REACTION-DIFFUSION ANALYSIS

MATH 350 - RENATO FERES
CUPPLES I - ROOM 17 - FERES@WUSTL.EDU - 5-6752

1. NOTE:

These are rough lecture notes for a course on applied math (Math 350), with an emphasis on chemical kinetics, for advanced undergraduate and beginning graduate students in science and mathematics. There are (Fall 2003) three students enrolled: a second year undergraduate with a major in applied mathematics, a senior with majors in mathematics, chemical engineering, and environmental sciences, and a first year graduate student in chemistry. During the course of the semester the students prepared projects on topics related to the course material and are presently scheduled to present them to each other in a four hour "mini-conference" on exam day, December 17, 2003. Their expositions will be about: pattern formation, implementation of a reaction diffusion system numerically, and use of group theory to study the spectrum of the Laplacian.

In parallel to this course (also in the Fall of 2003), Professor Gregory Yablonsky, of the Chemical Engineering Department, taught a course on "Transport in the Environment," CE 583, for chemical engineering students. Professor Yablonsky and I communicated almost daily during the semester and constantly exchanged ideas about the two courses. Our plan is to combine them into one course that contains both the mathematics at a similar depth as discussed here, but with a much more developed scientific content, emphasizing issues in biochemistry and environmental science. Our lecture notes will be integrated into a single text. Professor Yablosnky's lecture notes comprise the following main topics:

- Introduction. Environmental transport;
- fluid mechanics;
- ideal reactors;
- viscosity, transport in non-uniform media, diffusion;
- hydrodynamics and the Navier-Stokes equations;
- realistic models of surface water systems;
- non-steady state behaviors of surface water systems, system of lakes;
- incompletely mixed systems, propagations of polutants;
- conservation of energy, Bernoulli equation, heat balance;
- dissolved oxygen;
- chemical transformations in natural waters;
- eutrophication;
- modeling of the living matter;
- ground water and the subsurface environment;
- ground water motion;
- atmospheric transport and processes;
- transport of chemicals in the atmosphere.

A few of these topics are discussed below as well, but only in a very simplified and not specially coherent form. The unifying themes for the present notes are mathematical. The joint text will have a more developed scientific flesh for the mathematical bare bones described here.

2. What Is This Course About?

This course is centered around the study of systems of reactions (mostly chemical but not exclusively), possibly coupled with transport processes, mainly diffusion. Very broadly stated, we are interested in the temporal and spatial properties of reaction networks. This study brings together a wide variety of mathematical ideas and can be applied to a wealth of scientific problems.

Mathematically, much of the course will be dedicated to the study of 'dynamical systems,' a term that refers to the qualitative study of systems of non-linear (ordinary and/or partial) differential equations. Other subjects, particularly from algebra, will also be needed. To varying degrees of detail, we plan to cover the following topics:

(1) **Box models and steady state**
   (a) Stock, flows, residence time;
   (b) Basic concepts in chemistry and simple applications.
(2) **Abstract Stoichiometry**.
   (a) Linear algebra for complex reaction networks.
   (b) Graph representation.
(3) **Finite Dimensional Dynamical Systems**. This is the qualitative study of systems of ordinary differential equations and is comprised of topics such as:
   (a) Linear systems;
   (b) Generalities about vector fields and flows;
   (c) Local theory: linearization and local stability, stable and central manifolds;
   (d) Global theory: limit sets and attractors; Poincaré-Bendixon theorem, global phase portraits;
   (e) Bifurcation theory: structural stability; bifurcations at nonhyperbolic equilibrium points, Hopf bifurcations and bifurcations of limit cycles from a multiple focus, homoclinic bifurcations and chaos, the Melnikov's method.
(4) **Partial Differential Equations**.
   (a) The linear diffusion equation;
   (b) Diffusion and probability;
   (c) General properties of reaction-diffusion equations and special systems.

Reaction-diffusion equations are important to a wide range of applied areas such as cell processes, drug release, ecology, spread of diseases, industrial catalytic processes, transport of contaminants in the environment, chemistry in interstellar media, to mention a few. Some of these applications, particularly in chemistry and biology, will be considered along the course.

It is not meaningful to talk about a general theory of reaction-diffusion systems. This is a relatively recent subject of mathematical and applied research. Most of the work that has been done so far is concerned with the exploration of particular aspects of very specific systems and equations. This is because such systems are generally very complicated and display a wide array of poorly understood phenomena. As a result, there is no established syllabus for such a course, and no single textbook that would cover the topics that we wish to study. Topics and exercises will thus be drawn from a large number of different sources. The following list of references should prove useful. I will make clear throughout the course which sources I'm using at any particular moment.

(1) G. Nicolis and I. Prigogine, *Exploring Complexity - An Introduction*. W. H. Freeman and Company, New York, 1989.
(2) G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems - From Dissipative Structures to Order Through Fluctuations*. John Wiley & Sons, New York, 1977.
This is a more technical text, of which the first book is the popular, bedside version.
(3) Rutherford Aris, *Elementary Chemical Reactor Analysis*. Dover Publications, Mineola, 1999.
(4) John Harte, *Consider a Spherical Cow - A Course in Environmental Problem Solving*. William Kaufman, Inc, Los Angeles, 1985.
(5) Howard C. Berg, *Random Walks in Biology*. Princeton University Press, Princeton, 1993.
(6) S. I. Rubnow, *Introduction to Mathematical Biology*. Dover Publications, Mineola, 2002.
(7) L. Perko, *Differential Equation and Dynamical System*. Springer-Verlag, 1991. This is my main source for topics related to the theory of dynamical systems.
(8) S. H. Strogatz, *Nonlinear Dynamics and Chaos*. Westview, 1994.
Another good introduction to dynamical systems.
(9) S. Lynch, *Dynamical Systems with Applications using MAPLE*. Birkhäuser, Boston, 2001.
(10) J. D. Murray, *Mathematical Biology*. Springer-Verlag, 1993.
(11) K. Alhumaizi and R. Aris, *Surveying a Dynamical System: a Study of the Gray-Scott reaction in a Two-phase Reactor*. Longman, 1995.
A detailed modeling and bifurcation analysis of a specific system of chemical reactions.
(12) Bruce L. Clarke, *Stability of Complex Reaction Networks*. In Advances in Chemical Physics, Eds. I. Prigogine and S. A. Rice, vol. 43, John Wiley & Sons, 1980.
(13) P. Érdi and J. Tóth, *Mathematical Models of Chemical Reactions - Theory and Applications of Deterministic and Stochastic Models*. Princeton University Press, Princeton, 1989.
(14) V. I. Bykov, V. I. Elokhin, A. N. Gorban and G. S. Yablonskii, *Kinetic Models of Catalytic Reactions*. Comprehensive Chemical Kinetics, Ed. R. G. Compton, Elsevier, 1991.
(15) S. K. Scott, *Chemical Chaos*. International Series of Monographs on Chemistry 24, Clarendon Press, Oxford,1994.

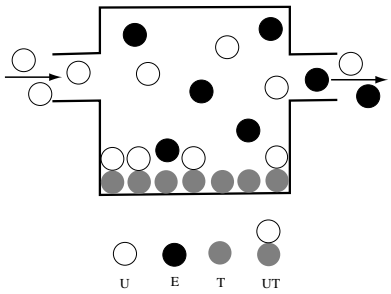
3. Examples of Reaction Networks

Very broadly stated, chemistry is the study of processes that involve *transformations* of substances. These transformations are expressed by *reaction* equations and they must satisfy certain physical restrictions, such as conservation of mass and energy. The object of a chemical reaction analysis is to determine how the amounts of the various substances involved in a given process vary in time and space. That analysis is actually very general and applies to situations that are related to chemistry only in a formal way, but for which the idea of transformation is still central.

I'd like here to suggest a few examples of reaction mechanisms, some of which unquestionably chemical, others not.

3.1. **Student-Teacher Interaction.**

Here the reactor is a "school," which contains a mixture of four "substances": (U) students without learning, (E) educated students, (T) teachers, and (UT) student-teacher "molecule."



