complex solutions as

$$\lambda_\pm = \frac{-\alpha \pm \sqrt{\alpha^2 - 4\beta}}{2}$$

Case i: $\alpha^2 - 4\beta > 0$. Two real roots.
In this case

$$\lambda_+ = \frac{-\alpha + \sqrt{\alpha^2 - 4\beta}}{2}$$

and

$$\lambda_- = \frac{-\alpha - \sqrt{\alpha^2 - 4\beta}}{2}$$

are both real and the solution is

$$h_n = c_1(\lambda_+)^n + c_2(\lambda_-)^n$$

Since the equation is linear we know that the superposition of solutions is again a solution. Notice that there are now two free parameters c_1 and c_2 to accommodate the two initial conditions x_0 and x_1 required for a second order difference equation. Notice also that $h_n \to 0$ for $n \to \infty$ if $\lambda_\pm| < 1$, but in general $|h_n| \to \infty$ if $|\lambda_+| > 1$.

EXAMPLE 6.13

$$x_{n+2} = x_{n+1} + x_n$$

The auxiliary equation is now

$$\lambda^2 - \lambda - 1 = 0$$

The solutions to this quadratic are

$$\lambda_{\pm} = \frac{1 \pm \sqrt{5}}{2}$$

Thus, the general solution to the homogeneous problem is

$$h_n = c_1\left(\frac{1+\sqrt{5}}{2}\right)^n + c_2\left(\frac{1-\sqrt{5}}{2}\right)^n$$

If we select $h_0 = h_1 = 1$ we have the *Fibonocci sequence* $\{1, 1, 2, 3, 5, 8, 13, \dots\}$. Employing this pair of initial conditions it is easily shown that the particular solution is

$$h_n = \left(\frac{\sqrt{5}+1}{2\sqrt{5}}\right)\left(\frac{1+\sqrt{5}}{2}\right)^n + \left(\frac{\sqrt{5}-1}{2\sqrt{5}}\right)\left(\frac{1-\sqrt{5}}{2}\right)^n$$

You might impress your friends by telling them the 50th number in this sequence $h_{50} = 20365011074$. It is also apparent that these numbers increase exponentially fast.

Case ii: $\alpha^2 - 4\beta = 0$. One real (double) root.

In this case

$$\lambda_+ = \lambda_- = -\frac{\alpha}{2}$$

so we only have one solution while we require two for the general solution of a second order difference equation.

It is not hard to verify that in this instance the second solution is actually

$$x_n = n(\frac{-\alpha}{2})^n$$

. (See Exercise 6.15). Now the general solution to this homogeneous equation is

$$h_n = c_1\big(-\frac{\alpha}{2}\big)^n + c_2 n\big(-\frac{\alpha}{2}\big)^n$$

EXAMPLE 6.14

$$x_{n+2} + 2x_{n+1} + x_n = 0$$

The auxiliary equation is

$$\lambda^2 + 2\lambda + 1$$

which has the solution $\lambda = -1$. Thus, the general solution to this homogeneous problem is

$$h_n = c_1(-1)^n + c_2 n(-1)^n$$

Case iii: $\alpha^2 - 4\beta < 0$. Two complex roots.

The solution to the auxiliary equation is again

$$\lambda_{\pm} = \frac{-\alpha \pm \sqrt{\alpha^2 - 4\beta}}{2}$$

Based on the fact $\alpha^2 - 4\beta < 0$ we rewrite this as

$$\lambda_{\pm} = \frac{-\alpha}{2} \pm i\frac{\sqrt{4\beta - \alpha^2}}{2}$$

where $i = \sqrt{-1}$.[1]

We could now write the solution

$$h_n = c_1\left(\frac{-\alpha}{2} + i\frac{\sqrt{4\beta - \alpha^2}}{2}\right)^n + c_1\left(\frac{-\alpha}{2} - i\frac{\sqrt{4\beta - \alpha^2}}{2}\right)^n$$

but this form would not provide much insight. Instead we employ Demoivre's theorem that states

$$\exp(inx) = \cos(nx) + i\sin(nx)$$

To exploit this formula we need to recall that each solution to the auxiliary equation can be written in its complex polar form

$$z = x + iy = r\exp(i\theta)$$

where $x = r\cos\theta$ and $y = r\sin\theta$. Thus, we take

$$x = \frac{-\alpha}{2}, \quad \text{and} \quad y = \frac{\sqrt{4\beta - \alpha^2}}{2}$$

To compute the polar form we need r and θ. Recall

$$r^2 = x^2 + y^2$$

so

$$\begin{aligned} r^2 &= (\frac{-\alpha}{2})^2 + (\frac{\sqrt{4\beta - \alpha^2}}{2})^2 \\ &= \beta \end{aligned}$$

So

$$r = \sqrt{\beta}$$

The angle satisfies

$$\tan\theta = \frac{y}{x} = \frac{\sqrt{4\beta - \alpha^2}}{-\alpha}$$

In polar form, the solution is

$$h_n = c_1 r^n \exp(in\theta) + c_2 r^n \exp(-in\theta)$$

[1] Unlike the previous cases we now assume familiarity with basic complex numbers.

The associated real form is

$$h_n = r^n(c_1 \cos(n\theta) + c_2 \sin(n\theta),$$

where we have used the facts that $\exp(in\theta) = \cos(n\theta) + i \sin(n\theta)$ and that the real and imaginary parts of a complex solution are real solutions (see problems). The form of the solution tells that $h_n \to 0$ for $n \to \infty$ if $r < 1$ and $|h_n| \to \infty$ if $r > 1$. If $r = 1$ the solution remains bounded, but does not approach zero.

EXAMPLE 6.15

Find the general solution to the homogeneous difference equation

$$x_{n+2} + 2x_{n+1} + 5x_n = 0$$

The auxiliary equation gives the solutions

$$\lambda_{\pm} = -2 \pm i$$

If we write these in polar form we have

$$h_n = 5^{n/2}(c_1 \exp(in\theta) + c_2 \exp(-in\theta))$$

where $\tan\theta = 1/2$. The associated real valued form is

$$h_n = 5^{n/2}(c_1 \cos(n\theta) + c_2 \sin(n\theta)).$$

6.3.2 The Cobweb Model Revisited

Consider a supply curve

$$p = m_s q + b_s$$

and a demand curve

$$p = m_d q + b_d$$

Here we derive a formula for the values (q_n, p_n) that are the iterations along the supply and demand curves that either converge to an economic equilibrium or spiral out of control. Let the starting point on the demand curve be (q_0, p_0). The next iteration is then given by

$$(q_1, p_1) = (\frac{p_0 - b_s}{m_s}, p_0)$$

Similarly,

$$(q_2, p_2) = (q_1, m_d q_1 + b_d),$$

$$(q_3, p_3) = (\frac{p_2 - b_s}{m_s}, p_2)$$

and

$$(q_4, p_4) = (q_3, m_d q_3 + b_d)$$

Thus, we have established the following pattern:

$$(q_{2n}, p_{2n}) = (q_{2n-1}, m_d q_{2n-1} + b_d)$$

and

$$(q_{2n+1}, p_{2n+1}) = (\frac{p_{2n} - b_s}{m_s}, p_{2n})$$

It is now possible to create a second order difference equation for both q_n and p_n. Since

$$q_{2n+1} = \frac{p_{2n} - b_s}{m_s}$$

it follows, upon substituting for p_{2n} that

$$q_{2n+1} = \frac{(m_d q_{2n-1} + b_d) - b_s}{m_s}$$

or,

$$q_{2n+1} = \frac{m_d}{m_s} q_{2n-1} + \frac{b_d - b_s}{m_s}. \tag{6.21}$$

A Nonhomogeneous Second Order Equation. The equation (6.21) is of the form

$$q_{2n+1} = \alpha q_{2n-1} + \beta$$

This is a nonhomogeneous second order difference equation whose general solution is, as in the first order case, given by

$$x_n = h_n + p_n,$$

where h_n is the general solution of the associated homogeneous equation and p_n is a particular solution of the nonhomogeneous equation.

The associated homogeneous equation is

$$q_{2n+1} = \alpha q_{2n-1}$$

and has the auxiliary equation

$$\lambda^2 = \alpha$$

so the general solution to the homogeneous problem is

$$h_n = c_1 \alpha^{n/2} + c_2 (-\alpha^{1/2})^n$$

As the nonhomogeneous term is a constant we first search for a particular solution of the form $p_n = A$. This must be an equilibrium solution, if it exists. Solving for A then gives

$$A = \alpha A + \beta$$

or

$$A = \frac{\beta}{1 - \alpha}$$

In terms of the original variables of the supply and demand problem, $\alpha = m_d/m_s$, $\beta = (b_d - b_s)/m_s$, the general solution to the nonhomogeneous equation now becomes

$$q_n = c_1\left(\frac{m_d}{m_s}\right)^{n/2} + c_2\left(-(\frac{m_d}{m_s})^{1/2}\right)^n + \frac{b_d - b_s}{m_s - m_d}$$

It is clear from our previous work that this equation will only converge if

$$|\frac{m_d}{m_s}| < 1$$

Note also that if this condition holds then the quantity supplied converges,

$$q_n \to \frac{b_d - b_s}{m_s - m_d}$$

and approaches the market equilibrium.

6.4 NONLINEAR DIFFERENCE EQUATIONS AND SYSTEMS IN POPULATION MODELING

In this section we will consider a sequence of modifications of a population model that characterize the modeling process and illustrate how including or deleting terms in equations can have dramatic effects on the predictive powers of a model.

The simplest model for population growth makes the assumption that there is no competition for resources such as nutrients or habitat. This exponential growth is readily captured by the simple difference equation

$$p_{n+1} - p_n = \Delta p_n = kp_n \tag{6.22}$$

where the growth constant $k > 0$ reflects the rate of reproduction. One would assume that for rabbits this constant would be larger than for elephants. Actual values for k must be determined empirically from the data using a data fitting technique such as least squares.

If instead of simply taking $k > 0$ in Equation (6.22) we could have modeled both the birth rate k_b and the death rate k_d such that

$$p_{n+1} - p_n = k_b p_n - k_d p_n \tag{6.23}$$

Clearly now we may write

$$k = k_b - k_d$$

and as we would expect, if $k > 0$ the model predicts that the population grows exponentially fast and if $-1 < k < 0$ then the population decays exponentially fast. Values of k in the rang $k < -1$ do not make sense because then the solution would oscillate between positive and negative values.