The school promotes the overall reaction $U \rightharpoonup E$, in which uneducated students are transformed into educated students. This reaction is mediated by the "catalyst" T through the reaction mechanism:

$$\begin{aligned} U + T &\rightleftharpoons UT \\ UT &\rightharpoonup T + E. \end{aligned}$$

Notice that T is not "consumed" in the reaction. Its function is to accelerate the transformation of U into E.

Exercise 1. *Suppose that the school "reactor" is operating under* steady-state *condition, that is, there is a constant* inflow *of students, which is equal to the* outflow*, so that the total* stock *of enrolled students is constant. Assume this stock is* 8000 *students and that they all graduate after four years. (We say that their* residence time *is four years.) Calculate the outflow, i.e., the number of graduating students per year.*

3.2. **Oxidation of Hydrogen.** Most chemical processes involve a large number of elementary reaction steps. The specification of a detailed reaction mechanism for the process requires spelling out those steps, which is usually a difficult experimental problem. Mathematics can help by providing ways to systematically obtain the possible reaction mechanisms and in exploring the consequences of adopting a particular mechanism as a working hypotheses.

As an example, consider the overall reaction

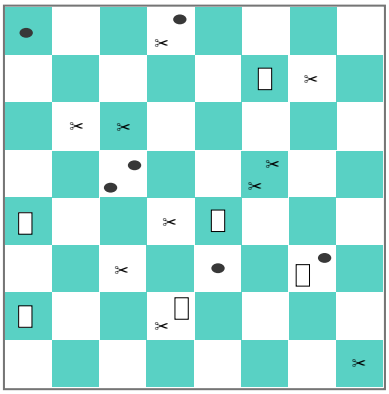
$$2H_2 + O_2 \rightleftharpoons 2H_2O.$$

It is perhaps to be expected that three molecules, two of hydrogen and one of oxygen, coming together at nearly the same time with the right amount of energy to produce two molecules of water would be a low probability event. In fact, it is believed that this overall reaction involves a large number of simpler steps. The following mechanism can be found in *Kinetic Models of Catalytic Reactions*, by Bykov et al. M is any substance which can exchange energy with the other molecules.

$$
\begin{aligned}
H_2 + O_2 &\rightleftharpoons 2OH \\
H_2 + O_2 &\rightleftharpoons HO_2 + H \\
OH + H_2 &\rightleftharpoons H_2O + H \\
H + O_2 &\rightleftharpoons OH + O \\
O + H_2 &\rightleftharpoons OH + H \\
O + H_2O &\rightleftharpoons 2OH \\
H + O_2 + M &\rightleftharpoons HO_2 + M \\
2H + M &\rightleftharpoons H_2 + M \\
2O + M &\rightleftharpoons O_2 + M \\
H + OH + M &\rightleftharpoons H_2O + M \\
2OH + M &\rightleftharpoons H_2O_2 + M \\
OH + O + M &\rightleftharpoons H_2O + M \\
HO_2 + H_2 &\rightleftharpoons H_2O_2 + H \\
HO_2 + H_2 &\rightleftharpoons H_2O + OH \\
HO_2 + H_2O &\rightleftharpoons H_2O_2 + OH \\
2HO_2 &\rightleftharpoons H_2O_2 + O_2 \\
H + HO_2 &\rightleftharpoons 2OH \\
H + HO_2 &\rightleftharpoons H_2O + O \\
H + HO_2 &\rightleftharpoons H_2 + O_2 \\
O + HO_2 &\rightleftharpoons OH + O_2 \\
H + H_2O_2 &\rightleftharpoons H_2O + OH \\
O + H_2O_2 &\rightleftharpoons OH + HO_2 \\
H_2 + O_2 &\rightleftharpoons H_2O + O \\
H_2 + O_2 + M &\rightleftharpoons H_2O_2 + M \\
OH + M &\rightleftharpoons O + H + M \\
HO_2 + OH &\rightleftharpoons H_2O + O_2 \\
H_2 + O + M &\rightleftharpoons H_2O + M \\
O + H_2O + M &\rightleftharpoons H_2O_2 + M \\
O + H_2O_2 &\rightleftharpoons H_2O + O_2 \\
H_2 + H_2O_2 &\rightleftharpoons 2H_2O \\
H + HO_2 + M &\rightleftharpoons H_2O_2 + M.
\end{aligned}
$$

The difficult task of determining the precise reaction mechanism for a given chemical process, such as the one given above, can be helped if there are ways of systematically enumerating all possible reactions involving an specified set of substances. We will see later how linear algebra can be a useful tool in this respect.

3.3. **The 'Rock-Paper-Scissors' Game.** Our next example may seem a bit artificial and is decidedly not chemistry. Consider the following game played by a large number, say N, of players over a checkerboard.



Each player chooses one piece from three kinds: rocks, papers, and scissors, and places it on a square of the checkerboard chosen at random. Let us say that the pieces are initially disposed as shown in the figure. Notice that there are 22 players in this example.

The game now evolves in discrete time, as follows. At time 1 we have the configuration of the game described by the figure. Inductively, suppose that we have obtained the configuration of the game at time n, for some $n = 1,2,3,\cdots$. Then the configuration at time $n+1$ is determined by a two-step rule:

(1) Each player chooses according to some random procedure from among the set

$$\{\text{North, South, East, West}\}$$

and moves its piece one square, accordingly. If the piece occupied a boundary square and the move would force it out of the checkerboard, then the player gets to draw again until it is given a move that keeps its piece within the board.

(2) Once all players have moved its piece as described in part one, we allow the pieces to 'interact' and 'transform.' The rules of transformation will apply to each square separately, as follows:
   (a) If a square is empty, or contains a single piece, nothing is done.
   (b) If a square contains more than one piece, we select two pieces in that square at random and transform them by the following rules ('reactions') below. We write S for 'scissors,' R for 'stone' and P for paper:

$$P+R \rightharpoonup P+P$$
$$S+R \rightharpoonup R+R$$
$$P+S \rightharpoonup S+S$$
$$R+R \rightharpoonup R+R$$
$$P+P \rightharpoonup P+P$$
$$S+S \rightharpoonup S+S.$$

For example, if the square contains one paper and one stone, the stone changes into paper and the paper remains unchanged.

Now and later, we reserve the 'harpoon' type arrow '$\rightharpoonup$' to indicate reactions. The ordinary arrow '$\to$' will be used in the way it is normally used in mathematics, to indicate functions and limit operations.

This is not a win-loose game. The purpose is simply to watch its evolution in time. In other words, we wish to understand how the number of stones, papers and scissors vary as a function of time. Here is a sample of questions we might be interested in answering. Let $\#_S(n)$, $\#_R(n)$, and $\#_P(n)$ represent, respectively, the proportion of scissors, stones, and papers in the population at time n.

(1) Do these numbers approach a well defined limit as $n\to\infty$, or do they oscillate in some unpredictable way? In particular, will one of the pieces go 'extinct' after a while?
(2) How do these possible asymptotic behaviors depend on the initial configuration of the game?
(3) We have defined the game by a number of random procedures. If we modify the probability parameters, how are the possible asymptotic behaviors affected?

Project 2. *Using your favorite math software (Matlab, Maple, Mathematica, etc.), simulate the stone-paper-scissors game and describe any interesting observation about how $\#_S(n)$, $\#_R(n)$, and $\#_P(n)$ vary with n.*

Here the reactor is the checker-board, which we can view as either an isolated system if no pieces are being added or taken out from the game, or as an open system if we allow "fluxes" in and out, through the sides. Notice that we may be interested not simply with changes in the numbers of S,P,R in time, but with their spatial distribution as well. The random motion described above (without regard to reaction) is a discrete approximation of a diffusion process. Thus in this case we are concerned with a reaction-diffusion system.

We could think of this game as representing a mixture of three interacting gases. More realistic, but similar, models of chemical (isomerization) processes will be shown later.

Although this example is discrete both in space and in time, and the rules of transformation are defined probabilistically, we will for the most part in this course study continuous systems with deterministic rules specified by differential equations. In the continuous version of the problem, the quantities $\#_S(n)$, $\#_R(n)$, and $\#_P(n)$ are replaced with the densities per unit area at time t and the change in time, say $\#_S(n+1)-\#_S(n)$, is replace with a time derivative.