The effect of adding to a population via immigration or subtracting via emigration is captured by

$$p_{n+1} - p_n = kp_n + \beta_n \tag{6.24}$$

where β_n is the net flux of population. Now we might expect that growth rates could be offset by immigration or emigration. For example $k < 0$ but $\beta_n = \beta$ can produce a positive equilibrium population.

Obviously ignoring competition for finite resources places significant limitations on this model. It will work well where the assumptions hold true but when the effects of competition for resources become important it will not capture them. To model competition we may argue as follows: competition occurs when there is interaction between two members of a species and the total amount of competition is the number of ways we can select subsets of 2 from a population p which is

$$\text{number of pairwise interactions} \propto \frac{p(p-1)}{2}$$

where we have divided by two to compute the number of combinations rather than permutations. Now we may modify the model to incorporate competition as

$$p_{n+1} - p_n = k_1 p_n - k_2 p_n (p_n - 1) \tag{6.25}$$

again ignoring effects due to migration. Here we are assuming $k_2 > 0$ and use the negative sign to reflect the fact that competition reduces the population. This equation can be simplified to

$$p_{n+1} - p_n = c_1 p_n - c_2 p_n^2 \tag{6.26}$$

This is the well-known *logistic* difference equation for population growth and it appears to correspond well to the growth of bacteria in agar jelly, for example.

Superficially we see that the difference between the model that does not model competition and the one that does is a quadratic term. A more fundamental difference is that Equation (6.22) is linear while Equation (6.26) is nonlinear. The only fixed point for Equation (6.22) is $\overline{p} = 0$. For Equation (6.26) there are now two fixed points $\overline{p_1} = 0$ and $\overline{p_2} = \frac{c_1}{c_2}$; see Figure 6.3. From the plot of $p_{n+1} - p_n$ it is clear that this new model predicts that the population will be limited, i.e., it can't grow unbounded to ∞ because as soon as $p_n > c_1/c_2$ then $\Delta p_n < 0$ so the population must decrease.

One is initially tempted to conclude that the equilibrium point $\overline{p_1} = 0$ is unstable while the equilibrium point $\overline{p_2} = c_1/c_2$ is stable. As we shall see in the numerical simulations this can be true, but for certain values of c_1 and c_2 the situation can be much more complicated including periodic and even chaotic solutions!

6.4.1 Systems of Equations and Competing Species

Now consider two species A and B whose populations are denoted a_n and b_n, respectively. If we assume that these species have infinite resources and compete neither with themselves or each other then we would propose the simple *system* of difference equations

$$a_{n+1} - a_n = k_1 a_n$$

$$b_{n+1} - b_n = k_2 b_n$$

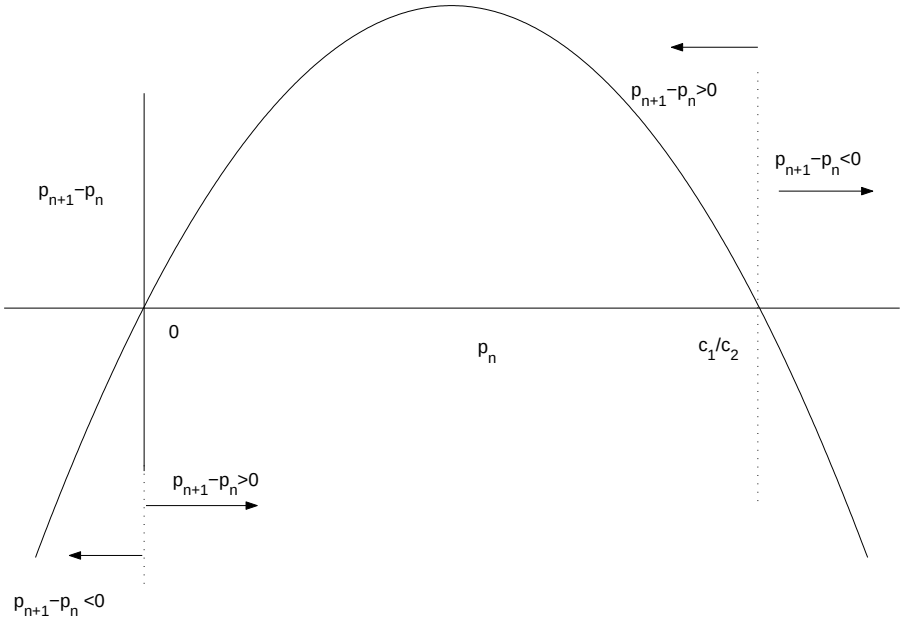


FIGURE 6.3: A plot of the change in population $p_{n+1} - p_n$ as a function of the population p_n.

This system is said to be *uncoupled* as the values of a_n do not influence b_n and, similarly, the values of b_n do not influence a_n.

If species B eats the same kind of food species A does, but species A does not eat the same kind of food species B does we have the model

$$\begin{aligned} a_{n+1} - a_n &= g_1 a_n - c_1 a_n b_n \\ b_{n+1} - b_n &= g_2 b_n \end{aligned}$$

If species A and B both like each others food we would employ the model

$$\begin{aligned} a_{n+1} - a_n &= g_1 a_n - c_1 a_n b_n \\ b_{n+1} - b_n &= g_2 b_n - c_2 a_n b_n \end{aligned}$$

See Figure 6.4 for a numerical simulation of this system. Note that this nonlinear system does not have a closed form solution. For the parameters selected we see that even though species A initially has a lower population it appears to grow without bound while population B becomes extinct. Here we may conclude that species A is more fit than species B and consequently survives.

If species A and B compete both with each other and with themselves the population model would then become

$$\begin{aligned} a_{n+1} - a_n &= g_1 a_n - c_1 a_n b_n - k_1 a_n^2 \\ b_{n+1} - b_n &= g_2 b_n - c_2 a_n b_n - k_2 b_n^2 \end{aligned}$$

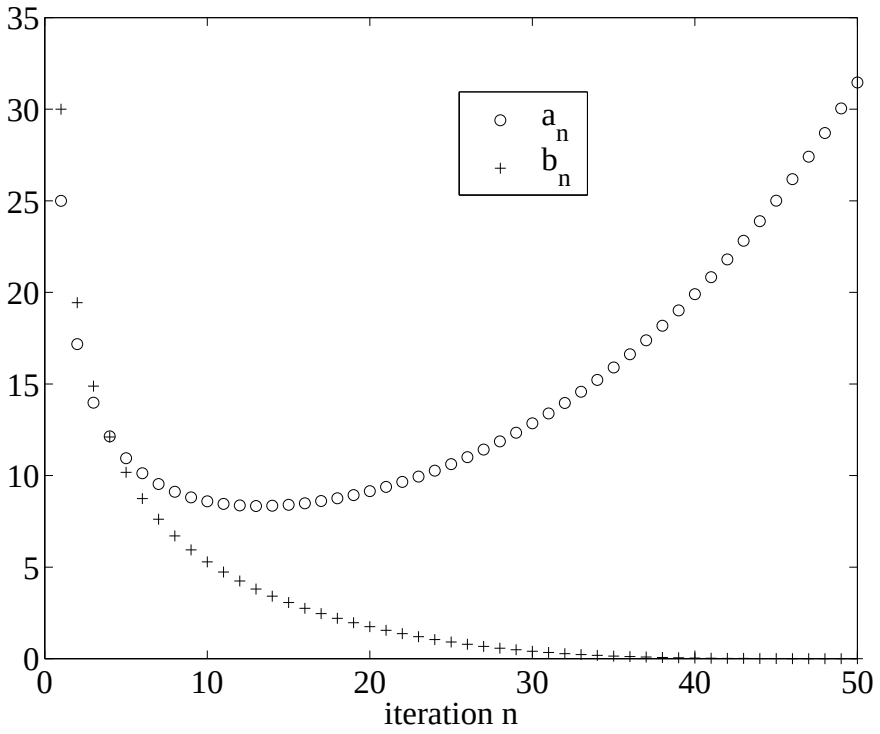


FIGURE 6.4: Competition for finite resources. We have selected the initial conditions $a_1 = 25$ and $b_1 = 30$. In addition the parameters were chosen to be $g_1 = .047, c_1 = .012, g_2 = .023, c_2 = .015$.

Notice that if population B becomes extinct while A survives then the model reduces to the logistic difference equation for a single species.

Predator Prey Model. Now consider modeling the interaction between natural predators and their prey. A classic example of this relationship is given by foxes and rabbits. The population of foxes and rabbits are intimately linked given that the rabbits are the food supply for the foxes. When the population of rabbits increases one can predict an associated, though possibly time lagged, increase in the number of foxes. Conversely, when the number of rabbits decreases the less food there is for the foxes. Of course an increase in the number of foxes will result in more rabbits being eaten and thus a reduction in the rabbit population.

Let's develop a model for this situation. First, denote the fox population by f_n and the rabbit population by r_n. If we assume that in the absence of rabbits the fox population becomes extinct we have the model

$$\Delta f_n = -g_1 f_n$$

where the constant $g_1 > 0$. If rabbits are available, then they should contribute positively to a change in the fox population. It seems reasonable to assume that the increase in the fox population will be proportional to the number of fox and

rabbit interactions which is given by the product $f_n r_n$. Thus, in the presence of rabbits we may model the change in the fox population to be

$$\Delta f_n = -g_1 f_n + c_1 f_n r_n$$

where the constant $c_1 > 0$

Now the rabbits should multiply in the absence of foxes

$$\Delta r_n = g_2 r_n$$

where the constant $g_2 > 0$. The impact of the foxes on the rabbits is presumably also proportional to the number of interactions but now this reduces the rabbit population.

$$\Delta r_n = g_2 r_n - c_2 f_n r_n$$

In summary we have the model

$$f_{n+1} = (1 - g_1) f_n + c_1 f_n r_n \tag{6.27}$$
$$r_{n+1} = (1 + g_2) r_n - c_2 f_n r_n \tag{6.28}$$

Note that we have omitted the competition amongst the foxes for the rabbits as well as the competition amongst the rabbits for their food. This is easily captured by extending the above system to

$$f_{n+1} = (1 - g_1) f_n + c_1 f_n r_n - d_1 f_n^2 \tag{6.29}$$
$$r_{n+1} = (1 + g_2) r_n - c_2 f_n r_n - d_2 r_n^2 \tag{6.30}$$

See Figure 6.5 for a simulation of the above equations. Note that the predicted oscillation is in fact there, however it is damped and the solution proceeds to a stable equilibrium.

6.5 EMPIRICAL MODELING

One may imagine that true observations, e.g., from populations in nature, will not be precise due to limitations in counting species in the wild. Thus, the data will contain what we refer to as a unknown *noise* component. In general, model selection can be arrived at by

1. Collect observations to build models
2. Propose models, e.g., predator prey or competing species
3. Compute model coefficients in each case
4. Compare models through validation and testing

Now we present the method of least squares as a means to determine our unknown model coefficients.

6.5.1 Non-Newtonian Fish?

Recall that Newton's Law of Cooling states that the temperature change in a body is proportional to the difference between the temperature of the body T_n and the

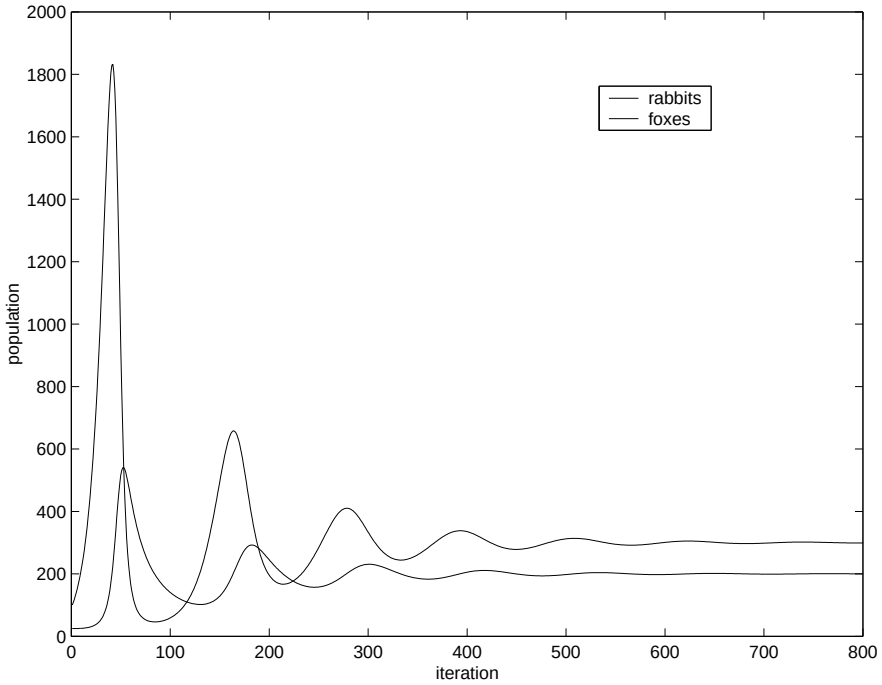


FIGURE 6.5: Simulation of preditor prey equations. We have selected the initial conditions $f_1 = 25$ and $r_1 = 100$. In addition the parameters were chosen to be $g_1 = 0.01, c_1 = 0.0001, g_2 = 0.1, c_2 = 0.0005$, $d_1 = 0.0001$ and $d_2 = 0$.

surrounding temperature M, i.e., as a difference equation

$$\Delta T_n = k(M - T_n)$$

After repeatedly overcooking a certain kind of fish based on this law a frustrated cook has decided to take science into her own hands. She speculates that the actual law of cooking for this fish has the more general form

$$\Delta T_n = k(M - T_n)^\alpha$$

and that for certain types of foods, call them *Non-Newtonian foods*, that $\alpha \neq 1$.

To test her hypothesis, our cook measures the temperature of a fish every minute until it approaches the temperature of the oven which is set to 425 degrees F. The results of her data collection are shown in Figure 6.6.

Now ΔT_n is known since T_n is known for $n = 1, \ldots, 200$. Thus, for any α and k we can compute a model error of

$$E(\alpha, k) = \sum_n \left(\Delta T_n - k(425 - T_n)^\alpha\right)^2$$

We recall from our previous work with least squares that computing α and k requires differentiating the error term $E(\alpha, k)$ with respect to α and k. For this particular

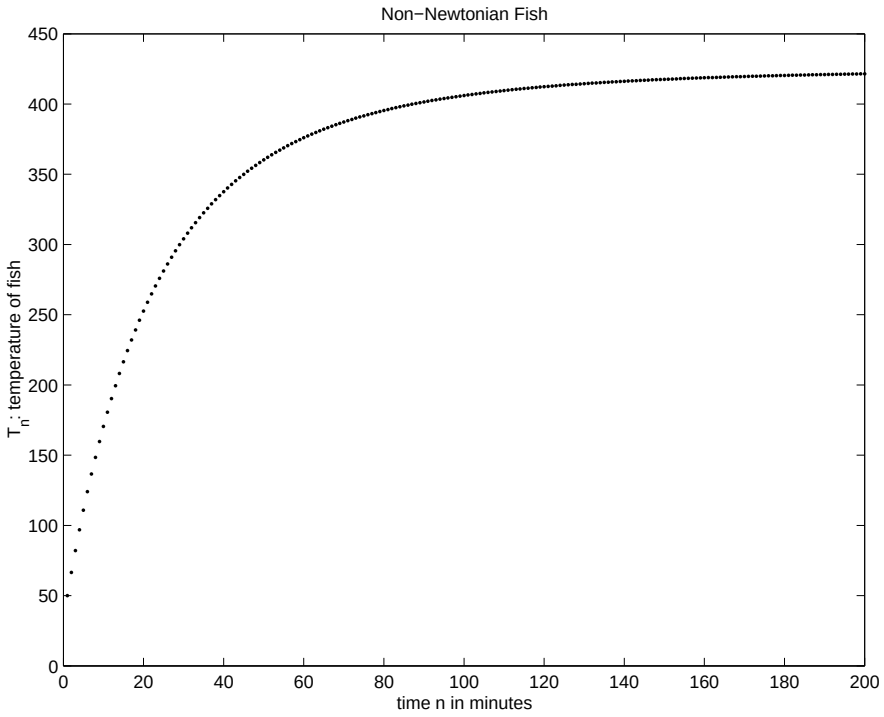


FIGURE 6.6: Observations of a Non-Newtonian fish. These are (synthetic) measurements of the temperature of the fish as a function of time.

model it is simpler to employ a logarithmic transformation

$$y_n = \ln \Delta T_n$$

$$b = \ln k$$

$$x_n = \ln(425 - T_n)$$

giving

$$E(\alpha, b) = \sum_n (y_n - b - \alpha x_n)^2$$

Differentiating these with respect to α and b and setting the results equal to zero produces the equations

$$\begin{pmatrix} \sum_n x_n^2 & \sum_n x_n \\ \sum_n x_n & P \end{pmatrix} \begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} \sum_n y_n x_n \\ \sum_n y_n \end{pmatrix}$$

Solving these equations using *only the first 101 observations* $T_0, T_1, \ldots, T_{100}$ and the MATLAB code provided produces the results

$$\alpha = 1.25$$

and

$$k = 0.01$$

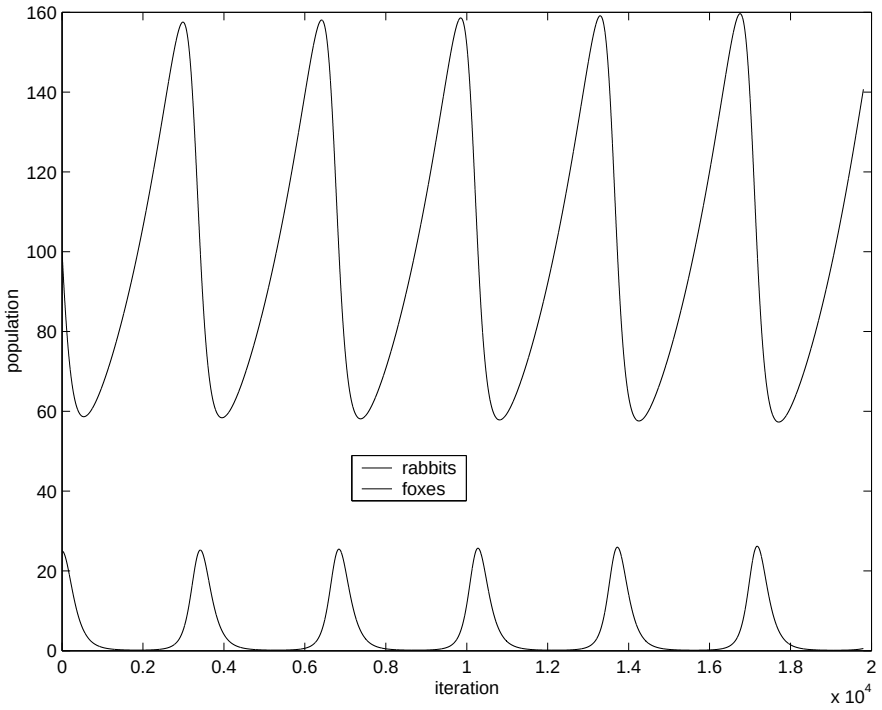


FIGURE 6.7: Simulation of preditor prey equations. We have selected the initial conditions $f_0 = 25$ and $r_0 = 100$. In addition the parameters were chosen to be $g_1 = .01, g_2 = .0005, c_1 = .0001, c_2 = .0001, d_1 = 0.0, d_2 = 0.0$

6.5.2 Predator or Prey?

Assume that the data in Figure 6.7 is provided. The goal is to see if we can calculate the coefficients of the preditor prey equations that will reproduce this data. Thus, given the tentative model

$$\Delta f_n = -g_1 f_n + c_1 f_n r_n$$
$$\Delta r_n = g_2 r_n - c_2 f_n r_n$$

the points $\{f_n, r_n\}$ are now observations while the equation coefficients $\{g_1, g_2, c_1, c_2\}$ are to be determined.