The continuous version of this game has the following mathematical description. Let C_R,C_P,C_S denote the concentrations of R, P and S over a square with coordinates x,y. Let ∇^2 denote the two-dimensional Laplacian. Then

$$\begin{aligned}\frac{\partial C_P}{\partial t} &= K(C_R-C_S)C_P + D\nabla^2 C_P\\ \frac{\partial C_R}{\partial t} &= K(C_S-C_P)C_R + D\nabla^2 C_R\\ \frac{\partial C_S}{\partial t} &= K(C_P-C_R)C_S + D\nabla^2 C_S.\end{aligned}$$

The constants K and D depend on the precise way in which the passage from discrete to continuous was taken.

3.4. **Evolutionary Game Theory.** The previous example can be studied from the perspective of *evolutionary game theory*, and is of interest to such areas as evolutionary and population biology,

ecology and economics. The SRP-game is treated, for example, in *Evolutionary Games and Population Dynamics*, by J. Hofbauer and K. Sigmund, Cambridge University Press, 1998; *Evolutionary Game Theory*, by J. W. Weibull, The MIT Press, 1995.

3.5. **Epidemic Models.** The following is a simplified form of a mathematical model describing the course of an epidemic. It was proposed and analyzed by Kermack and McKendrick in the 1930's. A population that has been exposed to some contagious disease is divided into three groups: those individuals *susceptible* of contracting the disease, denoted S; *infective* individuals, I; and *removed* individuals, R. Individuals of type S may change into type I, through contagion, while those of type I may change into type R, through death, recovery, or isolation from the rest of the population. Thus we have the following reactions:

(Infection) $$S + I \rightharpoonup 2I$$

(Removal) $$I \rightharpoonup R.$$

Exercise 3. *Devise rules for a checkerboard game that simulates the process of epidemic spread described above.*

There are many different ways to give mathematical flesh such set of equations. One way is to construct a probabilistic model along the lines of the rock-paper-scissors game. A deterministic model is obtained as follows. Denote by the same letters, S, R, and I, variables with (continuous) value range between 0 and 1, representing the fractions of the total population in the respective categories.

We interpret the infection reaction by the rule that S decreases in time at a rate proportional to the product of S and I, with proportionality constant σ; I increases due to the infection reaction at the same rate, and decreases due to removal at a rate proportional to I with constant ρ; R increases due to removal at the same rate as I decreases. Thus we have the following system of ordinary differential equations, where '$'$' denotes time derivative:

$$\begin{aligned} S' &= -\sigma SI, \\ I' &= \sigma SI - \rho I, \\ R' &= \rho I. \end{aligned}$$

Exercise 4. *Applying the coordinate changes*

$$x = \frac{\sigma}{\rho} S,\ y = \frac{\sigma}{\rho} I,\ z = \frac{\sigma}{\rho} R,\ \tau = \rho t$$

show that the differential equations reduce to

$$\begin{aligned} x' &= -xy \\ y' &= (x-1)y \\ z' &= y. \end{aligned}$$

Also show that the quantities $x+y+z$ and $z+\ln x$ do not change in time.

For much more on this subject, see *Mathematics in Population Biology*, by H. R. Thieme. Princeton Series in Theoretical and Computational Biology, Princeton University Press, 2003.

3.6. **Population Biology.** *A Primer of Population Biology*, by E. O. Wilson and W. H. Bossert. Sinauer, 1971.

The growth of a population of bacteria might be described by a simple minded model defined by the following reactions:

$$A \rightleftharpoons 2A.$$

The direct reaction represents the asexual reproduction of the organism whereas the reverse reaction might represent death due to resources competition. (Reactions that involve replication, as indicated by $A \rightharpoonup 2A$, are characteristic of life.)

Assuming that the reaction rates are given by the mass-action law (this will be explained later) the number density, u, of bacteria (number per unit area) satisfies the differential equation:

$$\frac{du}{dt} = k_1 u - k_2 u^2,$$

where k_1 and k_2 are the rate constants for the direct and reverse reactions, respectively. This (easily solvable) differential equation is called the *logistics equation*.

Suppose now that the population is allowed to disperse over a laboratory dish. If instead of bacteria we were considering small particles of some inert material, dispersion by random motion (brownian motion) would be described the *diffusion equation*:

$$\frac{\partial u}{\partial t} = D\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right).$$

(We will have much more to say about the diffusion equation later in the course.) The right-hand side of the equation is often written $D\nabla^2 u$, where ∇^2 is the Laplace operator in dimension 2.

For the bacteria population subject to reproduction, death, and random dispersal, its number density satisfies the reaction-diffusion equation

$$\frac{\partial u}{\partial t} = k_1 u - k_2 u^2 + D\nabla^2 u.$$

Exercise 5. *Show that by a suitable coordinate change we can write the above equation as*

$$\frac{\partial w}{\partial t} = sw(1-w) + k\nabla^2 w,$$

where s and k are positive constants. This is known as the Fisher equation. *Find a space-homogeneous solution. (That is, a solution that depends only on t and not on x, y.)*

Much more sophisticated models are used in population genetics. This simple equation, however, already exhibits some interesting behavior such as stationary waves. We'll return to this later.

3.7. **Environmental Science.** *Consider a Spherical Cow*, John Harte. We will discuss a number of processes of interest to environmental sciences very early in the course.

3.8. **Catalytic and Enzymatic Processes.** *Introduction to Mathematical Biology*, Rubinow.

One of the most basic enzymatic reactions, first proposed by Michaelis and Menten (1913), involves a substrate S reacting with an enzyme E to form a complex SE, which is then converted into a product P and the enzyme. Schematically, this is represented by

$$S + E \rightleftharpoons SE$$
$$SE \rightharpoonup P + E.$$

(This is the same reaction scheme as for the student-teacher catalytic process suggested earlier.)

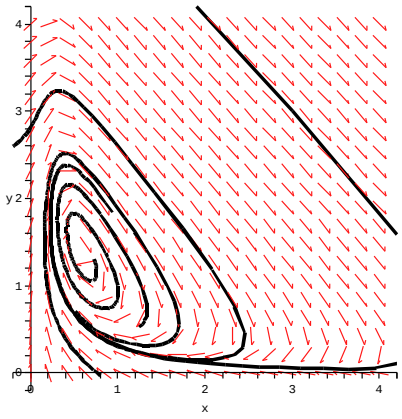
3.9. **Spontaneous Generation of Spatial Patterns - Turing Models.** *Mathematical Biology*, J. D. Murray.

3.10. **What is a 'Dynamical System'?** The differential equations that arise in chemical kinetics are often non-linear and do not admit the relatively simple and complete description of linear ODEs with which you may be familiar from Math 217. There are many new qualitative features displayed by solutions of non-linear equations that are not present in linear equations, and the systematic exploration of those new features require more sophisticated mathematical ideas. As a simple example, consider the following system of equations. They model a fundamental biochemical process called *glycolysis*, by which living cells obtain energy by breaking down sugar.

$$x' = -x + ay + x^2 y$$
$$y' = b - ay - x^2 y.$$

Here x and y are the concentrations of ADP (adenosine diphosphate) and $F6P$ (fructose-6-phosphate).

The following picture shows a typically non-linear phenomenon. (For this numerical example $a = 0.08$ and $b = 0.6$.)



Notice that trajectories are attracted to a limit cycle from either inside or outside. (Inside there is an unstable equilibrium point.) This shows that, according to this model, the concentrations of the two substances during glycolysis oscillate. In dimension 3 the structure of the attracting set can be considerably more complicated, and the trajectories can behave in fairly "chaotic" way.

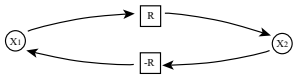
The theory of dynamical systems is concerned with the possible behaviors of trajectories of flows as in this case or discrete systems. Often, as in this example, we are given not a single systems, but a family of systems depending on a number of parameters (a and b in this case). We will be interested in finding out how the systems properties change with the parameters. This consideration will lead us to study so-called *bifurcations*.

Our main sources for the theory of dynamical systems will be: *Differential Equations and Dynamical Systems*, by L. Perko; and *Nonlinear Dynamics and Chaos*, by S. H. Strogatz.

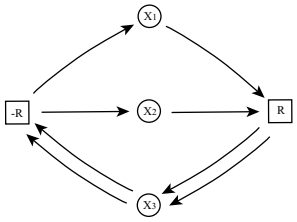
The above picture was produced with MAPLE using the following commands:

```
>with(plots);
>with(DEtools);
>eq1:= diff(x(t),t)=-x(t) + a*y(t) + x(t)^2*y(t);
>eq2:= diff(y(t),t)=b- a*y(t) - x(t)^2*y(t);
>a:=0.08;
>b:=0.6;
>ini1:=x(0)=0.4, y(0)=0.4;
>ini2:=x(0)=3, y(0)=3;
>ini3:=x(0)=0.1, y(0)=3;
>ini4:=x(0)=1,y(0)=1;
>DEplot({eq1, eq2}, [x(t),y(t)], -10..10,[[ini1], [ini2],[ini3], [ini4]],
 stepsize=0.1, x=0..4, y=0..4);
```

3.11. **Graph Representation of Reaction Mechanisms - I.** Mechanisms for complex chemical reactions can be represented by graphs in a number of different ways. One way is to draw the reactions and substances as nodes of a graph, with arrows connecting each substance to reactions of which it is a reactant, and arrows from each reaction to its products. The nodes are thus of two different kinds. We represent reactions by square and substances by circles. Here are a few examples. In each case, R represents the direct reaction ($\rightharpoonup$). The number of arrows between two nodes indicates how many molecules of a given type are either used or produced in a reaction.

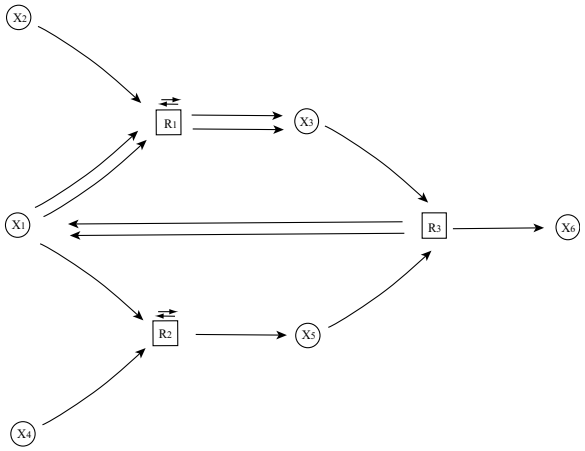


$$R : X_1 \rightleftharpoons X_2.$$



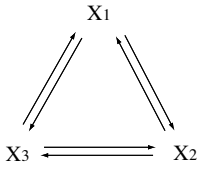
$$R : X_1 + X_2 \rightleftharpoons 2X_3.$$

So as not to overcrowd the picture we may collapse a pair of forward-back reactions by drawing only the forward reaction and add near or inside the square two opposite arrows, as in the figure below.



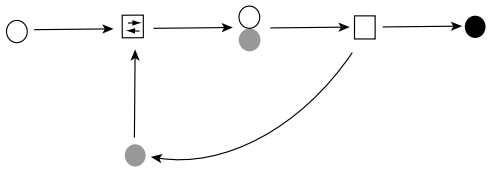
$$R_1 : 2X_1 + X_2 \rightleftharpoons 2X_3$$
$$R_2 : X_1 + X_4 \rightleftharpoons X_5$$
$$R_3 : X_3 + X_5 \rightharpoonup X_6 + 2X_1$$

When all reactions are monomolecular, that is, if each has only one reactant and one product species, we will typically omit the reaction squares. Thus we may draw:

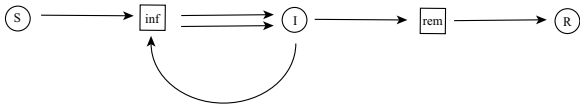


$$R_1 : X_1 \rightleftharpoons X_2$$
$$R_2 : X_2 \rightleftharpoons X_3$$
$$R_3 : X_3 \rightleftharpoons X_1.$$

Here are the graphs for some of the reactions described earlier: The graph of the student-teacher reaction might be drawn as



A graph for the epidemic reaction model might be:



Exercise 6. *Draw a graphical representation for the reaction mechanism of the SRP-game.*

4. Steady-State Box Models Without Reactions

This and the next few sections are shamelessly lifted from the excellent *Consider a Spherical Cow*, by John Harte. All data needed can be found in the appendix of that book.

Here is a reminder of some basic chemical terms. *Avogadro's number* (N) is defined as the number of carbon-12 atoms in exactly 12 g of carbon 12. Its value is approximately 0.60229×10^{24}. A *mole* of a substance is defined as Avogadro's number of molecules of the substance. Thus a mole of water is the quantity of water containing N molecules of H_2O. The molecular weight of water is $2 \times 1.01 + 1 \times 16.00 = 18.02$. It follows from the definitions that one mole of H_2O is 18.02 g. One mole of sulfur S weighs 32.06 g.

4.1. **Sulfur in Coal.** This problem is intended for gaining familiarity with the measurement of quantity of matter in units of mass and in moles.

Problem 7. *How many tonnes (metric tons) and how many moles of sulfur were contained in the coal consumed worldwide in 1980?*

In the appendix Harte's book we find the following information: the consumption of coal worldwide in 1980, in energy (heat) units, was 90×10^{18} joules (the number was 15×10^{18} J for the United States alone); the energy content of coal is 29.3×10^6 J/Kg; sulfur (S) makes up 2.5% of coal composition in weight.

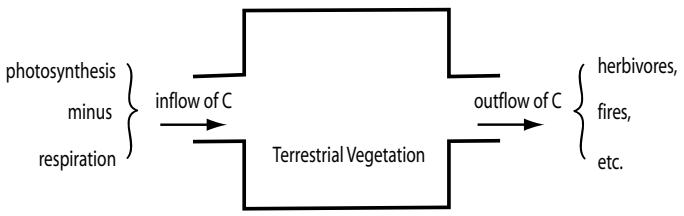
The amount of coal in metric tons consumed in 1980 can be calculated as the quotient

$$\frac{\text{energy derived from coal in 1980 in joules}}{\text{energy content per unit mass of coal in joules per tonne}}.$$

This number is 3.1×10^9 tonnes. Of this amount, sulfur makes up 2.5%, which gives 7.7×10^7 tonnes of S.

If one mole of S weighs 32.06 grams, one metric ton of S contains $10^6/32.06$ moles, or 3.1×10^4 moles. Multiplying this number by the amount of sulfur in tonnes gives 2.4×10^{12} moles of S.

4.2. **Carbon in the Biosphere.** Recall from the example of a school, given in class, the following concepts: for a system in steady-state the inflow and outflow of a given substance are equal and constant in time. The total amount of that substance in the system is the substance's total stock, and the residence time of that substance in the system is the total stock divided by the inflow.



Suppose that the system is the living continental (as opposed to marine) vegetation and that we are interested in the flow of carbon through it. The inflow of carbon from the atmosphere is mainly due to photosynthesis, and is called *gross primary productivity*. If we subtract from it the carbon loss due to respiration by the vegetation we obtain the *net primary productivity* of the system. The residence time of carbon in the system, obtained using the net primary productivity as inflow, serves as an estimate of the average life span of plants.

Problem 8. *Calculate the residence time of carbon in continental vegetation.*

From the appendix of Harte's book: the net primary productivity of continental vegetation is $(50 \pm 15) \times 10^{12}$ Kg of carbon per year. It also quotes the continental living biomass stock as $500(+300, -100) \times 10^{12}$ Kg of carbon. This number does not seem to distinguish between animal and vegetable biomass, but I'll make the assumption that most of the continental biomass is vegetable. Thus the residence time is

$$\frac{\text{stock of living continental biomass}}{\text{continental net primary productivity}}.$$

This gives a number anywhere between 7 and 23 years.