The least squares error is now

$$E(g_1, c_1, g_2, c_2) = \sum_n (\Delta f_n + g_1 f_n - c_1 f_n r_n)^2 + \sum_n (\Delta r_n - g_2 f_n + c_2 f_n r_n)^2 \quad (6.31)$$

Setting

$$\frac{\partial E}{\partial g_1} = \frac{\partial E}{\partial c_1} = \frac{\partial E}{\partial g_2} = \frac{\partial E}{\partial c_2} = 0$$

produces the necessary conditions for a minimum error. Taking the uncoupled

equations for c_1 and g_1 we have

$$\begin{pmatrix} -\sum_n f_n^2 & \sum_n f_n^2 r_n \\ -\sum_n f_n^2 r_n & \sum_n f_n^2 r_n^2 \end{pmatrix} \begin{pmatrix} g_1 \\ c_1 \end{pmatrix} = \begin{pmatrix} \sum_n (\Delta f_n) f_n \\ \sum_n (\Delta f_n) f_n r_n \end{pmatrix} \tag{6.32}$$

These must be solved simultaneously with the uncoupled conditions for c_2 and g_2, i.e.,

$$\begin{pmatrix} -\sum_n r_n^2 & -\sum_n r_n^2 f_n \\ \sum_n r_n^2 f_n & -\sum_n f_n^2 r_n^2 \end{pmatrix} \begin{pmatrix} g_2 \\ c_2 \end{pmatrix} = \begin{pmatrix} \sum_n (\Delta r_n) r_n \\ \sum_n (\Delta r_n) f_n r_n \end{pmatrix} \tag{6.33}$$

Solving these equations produces the exact coefficients that were used to generate the data. In principal, this procedure may be applied to direct observations from nature. One may conclude if a model *fits* the data and, if so, which species plays which role, i.e., by examining the computed signs of c_1, c_2, g_1 and g_2 one may infer which species is the predator and which is the prey. See the MATLAB code for these equations in the Appendix.

PROBLEMS

6.1. Consider the following equations and identify as

- linear or nonlinear
- homogeneous or nonhomogenous
- which order

(a) $x_{n+1}^2 + x_n = 1$.
(b) $x_{n+1} = x_{n-1} + 2$
(c) $x_{n+1} = \sin(x_{n-1})$
(d) $x_{n+3} = x_{n+1} + x_{n-3} + n^2$

6.2. Determine particular solutions to the following equations
(a) $x_{n+1} = x_n + 1$
(b) $x_{n+1} = 5x_n + n^2$
(c) $x_{n+1} = \frac{x_n}{2} + 6^n$

6.3. Show that the real and imaginary parts of a complex solution to a linear difference equation are also solutions to the same difference equation.

6.4. Determine general solutions to the following equations
(a) $x_{n+1} = x_n + 1$
(b) $x_{n+1} = 5x_n + n^2$
(c) $x_{n+1} = \frac{x_n}{2} + 6^n$
(d) $x_{n+1} = \frac{x_n}{2} + 4n^2 + 2n + 1$

6.5. You currently have \$5000 in a savings account that pays 6% interest per year. Interest is compounded monthly. You add another \$200 each month. What do you have on your savings account after five years, and what is the total interest earned during these five years?

6.6. You owe \$500 on a credit card that charges 1.5% interest in each month. You can pay \$50 each month and make no new charges. When is your loan paid off and what is your last payment? How much interest have you paid?

6.7. Your parents are considering a 30–year \$100,000 mortgage at an annual interest rate of 6%. What is the monthly payment, and what will be the total interest paid?

6.8. Mary receives \$5000 as graduation present from her parents when graduating from High School. She deposits the money on a savings account at an annual interest rate of 3%. Interest is compounded monthly. Before going to college she works for three years, and during this time she deposits each month a certain amount on the savings account. She plans to withdraw \$1200 each month in her first year on college, and to increase the monthly withdrawal in each of the following three years by \$100 (in her fourth year on college she withdraws \$1500 each month). What must the monthly payment during the first three years be in order that after Mary's four years on college the balance on the savings account is zero again, and what is the total interest Mary has earned after the seven years?

6.9. Redo Example 6.11 for the case that the annual salary s_m is paid during the first nine months (monthly payment $s_m/9$) in each year, i.e., there is no income and hence no payment on the retirement savings account during the last three months of the year. (This is the situation of university professors if they don't have additional income from grants.)

6.10. Assume the temperature of a roast in the oven increase at a rate proportional to the difference between the oven (set to 400 degrees F) and the roast. If the roast enters the oven at 50 degrees F and is measured one hour later to be at 90 when

should the table be set if the eating temperature is 166 degrees F? Hint: write down the difference equation and solve analytically.

6.11. *Computer.* This question concerns numerically exploring the solutions of the equation

$$p_{n+1} = p_n + \alpha p_n(1 - p_n)$$

First determine all the *equilibrium* solutions of this difference equation by setting $\overline{p} = p_{n+1} = p_n$. Now investigate the stability of these equilibrium numerically. Consider the initial conditions

- $p_0 = 0$
- $p_0 = 0.0001$
- $p_0 = 2$

Numerically simulate the difference equation using the following values of α

- $\alpha = .1$
- $\alpha = .7$
- $\alpha = 1.2$

Describe your results and comment on the stability of the equilibrium you found. Provide plots of all your results. It will make your comparisons easier if you plot all the results for one value of α on a single graph.

6.12. *Computer.* This question concerns numerically exploring the solutions of the equation

$$p_{n+1} = p_n + 0.1p_n(1 - p_n)(2 - p_n)$$

First determine all the *equilibrium* solutions of this difference equation. Numerically simulate the difference equation using the following initial conditions

- $p_0 = 0$
- $p_0 = 0.0001$
- $p_0 = .9999$
- $p_0 = 1$
- $p_0 = 1.0001$
- $p_0 = 1.9999$
- $p_0 = 2$
- $p_0 = 2.0001$

Describe your results and comment on the stability of the equilibrium you found. Provide plots of all your results. It will make your comparisons easier if you plot all the results on a single graph.

6.13. *Computer.* Simulate the fourth order difference equation

$$p_{n+4} = \sin(p_{n+3} + p_{n+2} + p_{n+1} - p_n) + 2$$

and compare to the related equation

$$p_{n+4} = \sin(p_{n+2} + p_{n+1} - p_n) + 2$$

using the initial coditions $p_1 = 6, p_2 = 1, p_3 = 2.5, p_4 = -3$. Explore modifications to these difference equations and see if you can find any interesting behavior. For example, what is the effect of varying the nonhomogeneous term? Plot your results in each case for 100 iterations.

6.14. *Computer.* Consider the system of difference equations

$$x_{n+1} = 0.3x_n + 0.8y_n$$

$$y_{n+1} = 0.7x_n + 0.2y_n$$

Simulate these equations numerically for a variety of initial conditions and attempt to determine any stable equilibrium. Verify that you have actually determined an equilibrium solution by substituting into the original system. (Note that the equilibrium solution in this problem actually depends on the initial condition.)

(a) How do the solutions change if you modify the first coefficient from 0.30 to 0.31?

(b) How do the solutions change if you modify the first coefficient from 0.30 to 0.31 *and* modify the 0.7 coefficient to 0.69.

(c) Compare the results in part a) and b). Can you explain?

6.15. Consider the difference equation

$$x_{n+2} + \alpha x_{n+1} + \beta x_n = 0$$

where it is assumed that $\alpha^2 - 4\beta = 0$. Show that $x_n = (-\frac{\alpha}{2})^n n$ is a solution.

6.16. Find the linear second order nonhomogeneous difference equation relating the price p_{2n+2} and p_{2n} in the cobweb model. Solve this equation and produce a convergence criterion. What does the equilibrium price converge to? Check your result by computing the point of intersection of the supply and demand curves.

6.17. Determine analytical solutions to the following difference equations assuming in each case that $x_1 = 1$ and $x_0 = -1$. Plot your results.

a) $x_{n+2} + 3x_{n+1} + x_n = 0$

b) $10x_{n+2} + x_{n+1} + x_n = 0$

c) $x_{n+2} + \sqrt{3}x_{n+1} + \frac{3}{4}x_n = 0$

6.18. Extend the population model with pairwise competition to include competition among groups of three. Furthermore, assume that the competition among groups of three is more intense than competition between pairs. Identity the new equilibrium solution(s). Use a plot of Δp_n versus p_n to argue whether the model predicts a bounded population.

6.19. Consider a clam population that obeys the logistic difference Equation (6.26). Modify this equation to account for constant harvesting of the clams. By computing the new equilibrium points of the population model describe the impact of harvesting on the clam population.

6.20. Consider three species A, B, C and the evolution of their populations a_n, b_n and c_n.

- Species A eats B and C
- Species B eats neither A nor B
- Species C eats only A.
- Species B eats waste products produced by species A and B.
- The population of both species A and B increase in the absence of other species.
- The population of species C decreases in the absence of A and B.

- Species C is competes with itself for food while this is not true for species A and B.

Write down a system of three coupled difference equations modeling the populations of the three species.

6.21. *Computer.* Provide a model for the bee colony population data in Table 6.2. What does your model predict the long-term population to be?

day	1	2	3	4	5	6	7	8	9	10
number	20	25	60	85	111	146	177	182	184	171
day	11	12	13	14	15	16	17	18	19	20
number	179	167	161	146	159	154	162	166	166	168

TABLE 6.2: Bee colony population data.

6.22. Find the equilibria of the difference equation

$$p_{n+1} = p_n - 0.1p_n(1 - p_n)(2 - p_n)$$

and determine which of them are stable.

6.23. *Computer.* Numerically compute and plot 50 iterates of the difference equation

$$p_{n+1} = p_n - 0.1p_n(1 - p_n)(2 - p_n)$$

for each of the initial conditions

(a) (a) $p_0 = 0.9$

(b) $p_0 = 1.1$.

Is the behavior of the iterates consistent with the stability calculation of Problem 6.22?

6.24. *Computer.* Find all the equilibrium solutions of the logistic difference equation

$$x_{n+1} = rx_n(1 - x_n)$$

as a function of r. Letting $x_0 = 0.2$ numerically iterate this difference equation for 200 iterations for the following values of r:

- $r = 2$
- $r = 3.2$
- $r = 3.8282$
- $r = 3.83$

Plot your results x_n as a function of n for each case and comment. Does this seem like a reasonable model for a population?