4.3. **A Polluted Lake.** The main concept here is *steady-state concentration*.

Problem 9. *A stable and highly soluble pollutant is dumped into a lake at the rate of* 0.16 *tonnes per day. The lake volume is* 4×10^7 *cubic meters and the average water flow-through rate is* 8×10^4 *cubic meters per day. Ignore evaporation from the lake surface and assume the pollutant is uniformly mixed in the lake. What eventual steady-state concentration will the pollutant reach?*

We need first to calculate the stock of pollutant, M_p, in the lake. This is the product $F_p \times T_p$, where F_p is the rate at which the pollutant is being dumped in the lake and T_p is the pollutant residence time. This residence time is equal to the water residence time, T_w, since water and pollutant are uniformly mixed. But

$$T_w = \frac{M_w}{F_w} = \frac{4 \times 10^7 \text{ m}^3}{8 \times 10^4 \text{ m}^3/\text{day}} = 500 \text{ days}.$$

Therefore

$$M_p = F_p \times T_p = 0.16 \text{ tonnes per day} \times 500 \text{ days} = 80 \text{ tonnes}.$$

The amount of water in the lake in weight is the volume times water density (= 1000 kg per cubic meter):

$$4 \times 10^7 \text{ m}^3 \times 1 \text{ metric ton per m}^3 = 4 \times 10^7 \text{ tonnes}.$$

Thus the steady-state concentration of the pollutant is

$$\frac{80 \text{ tonnes}}{4 \times 10^7 \text{ tonnes}} = 2.0 \times 10^{-6}.$$

This is 2 parts per million by weight.

Aqueous concentrations are often specified in units of molarity, or moles per liter. Suppose the pollutant has a molecular weight of 40 (that is, there are a total of 40 protons and neutrons in the atoms of each molecule). Then the number of moles of pollutant is the weight in grams divided by 40 g:

$$\text{number of moles} = \frac{\text{weight in grams}}{\text{molecular weight}} = \frac{80 \times 10^6}{40} = 2.0 \times 10^6 \text{ moles}.$$

The volume of water in liters is 4×10^{10} liters. So the molarity of pollutant is 50×10^{-6} moles per liter, or 50 micromoles per liter. One mole per liter is called a *molar*, and designated M. We can write the solution as 50μM (micromolars).

You may have encountered similar problems in Math 217 (differential equations) if you took it before. Let us stop to look at how the same problem is approached in that course. Let x denote the concentration of pollutant (in tonnes per cubic meter). The rate of change of the amount of pollutant, which is the time derivative of $M_w x$, is the difference between the inflow (0.16 tonnes per day) and the outflow of pollutant. The latter is x times the outflow of water measured in volume per time: $8\times10^4 x$. Therefore the concentration x satisfies the differential equation:

$$4\times10^7\dot{x} = 0.16 - 8\times10^4 x.$$

In steady-state, the concentration is constant in time so $\dot{x}=0$, and we arrive at

$$x = \frac{0.16}{8\times10^4} = 2.0\times10^{-6} \text{ tonnes per cubic meter.}$$

If instead of steady-state condition we assume that the lake water is initially clean, then $x(0)=0$, and we have the initial value problem:

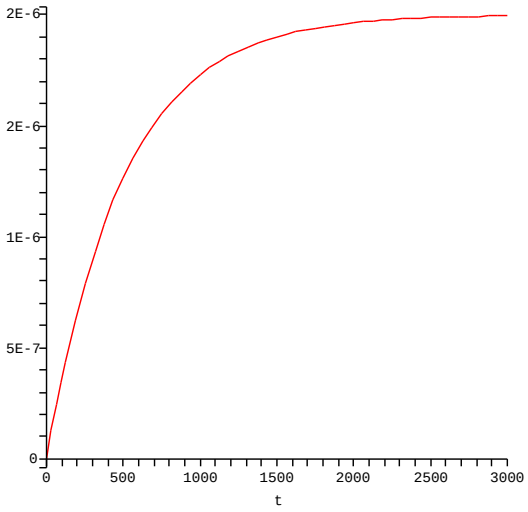
$$\dot{x} = ax+b$$
$$x(0) = x_0$$

where $a=-2\times10^{-3}, b=4\times10^{-9}$, $x_0=0$.

Exercise 10. *Show that the solution to the above initial value problem is:*

$$x(t) = \frac{b}{a}\left(e^{at}-1\right) + x_0 e^{at}.$$

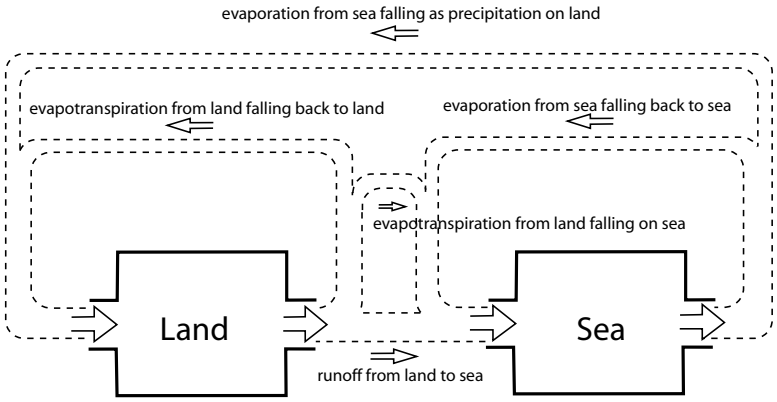
What happens to $x(t)$ as $t\to\infty$?



The above graph of the concentration function $x(t)$ was obtained using the MAPLE commands:

```
>f:= t->2*10^(-6)*(1-exp(-2*10^(-3)*t));
>plot(f(t), t=0..1000);
```

4.4. **The Global Hydrocycle.** Here we are concerned with the flows of water between land and sea. A simplified global water budget is described schematically in the next figure.



Land water that evaporates from rivers and lakes as well as water from transpiration of the vegetation cover (evapotranspiration) falls as precipitation, partly back to land and partly to sea. Sea water evaporates and falls as precipitation, partly back to sea and partly to land. River water flows from earth to sea (runoff).

Problem 11. *If evapotranspiration from Earth's land area were to diminish by* 20% *uniformly over the land area, as might result from widespread removal of vegetation, what changes would occur in the globally averaged precipitation on the land surface and in the globally averaged runoff from the land to the sea?*

It will be helpful to introduce the following rates:

$$
\begin{aligned}
e_{LL} &= \text{ rate of evapotranspiration from land that falls as precipitation on the land;}\\
e_{LS} &= \text{ rate of evapotranspiration from land that falls as precipitation on the sea;}\\
e_{SL} &= \text{ rate of evaporation from sea that falls as precipitation on the land;}\\
e_{SS} &= \text{ rate of evaporation from sea that falls as precipitation on the sea;}\\
r &= \text{ rate of runoff from the land to the sea;}\\
p_L &= \text{ rate of precipitation on the land;}\\
p_S &= \text{ rate of precipitation on the sea.}
\end{aligned}
$$

Then our problem is to find out how r and p_L will change if e_{LL} and e_{LS} are reduced by 20%.