6.25. Consider the logistic difference equation with $r > 0$:

$$p_{n+1} = rp_n(1 - p_n).$$

(a) Show that $\overline{p}_1 = 0$ is an equilibrium.

(b) Find the second equilibrium $\overline{p}_2(r)$. For which values of r is $\overline{p}_2(r) \geq 0$?

(c) For which values of r is $\overline{p}_1 = 0$ stable?

(d) For which values of r is $\overline{p}_2(r)$ stable?

6.26. *Computer.* Numerically compute and plot 50 iterates of the difference equation

$$p_{n+1} = rp_n(1 - p_n)$$

for $p_0 = 0.5$ and each of the following values of r:
(a) $r = 0.8$
(b) $r = 2.9$
(c) $r = 3.1$
(d) $r = 3.5$
(e) $r = 3.9$.
Describe the behavior of the iterates and relate it, where possible, to the stability calculation of Problem 6.25.

6.27. *Computer.* Use a least squares approach to determine k in Newton's Law of cooling

$$T_{n+1} = T_n + k(M - T_n)$$

using the data generated by our empirical fish model

$$T_{n+1} = T_n + 0.01(M - T_n)^{1.25}$$

First generate 200 points using this equation and compute k based on these points. Now predict the next 200 points and calculate the error. If a fish is well cooked at 170 degrees F how long does each model predict it will take to cook the fish? Use the values $M = 425$ and $T_0 = 50$.

6.28. *Computer.* Using the data provided in Table 6.3 estimate via least squares the coeffients c_1, c_2, d_1, d_2 in the model

$$a_{n+1} = a_n + c_1 a_n + d_1 b_n$$

$$b_{n+1} = b_n + c_2 a_n + d_2 b_n$$

Include your equations for the unknown coefficients in your write-up.

n	a_n	b_n
1	15.00	45.00
2	30.00	30.00
3	24.00	36.00
4	26.40	33.60
5	25.44	34.56
6	25.82	34.18
7	25.67	34.33
8	25.73	34.27
9	25.71	34.29
10	25.72	34.28
11	25.71	34.29

TABLE 6.3: Did this data come from a linear system?

6.29. Extend the Equations (6.32) and (6.33) provided for computing the coefficients c_1, c_2, g_1, g_2 for the predator-prey model with no intra-species competition given by Equation (6.27) to the case of Equations 6.29 where intraspecies competition is accounted for. Your equations should now provide estimates for $c_1, c_2, g_1, g_2, d_1, d_2$

6.30. Consider the differential equation for the unforced damped nonlinear pendulum

$$\frac{d^2x}{dt^2} + \alpha\frac{dx}{dt} + \sin x = 0$$

where $x(t)$ represents the angular displacement from the equilibrium in radians. Using the expressions for the numerical estimates of the derivatives

$$\frac{d^2x}{dt^2} = \frac{x_{n+1} + x_{n-1} - 2x_n}{(\Delta t)^2}$$

and

$$\frac{dx}{dt} = \frac{x_n - x_{n-1}}{\Delta t}$$

where $x_n \equiv x(n\Delta t)$.

(a) Show that the differential equation can be approximated by the second order difference equation

$$x_{n+1} = (2 - \alpha\Delta t)x_n + (\alpha\Delta t - 1)x_{n-1} - (\Delta t)^2 \sin(x_n) \tag{6.34}$$

(b) Simulate this difference equation for 1000 iterations using the values $\Delta t = 0.05, \alpha = 0.1, x_1 = 0, x_2 = 0.0001$ and plot your result. Repeat this calculation for $x_1 = 2, x_2 = 2.0001$ and compare your results.

(c) Redo this simulation using the *small angle approximation* $\sin x = x$, i.e., simulate

$$x_{n+1} = (2 - \alpha\Delta t)x_n + (\alpha\Delta t - 1)x_{n-1} - (\Delta t)^2 x_n \tag{6.35}$$

using the values $\Delta t = 0.05, \alpha = 0.1, x_1 = 0, x_2 = 0.0001$ and plot your result. Again, repeat this calculation for $x_1 = 2, x_2 = 2.0001$ and compare your results with those found in part (c).

(d) Rewrite the second order Equation (6.34) as a system of two first order equations via the substitution $y_{n+1} = x_n$ and determine all equilibria. Note that the equilibria can also be determined directly from Equation (6.34).

(e) By computing the eigenvalues of the Jacobian matrix of this system, ascertain which equilibria are stable and unstable. Discuss.

6.31. Repeat parts (d) and (e) of Problem 6.30 for the small angle approximation Equation (6.35) and compare.

6.32. Analytically solve the linear difference equation from the previous problem

$$x_{n+1} = (2 - \alpha\Delta t)x_n + (\alpha\Delta t - 1)x_{n-1} - (\Delta t)^2 x_n$$

and compare with your numerical simulation above. For simplicity you may take $\Delta t = 0.05, \alpha = 0.1, x_1 = 2, x_2 = 2.0001$.

6.33. Analytically solve the linear nonhomogeneous difference equation

$$x_{n+1} = (2 - \alpha\Delta t)x_n + (\alpha\Delta t - 1)x_{n-1} - (\Delta t)^2 x_n + 0.01\sin(n/50)$$

Simulate this problem numerically and compare with your analytical solution for 2000 iterations. Can you identify a transient (i.e., a term that goes to zero) and steady state (persistent) components of your solution? Again, for simplicity you may take $\Delta t = 0.05, \alpha = 0.1, x_1 = 2, x_2 = 2.0001$. Hint: combine your solution to the homogeneous problem found above with a particular solution of the form

$$p_n = A\cos(n/50) + B\sin(n/50)$$

Solve for the undetermined coefficients A and B.

REFERENCES

[**1**] Bagnet, G. C., 2206, *The widget maker's guide to snarfle splatting and freen wongling*, 17th Edition, Buena Free Press, Crawdadsville, South Vermont.

[**2**] D. Knuth, Notices Amer. Math. Soc. **49** (2002), no. 3, 318–324.

[**3**] I. Lepper, Theoret. Comput. Sci. **269** (2001), no. 1-2, 433–450.

[**4**] R. R. Fletcher, III, Congr. Numer. **147** (2000), 17–31.

CHAPTER 7

Simulation Modeling

It is not unusual that the complexity of a phenomenon or system makes a direct mathematical attack time-consuming, or worse, intractable. An alternative modeling approach consists of the literal execution of rules by the computer. Such *simulation* approaches occur in great variety but share the common feature that a computer is the central vehicle for knowledge, or process discovery. Here are some problems where simulation modeling appears to be successful, if not indispensible:

- Develop strategies in games with simple rules but stochastic components such as blackjack, checkers, and solitaire.
- Numerically approximate solutions to complex models such as the Navier-Stokes equations of fluid dynamics. (These simulations are typically deterministic).
- Modeling phenomenon with inherent probabilistic components such as processing queues, traffic problems, and inventory problems.

This above list, while certainly not complete, gives an indication of the range of problems that lend themselves naturally to simulation modeling.

7.1 THE TIRE DISTRIBUTOR PROBLEM

Consider the inventory problem confronting the distributor of a commodity with random demand. To be concrete we will consider the situation of a car tire distributor. Our assumption is that this distributor supplies tires to a large number of clients and in turn purchases its supply of tires from a major tire producer. The distributor itself does not fabricate tires but relies on deliveries from the factory. Based on observed daily demand it is up to the distributor to determine a quantity X of tires to be delivered at an integer interval N in days.

The problem is driven by costs. It is the desire of the distributor to select an X and N to minimize total costs. These costs are assumed to consist of two components:

- tire delivery costs
- interest costs on money borrowed to pay for tires in stock

The tire delivery costs may be modeled in several different ways. We will assume that the factory delivers and charges by the truckload and each truckload can contain up to 1,000 tires. Furthermore, the delivery charges for a truckload with no discount if the truck is not full. Hence, if a truckload costs $\$\alpha$, then the delivery costs are $d(X)$ are

$$d(X) = \alpha\left([\frac{X-1}{1000}] + 1 \right)$$

where the notation $[a]$ means the the number a rounded down to the nearest integer. So, for example,

$$\left[\frac{1500}{1000}\right] = 1$$

The second cost is due to the interest charged by the bank for the load the company has to purchase the tires. Note that even if the distributor had enough cash flow to not need to borrow, we would still view the investment in the tires sitting in the warehouse as a cost as this money could presumably be invested to earn a dividend.

Let's assume the interest rate per day per tire is β. If x_n tires are in stock at the end of day n then we are assessed the interest fee βx_n so that the total cost of interest over an interval N is

$$I = \sum_{n=1}^{N} \beta x_n$$

This problem is complicated by the fact that the number of tires purchased by customers varies randomly from day to day as shown in Table 7.1. Furthermore, the average demand was calculated to be 997 tires/day. Thus, the distributor must attempt to simulate this statistical demand in the model. To accomplish this we need to develop a demand subroutine that will mimic the observed daily demand. This may be achieved by associating the demand intervals with segments of the unit interval the length of which is determined by the actual frequence of demand. So, for example, the demand interval $0 \leq x_n < 100$ occurs 12 days out of 365 (presumably due to holidays). So, we associate the interval

$$I_1 = [0, 12/365]$$

to the probability that the daily demand will be in the interval $0 \leq x_n < 100$ tires. Since we are mapping the interval [0,1] to daily demand ranges we must require that these subintervals be nonoverlapping. Thus, the fact that the interval $100 \leq x_n < 300$ occurred on exactly 4 days means that we should reserve $4/365$ of the unit interval for this demand range, i.e., $I_2 = [12/365, 16/365]$. We may develop the rest of the intervals in a similar fashion, i.e.,

$$I_3 = [16/365, 43/365]$$

$$I_4 = [43/365, 86/365]$$

and so on.

Now we have partitioned the unit interval, i.e.,

$$[0, 1] = I_1 \cup I_2 \cup \cdots \cup I_{11}$$

Thus, if we pick a random number $z \in [0, 1]$ (always picking this number uniformly from the interval) we can map that number to an appropriate daily demand interval.

For example, if our uniform random number generator returns $z = .0768$ then we select the interval containing this point, i.e., I_3. Now the question remains how to pick an actual daily demand? If we are in interval I_3 we only know that

the demand should be between 300 and 500 tires. So we may select this demand randomly from the integers in this interval. This model we have constructed for simulating the demand can be simply tested by running the model for 100 years and seeing if we reproduce the demands (now averaged over the 100 years). The results of doing this, shown in the last column of table 7.1 suggest this model is rather good.

daily demand	frequency in days	cumulative distribution	simulated freq.
$0 \leq x_n < 100$	12	12	12.22
$100 \leq x_n < 300$	4	16	4.04
$300 \leq x_n < 500$	27	43	26.67
$500 \leq x_n < 700$	43	86	43.05
$700 \leq x_n < 900$	48	134	47.20
$900 \leq x_n < 1100$	67	201	66.31
$1100 \leq x_n < 1300$	78	279	78.92
$1300 \leq x_n < 1500$	55	334	55.77
$1500 \leq x_n < 1700$	22	356	21.56
$1700 \leq x_n < 190$	7	363	7.24
$1900 \leq x_n < 2100$	2	365	2.02

TABLE 7.1: Number of days certain quantities of tires were demanded. Total number of days of collected data is 365.

Summary of Tire Distributor Simulation

We run the simulation for 365 days and compute the average daily cost. This calculation is repeated 100 times and the average cost is now averaged again. This produces a more reliable stochastic estimate of the cost.

- Make an initial delivery of tires of size deliver_quantity.
- Compute the stochastic demand and subtract sales from stock NUM_TIRES
- If the stock is zero then add a penalty
- Accrue interest costs every iteration.
- If delivery_interval counter indicates delivery then increment NUM_TIRES by delivery_quantity.

7.2 BLACKJACK STRATEGY

Blackjack is a poker game pitting the Dealer, or Bank, against one or more players. For simplicity we will assume there is only one player, as in video blackjack. The object of the game is to score higher than the dealer without going over 21 points. If this occurs the player is paid the value of his bet. If the Dealer has a higher score than the player the Dealer collects the bet. Ties result in no loss of bet, or a *push*.

Card Values. The 10, Jack, Queen, and King are all valued at 10 while the cards from 2 through 9 are valued as indicated. The score of a hand is obtained

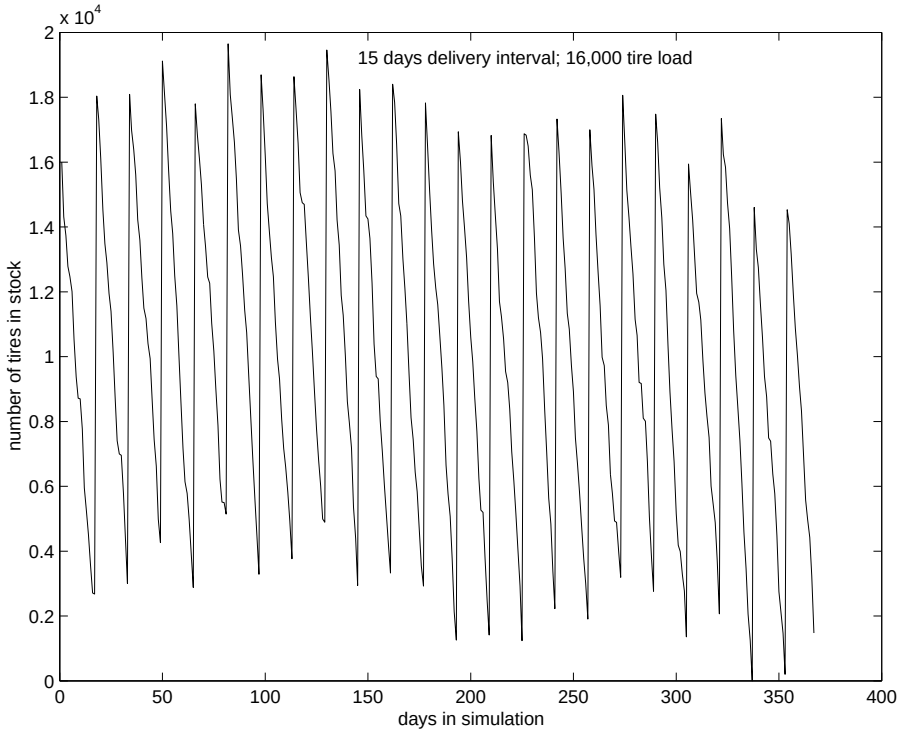


FIGURE 7.1: The average daily cost for this simulation averaged over 365 days was $228. This simulation has no days without tires and does not tend to accumulate tires that would be subject to interest.

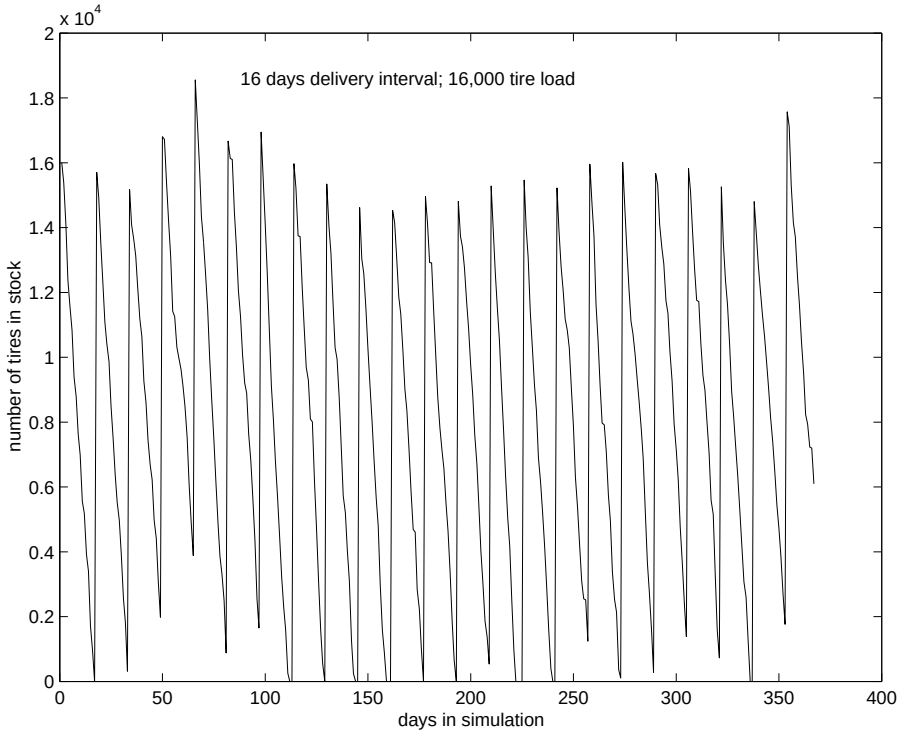


FIGURE 7.2: The average daily cost for this simulation averaged over 365 days was $703. The increased cost is due to number of days (19) where the stock went to zero.

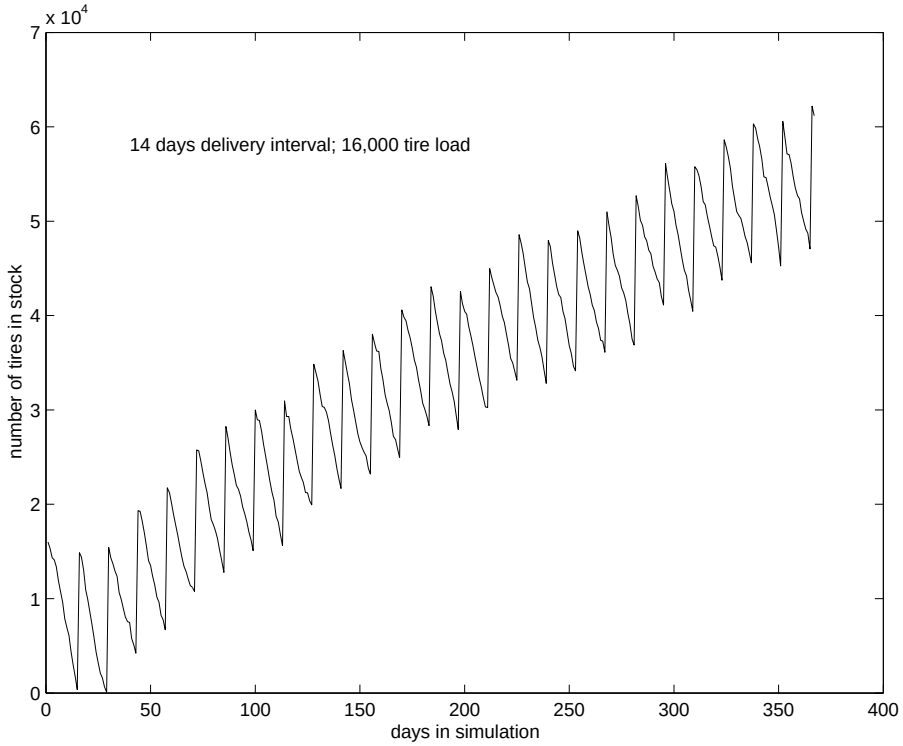


FIGURE 7.3: The daily cost for this simulation averaged over 365 days was $447. This increased cost is apparently due to the interest costs on the increasing stock.