Here are some numerical values available in the appendix of Harte's book:

$$\begin{aligned}
\text{world evapotranspiration from land} &= 62 \times 10^{12} \text{ cubic meters per year;} \\
\text{world evaporation from sea} &= 456 \times 10^{12} \text{ cubic meters per year;} \\
\text{world precipitation on sea} &= 410 \times 10^{12} \text{ cubic meters per year;} \\
\text{world precipitation on land} &= 108 \times 10^{12} \text{ cubic meters per year;} \\
\text{world runoff} &= 46 \times 10^{12} \text{ cubic meters per year.}
\end{aligned}$$

This means that, in units of 10^{12} cubic meters per year:

$$\begin{aligned}
\text{(eq. 1)} \qquad & e_{LL} + e_{LS} = 62 \\
\text{(eq. 2)} \qquad & e_{SL} + e_{SS} = 456 \\
\text{(eq. 3)} \qquad & e_{LS} + e_{SS} = 410 \\
\text{(eq. 4)} \qquad & e_{LL} + e_{SL} = 108. \\
\text{(eq. 5)} \qquad & r = 46
\end{aligned}$$

There are here 5 equations and 5 unknowns, but notice that these are not all independent equations. In fact, it can be checked that

$$(\text{eq. } 1) + (\text{eq. } 2) = (\text{eq. } 3) + (\text{eq. } 4).$$

This is to be expected since all the water that evaporates eventually falls back as precipitation. We need to find another relationship among these rates. For that purpose we will make the assumption that three-fourths of the evapotranspiration from the land precipitates back on the land. (Is this a reasonable estimation?) In other words, by a rough estimate we have:

$$\text{(eq. 6)} \qquad e_{LL} = 3e_{LS}.$$

Solving equations 1 to 6 we obtain:

$$\begin{aligned}
e_{LS} &= 15.5 \\
e_{SS} &= 394.5 \\
e_{SL} &= 61.5 \\
e_{LL} &= 46.5 \\
r &= 46.
\end{aligned}$$

There are several relations among these quantities that must always hold, even after some hypothetical change in the evapotranspiration rate. First, by definition we have:

$$\begin{aligned}
p_L &= e_{LL} + e_{SL} \\
p_S &= e_{LS} + e_{SS}.
\end{aligned}$$

Also, under the steady-state assumption, the total stocks of water on sea and on land remain constant. Conservation of the sea water gives the equation:

$$p_S + r = e_{SS} + e_{SL};$$

and conservation of water on land gives the equation:

$$p_L = r + e_{LL} + e_{LS}.$$

If evapotranspiration from earth decreases by 20%, we have the following changes (primed letters represent the new values):

$$e'_{SS} = e_{SS};$$
$$e'_{SL} = e_{SL};$$
$$e'_{LL} = 0.8e_{LL};$$
$$e'_{LS} = 0.8e_{LS}.$$

A little algebra now gives

$$p'_L - p_L = -0.2e_{LL} = -9.3 \times 10^{12} \text{ cubic meters per year;}$$
$$r' - r = \ 0.2e_{LS} = 3.1 \times 10^{12} \text{ cubic meters per year.}$$

The result is that r increases by about 7% and p_L decreases by about 9%.

4.5. **Cooling off.** We have so far considered box models involving matter flows. A similar accounting of inflow and outflow of energy instead of matter is what thermodynamics is mainly about. For a very nice and good humored overview of classical thermodynamics see *Understanding Thermodynamics*, by H. C. Van Ness. Dover, 1983.

Problem 12. *At what rate is water used to cool a* 1000*-megawatt coal-fired power plant?*

We first need to have some idea of how a power plant operates. (See the figure on the next page.) There are four primary devices in the power cycle. The boiler serves to convert liquid water into steam at a high pressure and a high temperature. This requires heat from some high-temperature source. The steam is fed to a turbine which drives an electric generator. Through the turbine steam expands causing the turbine blades to rotate, and then exhausts at a low pressure. At this point, heat has been converted to mechanical energy.

The next step is the conversion of mechanical energy to electric. This is accomplished using the principle that a variable magnetic field generates current on a metal wire immersed in that field. To take advantage of this fact, the rotating turbine shaft is wrapped with conducting wires and surrounded by a strong magnetic field. This is what takes place inside the generator.

There is no limit in principle to how efficiently mechanical energy can be converted to electric energy. On the other hand, there is a fundamental thermodynamical limit on the conversion of heat into mechanical or any other type of work, which was discovered in 1824 by Carnot. It says, in the present context, that the maximum efficiency of the heat-to-mechanical energy conversion is given by the equation

$$\eta_{\max} = \frac{T_H - T_C}{T_H},$$

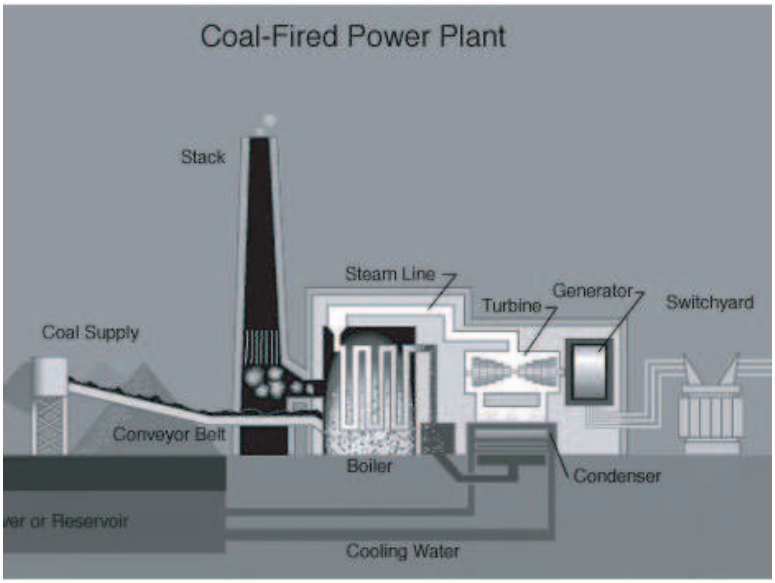
where the hotter temperature, T_H, is the temperature of the boiler and the cooler temperature, T_C, is the temperature of the condenser. Therefore the greater the gap of temperature across the turbine the greater the efficiency at which it can operate.

The actual efficiency of such a power plant will be less than the ideal, $\eta_{\max}$. Inefficiencies in both conversion processes, resulting from mechanical friction, electrical resistance, loss of heat up the smoke stack, etc., result in energy loss.

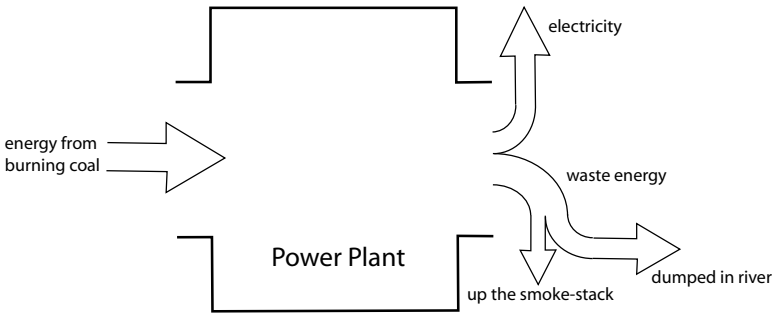
For a typical coal-fired electric generating plant T_H is about 800 K, and T_C is about 300 K. Thus the ideal efficiency of conversion is $\eta_{\max} = 500/800 = 62.5\%$. The actual conversion efficiency for

a modern coal-fired electricity generating plant is about 40%, which means that

$$\frac{\text{electric power output}}{\text{rate of heat input from coal burning}} = 0.40.$$



(The above figure is taken from the website `http://www.tva.gov/power/coalart.htm` maintained by the Tennessee Valley Authority.) The energy budget of our power plant can be described as follows. Energy obtained from burning coal is partly transformed into electric energy and partly wasted as heat going up the smoke-stack as well as heat that is transferred to the river water as it is used to keep the condenser at its operating temperature of 300 K. A box diagram is shown next.



The difference between the heat input and the electric output is waste heat. The rate at which this waste heat is produced, R, is

$$\begin{aligned} R &= (\text{rate of heat input}) - (\text{reate of electrical energy output}) \\ &= \frac{(\text{electric power output})}{0.40} - (\text{electric power output}) \\ &= 1.5 \times (\text{electric power output}) \\ &= 1.5 \times 1000 \text{ megawatt} \\ &= 4.7 \times 10^{16} \text{joules per year.} \end{aligned}$$

(We have used above that 1 joule equals 1 watt-sec.)

Typically, about 15% of this waste heat is removed via the smoke-stack, in the form of hot effluent gases. The remaining 85% is the heat that must be discharged from the turbine by some cooling process. This amounts to 4.0×10^{16} joules per year.

We are assuming that for the cooling process, cool water flowing past the power plant and through the turbine condenser is warmed by the waste heat and then discharged to the environment. Let us say that cooling water enters at an average temperature of 290 K. Then it will typically be heated by about 10 K (to 300 K) as it passes through the turbine condenser.