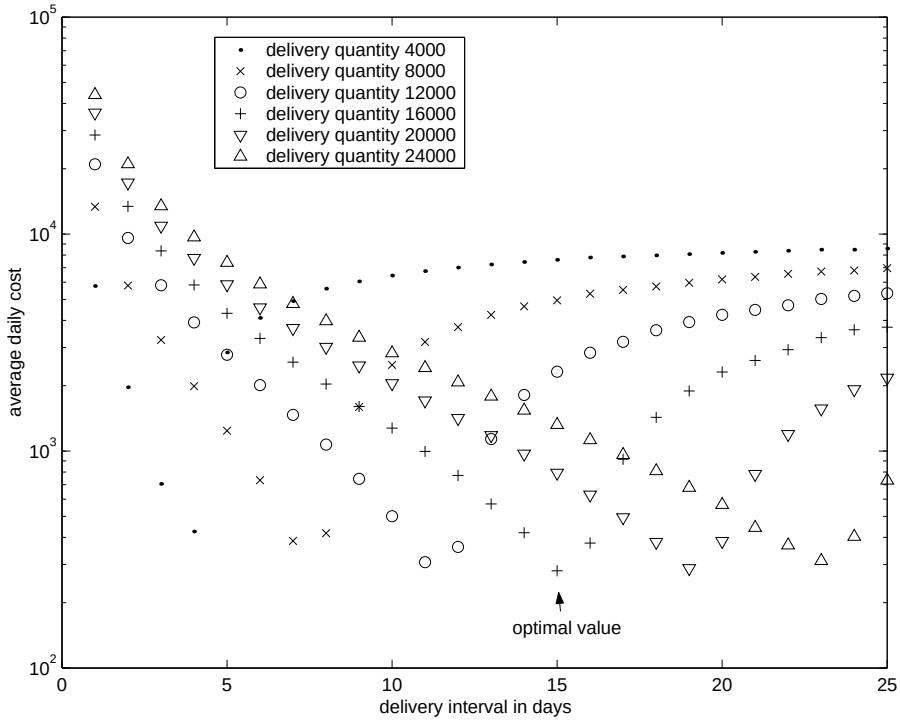


FIGURE 7.4: Average daily cost for a range of delivery intervals and delivery quantities. The minumum costs are themselves averaged over 365 day cycles and in turn these cycles are rerun 100 times and the result of the minimum daily cost averaged. The minimum average cost in these curves is (rounded to the nearest dollar) \$415, \$416, \$354, \$331, \$340 and \$341. A deliver interval of 15 days and quanity 16,000 tires appears optimal from our simulations.

by adding the values of the individual cards in the hand. The Ace is worth 1 or 11 points. The player can choose the value of the Ace whereas the dealer must always take the high value unless the total is over 21.

Blackjack occurs when a total of 21 is obtained with the first two cards in a hand (a black Jack is not necessary). If the Dealer has Blackjack and the Player has 21, the Dealer wins. If the Dealer has 21 and the Player has Blackjack, the Player wins. If a Player wins with a Blackjack then he is paid 1.5 times the bet placed. A total of 21 with more than two cards is not black jack.

Rules of Play. The dealer distributes two cards to the Player and to himself. The Dealer shows the value of one of his cards to the player. The Player then requests cards one at a time until he decides to stay pat (receive no more cards) or goes bust (exceeds 21). If the Player stands pat then the Dealer plays. The Dealer has no choices in how he plays his cards but must follow a specific set of rules.

Dealer Play. Dealer must stand pat on 17. Dealer must continue to take cards until his total is 17 or higher An Ace in the dealer's hand is always counted as 11 unless counting it as 1 prevents the dealer from going over. Thus if the dealer holds (ace,7) then he must stand. If the dealer holds (10,6,Ace) where the Ace is the third card picked then the dealer holds 17 and must stand.

PROBLEMS

7.1. The tire factory that supplies our distributor is overstocked and has decided to discount shipping such that the first truck costs the usual amount $\alpha = \$400$, the second truck costs $\$\alpha/2$, the third truck costs $\$\alpha/3$ and so on. Modify the tire distributor problem to account for this and determine new optimal values of delivery_amount and delivery_interval. Are your answers what you would expect?

7.2. Modify the ordering of a new shipment of tires to occur when the stock is at 20% of the delivery load. How does this effect the optimal delivery amount and interval? Does the smallest average daily cost go down?

7.3. Modify your code so that the number of tires in stock can't excede $T_{\max} = 17,000$ tires. Include graphs of your new simulations to demonstrate this limited capacity. What is the new optimal delivery_interval and delivery_quantity?

7.4. Modify the ordering of a new shipment of tires to occur when the stock is at 20% of the delivery load. How does this effect the optimal delivery amount and interval? Does the smallest average daily cost go down?

7.5. Now assume that the placement of orders for new shipments of tires have a random component based on the routines of the three rotating managers. Assume manager I works 30% of the shifts and that he will place the order when stock dips under 35% *of the delivery_quantity*; manager II works 50 % of the shifts and he places orders will when the stock dips below 20% of the delivery_quantity; manager III works 20% of the shifts and he places orders when the stock is just under 5% of the delivery_quantity. Modify your code to account for this. What is the new optimal deliver_quantity? (Note: there is now no delivery_interval.) Hand in your modified code for grading.

7.6. In the simulation code provided the days without tires are all charged at the same flat rate. Modify the code to account for the fact that some tires may have been sold before the stock ran out. For example, if there were 700 tires at the beginning of the day and 800 were sold the penalty for not filling 100 orders should be smaller than if the whole day were spent without tires.

7.7. Adapt the Blackjack computer program in the Appendices to exploit knowledge of the Dealer's showing card when deciding when to stop getting new cards. Run the simulation 1000 times (i.e., 2000 decks of cards) picking my_stay_value based on the value of the Dealer's face card and compare the results. Compare your winnings with the provided code that does not consider the Dealer's face card. You should win a higher percentage after the modification.

7.8. Experiment with the player's strategy for deciding whether to select an ace as a one or eleven based on the Dealer's showing card. Can you improve over your results in the previous problem?

APPENDIX A

Matlab Code for Data Fitting

A.1 MAMMALIAN HEART RATE PROBLEM

```
File: ls_mammals.m
----------------------Start of actual file contents------------------------
%LEAST SQUARES ANALYSIS OF MAMMALIAN HEART RATE

%w body weights
%r corresponding heart rates
%data is a row vector
w = [3.5 4 6 25 103 117 200 252 300 437 1340 2000 2700 5000
     22500 30000 33000 50000  70000 100000 415000 450000 500000 3000000];
r = [787 660 588 670 347 300 420 352 300 269 251 205 187 120 100
     85 81 70 72 70 45 38 40 48];

x1 = w.^(-1/3) %raise each component of vector w to -1/3 power.
x2 = w.^(-2/3) %the result of this operation is a vector the same size as w
```

Now calculate the slope given by the formula:

$$k = \frac{\sum_{i=1}^{P} r_i w_i^{-1/3}}{\sum_{i=1}^{P} w_i^{-2/3}}$$

We will use the variables numerator and denominator to split up the calculation in the obvious fashion. The numerator is expressed as the vector dot product of r, the row vector of heart rates, and x1 as found above.

```
numerator = r*x1';%apply transpose operator ' to x1 to compute dot product.
denominator = sum(x2);%compute the sum of each component
k1 = numerator/denominator;
```

Now we reproduce the above calculation reproducing all the steps but by using a different data set to compute the slope. It would be more effiecient to pass the data to a subroutine rather than repeat all the code. We examine this in the next section.

```
%%Now build the model on the first 2/3 of the data (16 points)
ws = w(8:24)%the notation 8:24 is equivalent to [8 9 10 11 12 .... 24]
rs = r(8:24)

x1 = ws.^(-1/3)%raise each component of vector w to -1/3 power.
x2 = ws.^(-2/3)

numerator = rs*x1';%apply transpose operator ' to x1 to compute dot product.
denominator = sum(x2);%compute the sum of each component
k2 = numerator/denominator;%see formula in section 3.1.1

hold on
plot(w.^(-1/3),r,'o')%plot raw data
plot(w.^(-1/3),k1*w.^(-1/3),'--x')%plot first model
plot(w.^(-1/3),k2*w.^(-1/3),'--v')%plot second model
title('mammalian heart rate model')
xlabel('weight w^{(-1/3)}')
ylabel('pulse rate')
legend('raw data','least squares fit (all data)', 'least squares fit 2/3 data')
-----------------------End of actual file contents--------------------------------
```

A.2 LEAST SQUARES WITH NORMAL EQUATIONS

This program consists of two parts: a subroutine called ls_normal.m and a driver called run_ls.m.

```
File:  ls_normal.m
---------------------Start of actual file contents-----------------------
%Input:
%   x is a column vector of domain (input) variables
%   y is a column vector of range (output) variables
%
%Output:
%   m is the slope of the line
%   b is the intercept of the line

function [m,b] = ls_normal (x,y)

P = size(x,1)%how many points are there in this column vector?
```

Now we compute the terms required in the evaluation of m and b in the normal equations. Recall

$$m = \frac{(\sum y_i)(\sum x_i) - P\sum y_i x_i}{(\sum x_i)^2 - P\sum x_i^2}$$

$$b = \frac{-(\sum y_i)(\sum x_i^2) + (\sum x_i)(\sum y_i x_i)}{(\sum x_i)^2 - P\sum x_i^2}$$

We set sy $= \sum y_i$, dp_xy $= \sum y_i x_i$ and x_sq $= \sum x_i^2$.

```
sy = sum(y);%sum y_i (scalar)
sx = sum(x);%sum x_i (scalar)
dp_xy = x'*y; %y dot product x (scalar)
x_sq = sum(x.*x);%sum x_i^2 term (scalar)
denom = sx^2 - P*x_sq; %(scalar)

m = (sy*sx-P*dp_xy)/denom
b = (-sy*x_sq+sx*dp_xy)/denom

%plot results
plot(x,y,'o')
hold
plot(x,m*x+b,'--v')
legend('data','model')
-----------------------End of actual file contents--------------------------
```

The above subroutine is called using the following driver:

```
File: run_ls.m
----------------------Start of actual file contents------------------------
w = [3.5 4 6 25 103 117 200 252 300 437 1340 2000 2700 5000
     22500 30000 33000 50000 70000 100000 415000 450000 500000 3000000];
r = [787 660 588 670 347 300 420 352 300 269 251 205 187 120 100
     85 81 70 72 70 45 38 40 48];

x = w.^(-1/3)'%note that the transpose operator ' turns the row vec into col vec.
y = r'% the subroutine expects column vectors by design.

[m,b] = ls_normal (x,y)
-----------------------End of actual file contents--------------------------------
```

A.3 LEAST SQUARES WITH OVERDETERMINED SYSTEM

```
File: ls_interp.m
----------------------Start of actual file contents------------------------
%Input:
%   x is a column vector of domain (input) variables
%   y is a column vector of range (output) variables
%
%Output:
%   m is the slope of the line
%   b is the intercept of the line

function [m,b] = ls_interp(x,y)

%Compute the matrix X
```

Recall the equation we are solving in this problem:

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_P \end{pmatrix} \begin{pmatrix} b \\ m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_P \end{pmatrix}$$

In terms of matrices we can summarize the above as

$$\mathrm{Xvec} = \mathrm{y}$$

This equation will be solved via matlab's backslash routine.

```
P = size(x,1)%how big is the data set?

c1 = ones(P,1)%create a column vector of ones of length P

X = [c1 x]% construct the "interpolation matrix"

vec = X\y% solve the least squares problem

b = vec(1)%obtain the first component (intercept)
m = vec(2)%obtain the slope

%plot results
plot(x,y,'o')
```

```
hold
plot(x,m*x+b,'--v')
legend('data','model')
----------------------End of actual file contents--------------------------------
```

A.4 NON-NEWTONIAN FISH

See the discussion in Section 6.5.1.

```
File: fishfit.m
---------------------Start of actual file contents-------------------------
y = log(T(2:101)-T(1:100))%use only the first 100 points
x = log(425-T)%


m11 = x(1:100)*x(1:100)'

m12 = sum(x(1:100))
m21 = m12
m22 = 100
r1 = (x(1:100))*y'
r2 = sum(y(1:100))

R = [r1;r2]

M = [m11 m12; m21 m22]

XX = inv(M)*R%this vector contains the results [b alpha]
k = exp(XX(2))%transforming back to original model equation
----------------------End of actual file contents--------------------------------
```

A.5 PREDITOR OR PREY?