We can now calculate the rate at which water must flow through the system. The key fact is this: it takes 1 cal, or 4.18 joules, to heat 1 gram of water 1 K. Thus 4.0×10^{16} joules per year will heat $\frac{4.0 \times 10^{16}}{4.18} = 9.6 \times 10^{15}$ grams of water 1 K per year. If the water is 10 K hotter after passing through the condenser, then we need 9.6×10^{14} grams of water per year. This amounts to about 30 cubic meters of water per second.

4.6. **Stoichiometry of Burning Fossil Fuels.** Stoichiometry is concerned with the relative amounts of the elements that constitute a chemical substance. The expressions H_2O and $HO_{0.5}$ are equivalent stoichiometric formulas for water. Note that one mole of H_2O weighs $2+16=18$ grams, whereas one mole of $HO_{0.5}$ weighs $1+8=9$ grams.

Problem 13. *In 1980, how much O_2 was removed from the atmosphere due to the combustion of fossil fuel on Earth, and how much CO_2 and H_2O were produced in the combustion process?*

The primary elemental constituents of the three majors types of fossil fuels (natural gas, petroleum, and coal) are carbon and hydrogen. When fossil fuel is burned, oxygen from the atmosphere combines with the carbon to make CO_2 and with the hydrogen to make H_2O. In addition, coal contains some water (typically 10-15% by weight), which is released to the atmosphere upon combustion of the coal. All three types of fossil fuel contain various other substances such as ash, sulfur, and trace metals in even lower concentrations.

We need to consider petroleum, coal, and natural gas separately.

Here is some pertinent information from the appendix of Harte's book regarding petroleum (crude): 98% of its weight consists of $CH_{1.5}$; the world consumption in 1980 was 135×10^{18} joules, and its heat content is 43×10^6 joules per kilogram.

The combustion of $CH_{1.5}$ follows the overall reaction

$$CH_{1.5} + xO_2 \rightharpoonup yCO_2 + zH_2O,$$

where x, y, z are the stoichiometric constants, which we need to determine. These constants provide an answer to the problem: for each mole of $CH_{1.5}$ x moles of O_2 is consumed, and y and z moles of CO_2 and H_2O, respectively, are produced.

To determine x,y,z, we equate the number of moles of each element, C,O,H, on the left and right sides of the reaction:

(C) $$1 = y$$

(O) $$2x = 2y + z$$

(H) $$1.5 = 2z.$$

This gives

$$CH_{1.5} + 1.375O_2 \rightharpoonup CO_2 + 0.75H_2O.$$

To determined the number of moles of $CH_{1.5}$ burned worldwide in 1980 we calculate:

$$\text{weight fraction of } CH_{1.5} \text{ in crude composition} \times \frac{\text{world consumption (joules)}}{\text{heat content (joules per kilogram)}}.$$

This number is

$$0.98 \times \frac{135 \times 10^{18}}{43 \times 10^6} \text{ kilograms} = 3.08 \times 10^9 \text{ tonnes}.$$

Since one mole of $CH_{1.5}$ weighs $12 + 1.5 = 13.5$ grams, the amount of $CH_{1.5}$ burned is

$$\frac{3.08 \times 10^{15} \text{ grams}}{13.5 \text{ grams}} = 2.28 \times 10^{14} \text{ moles of } CH_{1.5}.$$

Therefore, the amounts of O_2, CO_2, and H_2O consumed or released into the atmosphere due to petroleum burning are:

$$n(O_2) = 1.375 \times 2.28 \times 10^{14} = 3.14 \times 10^{14} \text{ moles (consumed)}$$
$$n(CO_2) = 2.28 \times 10^{14} \text{ moles (released)}$$
$$n(H_2O) = 0.75 \times 2.28 \times 10^{14} = 1.71 \times 10^{14} \text{ moles (released)}.$$

We repeat the analysis, now for natural gas. The composition of natural gas in molar fractions is: CH_4 (75%), C_2H_6 (6%), C_3H_8 (4%), C_4H_{10} (2%), and C_5H_{12} (1%). The remaining 12% is noncombustible. Thus one mole of natural gas consists of:

$$0.75 \times 1 + 0.06 \times 2 + 0.04 \times 3 + 0.02 \times 4 + 0.01 \times 5 = 1.12 \text{ moles of } C$$
$$0.75 \times 4 + 0.06 \times 6 + 0.04 \times 8 + 0.02 \times 10 + 0.01 \times 12 = 4 \text{ moles of } H.$$

Thus, the effective formula for the combustible portion of natural gas is $C_{1.12}H_4$, and this makes up 88% of natural gas in moles. After finding the stoichiometric constants as before, we obtain the reaction formula

$$C_{1.12}H_4 + 2.12O_2 \rightharpoonup 1.12CO_2 + 2H_2O.$$

The world consumption of natural gas in 1980 was 60×10^{18} joules and the heat content at standard temperature and pressure is 3.9×10^7 joules per cubic meter. This gives 1.54×10^{12} cubic meters.

One mole of any gas at standard temperature and pressure comprises 22.4 liters, or 22.4×10^{-3} cubic meters. We will return to this point later. Accepting it for now, we obtain that 1m^3 of any gas at STP contains 44.6 moles. Therefore, the world consumption of natural gas in moles was $44.6 \times 1.54 \times 10^{11} = 6.9 \times 10^{13}$ moles of natural gas. Multiplying by 0.88 (the combustible fraction)

gives 6.0×10^{13} moles of $C_{1.12}H_4$. It follows that the amounts of O_2, CO_2 and H_2O consumed or released by burning natural gas are:

$$n(O_2) = 2.12 \times 6.0 \times 10^{13} = 1.27 \times 10^{14} \text{ moles (consumed)}$$
$$n(CO_2) = 1.12 \times 6.0 \times 10^{13} = 0.67 \times 10^{14} \text{ moles (released)}$$
$$n(H_2O) = 2 \times 6.0 \times 10^{13} = 1.20 \times 10^{14} \text{ moles (released).}$$

I leave the calculation for coal to you. Notice that 13% of coal is water, which is liberated to the atmosphere upon combustion and must be included in the calculation. The result for coal is

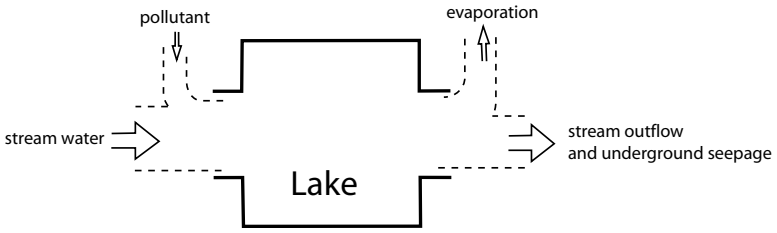
$$n(O_2) = 2.16 \times 10^{14} \text{ moles (consumed)}$$
$$n(CO_2) = 1.80 \times 10^{14} \text{ moles (released)}$$
$$n(H_2O) = 0.94 \times 10^{14} \text{ moles (released).}$$

4.7. **Pollution Buildup in a Lake.** We have so far considered systems in steady state. If this condition is not present, then the stock, M, of a substance will grow or decrease in time at a rate equal to

$$\frac{dM}{dt} = F_{\text{in}} - F_{\text{out}}.$$

Problem 14. *A lake has a volume of* 10^6 *cubic meters and a surface area of* 6×10^4 *square meters. Water flows into the lake at an average rate of* 0.005 *cubic meters per second. The amount of water that evaporates yearly from the lake is equivalent in volume to the lake's top meter of water. Initially, the lake water is pristine, but at a certain time a soluble, noncodistilling pollutant is discharged into the lake at a steady rate of* 40 *tonnes per year. Derive a formula for the concentration of pollutant in the lake as a function of time since the pollutant discharge began.*

A *noncodistilling* pollutant is a substance that does not evaporate away with evaporating water. Therefore, evaporation of lake water is not an exit pathway for the pollutant. However, if the lake water flows out of the lake in an outlet stream or via underground seepage, that water outflow will remove pollutant.