See the discussion in Section 6.5.2.

```
File: predpreyfit.m
---------------------Start of actual file contents-------------------------
a = f(1:19800);%init data representing the population of species A
b = r(1:19800);%init data representing the population of species B

%%FOXES matrix equation coefs
m11 = -sum(a*a')
m12 = sum(a.*a.*b)
m21 = -m12
m22 = sum(a.*a.*b.*b)

z1 = (f(2:19801)-a)*a'
z2 = sum((f(2:19801)-a).*a.*b)
```

```
M = [m11 m12; m21 m22];%now compute the matrix by assembling the components
R1 = [z1;z2]%this is the RHS of equation for (g_1, c_1)

cg1 = inv(M)*R1%solve for (g_1, c_1)


%%rabbits equation coefs
m11 = sum(b*b')
m12 = -sum(b.*b.*a)
m21 = -m12
m22 = -sum(a.*a.*b.*b)

z1 = (r(2:19801)-b)*b'
z2 = sum((r(2:19801)-b).*a.*b)

M = [m11 m12; m21 m22];
R1 = [z1;z2]

cg2 = inv(M)*R1%solve for (g_2, c_2)
-----------------------End of actual file contents-------------------------------
```

A.6 TIRE DISTRIBUTOR

See the discussion in Chapter 7 Section 7.1.

```
File: flatdemand.m
---------------------Start of actual file contents-----------------------

function [num_tires] = flatdemand

u = rand(1);
u1 = rand(1);
if u < 12/365
    num_tires = floor(u1*100);
elseif u >= 12/365 & u < 16/365
    num_tires = 100 + floor(u1*200);
elseif u >= 16/365 & u < 43/365
    num_tires = 300 + floor(u1*200);
elseif u >= 43/365 & u < 86/365
    num_tires = 500 + floor(u1*200);
elseif u >= 86/365 & u < 134/365
    num_tires = 700 + floor(u1*200);
elseif u >= 134/365 & u < 201/365
    num_tires = 900 + floor(u1*200);
elseif u >= 201/365 & u < 279/365
    num_tires = 1100 + floor(u1*200);
elseif u >= 279/365 & u < 334/365
    num_tires = 1300 + floor(u1*200);
```

```
elseif u >= 334/365 & u < 356/365
    num_tires = 1500 + floor(u1*200);
elseif u >= 356/365 & u < 363/365
    num_tires = 1700 + floor(u1*200);
else u >= 363/365 & u <1
    num_tires = 1900 + floor(u1*200);
end
----------------------End of actual file contents--------------------------------

File: simulate.m
---------------------Start of actual file contents-----------------------


function [AVE_DAILY_COST] = simulate(num_runs, delivery_interval, delivery_quantity)

NUM_DAYS = 365
TRUCK_CAPACITY=4000;%tires
truck_charge = 400;%delivery cost per truck
delivery_charge = truck_charge*(floor((delivery_quantity-1)/TRUCK_CAPACITY)+1);
penalty = 1000;%10 dollar penalty per day * 1000 tires

for i = 1:num_runs
    day_counter = 0;
    %Assume there is a delivery at the outset
    NUM_TIRES = delivery_quantity;%init
    COST =delivery_charge;
    interest_rate = 0.01;
    days_without_tires =0;
    delivery_counter = 0;

    daily_inventory(1) = NUM_TIRES;

    while day_counter <= NUM_DAYS;
        if delivery_counter == delivery_interval;%add delivery charge
            NUM_TIRES = NUM_TIRES + delivery_quantity;
            COST = COST + delivery_charge;
            delivery_counter = 0;
        end
        NUM_TIRES = NUM_TIRES - flatdemand; %sell tires for day

        if NUM_TIRES <=0;%out of stock?
            days_without_tires = days_without_tires + 1;
            NUM_TIRES = 0;
        end
        COST = COST + interest_rate*NUM_TIRES;%add charges for unsold tires
        day_counter = day_counter + 1;
        daily_inventory(day_counter+1) = NUM_TIRES;
```

```
        delivery_counter = delivery_counter + 1;
    end
    %add daily charge due interest at end of day for remaining tires
    cost_penalty = days_without_tires*10000;
    AVE_DAILY_COST(i) = (COST + cost_penalty)/NUM_DAYS;
end
-----------------------End of actual file contents--------------------------------
```

```
File: run_simulate.m
----------------------Start of actual file contents------------------------

num_runs = 100;
delivery_quantity = 4000; %number to have delivered in each shipment (X in notes)


for j=1:6
    for i = 1:25
    delivery_interval = i;
    DAILY_COST(:,i,j) = simulate(num_runs, delivery_interval, delivery_quantity);
    end
    delivery_quantity = delivery_quantity + 4000
end

for j=1:6
    for i = 1:25
    AVES(i,j) = sum(DAILY_COST(:,i,j))/num_runs
    end
end



semilogy(1:25,AVES(:,1),'.',1:25,AVES(:,2),'x',1:25,AVES(:,3),'o',1:25,AVES(:,4),'+',1:25,
legend('delivery quantity 4000','delivery quantity 8000','delivery quantity 12000','delive
ylabel('average daily cost')
xlabel('delivery interval in days')
-----------------------End of actual file contents--------------------------------
```

A.7 BLACKJACK

```
File: blackjack.m
----------------------Start of actual file contents------------------------
%blackjack simulation
%bet one dollar on each hand; start with $100
my_money = 100;
iwin=0;%counter for number of wins
ilose = 0;%counter for number of losses
ties = 0;%counter for number of ties
hand = 0;%counter for number of hands played.
my_stay_value = 14; %don't take another card if hand is worth 14 or more.

%deal cards from 2 decks
for i = 1:100
%note 11= jack, 12 = queen, 13 = king, and 14 = ace.
d1 = [2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5 6 6 6 6 7 7 7 7 8 8 8 8 9 9 9 9 10 10 10 10
      11 11 11 11 12 12 12 12 13 13 13 13 14 14 14 14];
```

```
deck = [d1 d1];
numcards = size(deck,2);
bigindexset = 1:numcards;
permuted_indices = bigindexset(randperm(numcards));
shuffled_cards = deck(permuted_indices);


%One game (of two decks)
card_counter = 1;
while card_counter < numcards - 10%dont let deck run out!
   %new hand
   hand = hand +1;
   my_count = 0;
   dealer_count = 0;

   %make initial deal of two cards to Player and Dealer.
   my_count = my_card_value(card_counter, shuffled_cards, my_count);
   card_counter = card_counter+1;%advance deck index
   dealer_count = dealer_card_value(card_counter, shuffled_cards, dealer_count);
   card_counter = card_counter+1;
   my_count = my_card_value(card_counter, shuffled_cards, my_count);
   card_counter = card_counter+1;
   dealer_count = dealer_card_value(card_counter, shuffled_cards, dealer_count);
   card_counter=card_counter+1;

   while my_count < my_stay_value  & card_counter < numcards
      my_count = my_card_value(card_counter, shuffled_cards, my_count);
      card_counter = card_counter +1;
   end

   while dealer_count < 17  & card_counter < numcards & my_count < 22
      dealer_count = dealer_card_value(card_counter, shuffled_cards, dealer_count);
      card_counter = card_counter +1;
   end

%who wins?
   if my_count > 21% I am bust
      my_money = my_money - 1;
      ilose = ilose +1;
   elseif dealer_count > 21%dealer is bust
      my_money = my_money + 1;
      iwin = iwin +1;
   elseif my_count == dealer_count
      %push--my winnings don't change
      ties = ties +1;
   elseif my_count > dealer_count
      my_money = my_money +1;
```

```
      iwin = iwin +1;
   else
      my_money = my_money -1;
      ilose = ilose + 1;
   end%if
   %construct an array that computes a running fraction of losses
   perclose(hand) = ilose/(iwin+ilose+ties);
   end%while
end%for

   iwin
   ilose
   ties
   plot(perclose)
   my_money
-----------------------End of actual file contents-------------------------------

File: dealer_card_value.m
---------------------Start of actual file contents-------------------------

   function newcount = dealer_card_value(card_counter, shuffled_cards, dealer_count)

   card_value = shuffled_cards(card_counter);
   if card_value >= 10 & card_value < 14
        newcount = dealer_count + 10;
   elseif card_value < 10
      newcount = dealer_count + card_value;
   else%card is an ace and has value of 11 for dealer unless he goes bust.
      bigcount = 11 + dealer_count;
      smallcount = 1 + dealer_count;
      if bigcount <22
         newcount = bigcount;
      else
         newcount = smallcount;
      end
   end
-----------------------End of actual file contents-------------------------------

File: dealer_card_value.m
---------------------Start of actual file contents-------------------------

   function newcount = my_card_value(card_counter, shuffled_cards, my_count)

   card_value = shuffled_cards(card_counter);
   if card_value >= 10 & card_value < 14
        newcount = my_count + 10;
   elseif card_value < 10
      newcount = my_count + card_value;
```

```
   else%card is an ace and has value one or 11.
      bigcount = 11 + my_count;
      smallcount = 1 + my_count;
      if bigcount >18 & bigcount <22
         newcount = bigcount;
      else
         newcount = smallcount;
      end
   end
-----------------------End of actual file contents--------------------------------
```