The volume of the lake is assumed constant, so the rate of outflow stream and seepage must equal the difference between the rate of inflow and rate of water lost to evaporation. (We disregard the change in volume due to the pollutant. To get a sense of how much error is involved in this approximation, an amount of water of same weight as the pollutant added in a year, 40 tonnes, would occupy a volume of 40 cubic meters, which is less than 0.1% of the 60000 cubic meters of water that evaporates in a year.) So we have the following information (recall that 1 cubic meter of water weighs 1 metric ton):

$$\text{volume of lake} = 10^6 \text{m}^3$$
$$\text{rate of evaporation} = 6 \times 10^4 \text{m}^3/\text{yr}.$$
$$\text{inflow of water} = 0.005 \times (365 \times 24 \times 60 \times 60) = 1.6 \times 10^5 \text{m}^3/\text{yr}.$$

outflow stream and under ground seepage $= 1.6 \times 10^5 - 6 \times 10^4 = 10^5 \text{m}^3/\text{yr}$.

Let $x(t)$ represent the concentration of pollutant in the lake at time t as a fraction of weight. We set $t = 0$ as the time when pollutant started to be dumped in the lake. Thus $x(t) = M(t)/M_w$, where $M(t)$ is to total amount of pollutant (in tonnes), M_w is the total amount of water (in tonnes), which is 10^6 tonnes, and $x(0) = 0$. The flows of pollutant are

$$F_{\text{in}} = 40\text{tonnes per year}$$
$$F_{\text{out}} = \text{outflow of water by stream and seepage} \times \text{fraction of pollutant} = 10^6 \times x(t).$$

Therefore,

$$\frac{dx}{dt} = 4 \times 10^{-5} - 0.1x.$$

The solution to the initial value problem $\dot{x} = ax + b$, $x(0) = x_0$, was given in an earlier exercise. It is:

$$x(t) = \frac{b}{a}(e^{at} - 1) + x_0 e^{at}.$$

In the present case,

$$x(t) = 4.0 \times 10^{-4}(1 - e^{-0.1t}).$$

Note that, at equilibrium ($t \to \infty$) x approaches 400 parts per million.

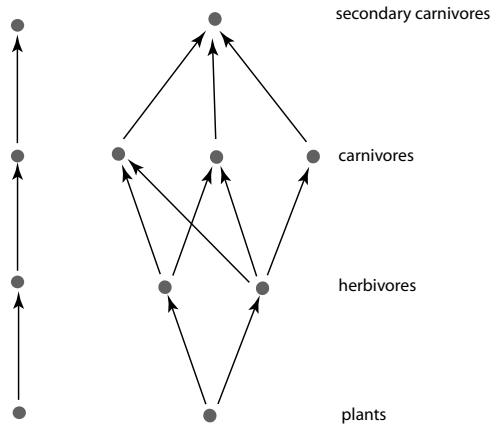
4.8. **Biomagnification of Trace Substances.**

Problem 15. *How does biomagnification of a trace substance occur? Specifically, identify the critical ecological and chemical parameters determining bioconcentrations in a food chain. In terms of these parameters, derive a formula for the concentration of a trace substance in each link of a food chain.*

Imagine a young fish growing up in a lake and eating nothing but plankton. As the fish grows from, say, 10 g at birth to its average weight at death of 10^3 g, all of its newly acquired flesh and bone will be derived from the plankton it eats. To add on 990 g of tissue, it will have to eat far more than 990 g of plankton, because of excretory and metabolic losses. Excretory losses, in the sense used here, include the sloughing off of old tissue as it is replaced with new. The typical growing fish eats 10 g of plankton to grow by about 1 g of body weight. In the ecological literature this factor of 10% (the ratio of the weight it gains to the weight of food it eats) is called an *incorporation efficiency*. The remaining 90% of the food the fish eats is excreted or metabolized.

Suppose, now, that the plankton contains a trace substance, such as DDT or mercury. If a greater proportion of the trace substance than of the plankton is retained in the fish rather than excreted and metabolized, then the concentration of the trace substance in the animal will build up to a level greater than that in the plankton. Similarly, if a population of osprey prey upon the fish, they too may preferentially retain the trace substance in their food; over the lifetime of the osprey, the concentration of the substance can build up to an even greater level than that in the fish. The fraction of ingested trace substance retained by an organism is called the *retention factor*.

The next figure shows two food webs with four trophic levels (not counting the sun). The linear web is a food chain.



We would like to devise a mathematical model to study the magnification of trace substances along the various trophic levels of a food chain. The model should incorporate the following features.

(1) **Preferential Retention of the Substance in the Body.** In the extreme case, the rate of bioconcentration is greatest when all the trace substance an animal ingests is stored in body tissue and none is metabolized or excreted.

(2) **Fraction of Ingested Food Incorporated into New Tissue.** If an animal is very inefficient in building new tissue out of its food source, it must consume a lot of food to grow by any specified amount. That larger amount of ingested food is accompanied by a larger amount of trace substance. Thus, for a given retention factor, the lower the incorporation efficiency, the greater the rate of bioconcentration.

(3) **Ratio of Weight at Death to Weight at Birth.** For a given retention factor and incorporation efficiency, the more weight an animal puts on during its lifetime, compared to its weight at birth, the greater is the percentage increase in the concentration of a trace substance over the lifetime of the organism. Note that while this ration, the *relative growth factor*, is sometimes correlated with the lifetime of the organism, it need not be. Fish generally don't live as long as humans, but their proportional weight gain is much larger. If an organism is fully grown well before death (like humans or birds, but not fish or trees), and if that organism lives many years in its fully grown state, then during these mature years its intake of the trace substance continues but its body weight does not change. Therefore, The longer the period in which the organism is fully grown, the greater the concentration of the trace substance in the organism at death. However, that effect is already described by the incorporation efficiency, which, if given as an average over the lifetime of the organism, will reflect a possible lack of growth during later years.

(4) **Location in the Food Chain.** The osprey bioconcentrates the trace substance from the fish it eats and, if other factors are equal, accumulates a higher concentration than that in the fish. The higher in a food chain an organism feeds, the greater is the concentration effect for that organism.

(5) **Environmental Contamination.** The concentration of the trance substance in the plankton reflects that in the water. The contamination of soil and water initiates the food-chain effect

and its amount is therefore an important determinant of the ultimate concentration in all organisms.

From this qualitative picture we would like to build a plausible mathematical model to determine the quantitative importance of the various factors enumerated above. Consider N populations, each in a steady state, whose feeding pattern is described by a linear food chain. Ignore immigration and emigration. Let $X_1, \ldots, X_N$ be the biomasses of the populations, where X_j eats X_{j-1}, for each $j \geq 2$. None of the N populations feed on X_N; dead individuals from that population are decomposed by organisms not included among the N populations here.

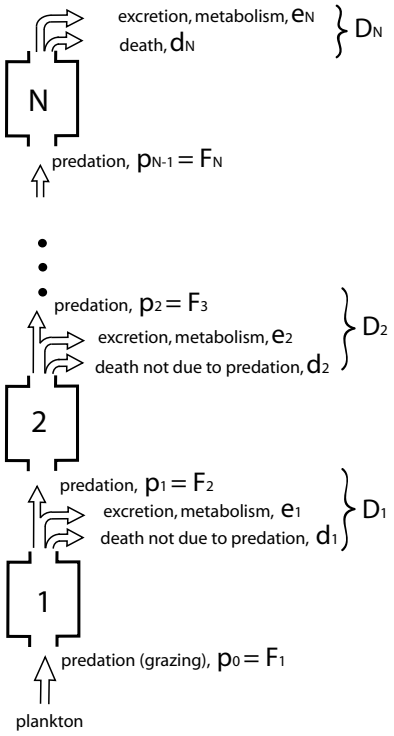
Assuming that D_i is the total rate at which biomass exits from the ith population. This total rate is the sum of metabolism plus excretion, e_i, predation, p_i, and other causes of death, d_i:

(eq. 1) $$D_i = e_i + p_i + d_i.$$

The total inflow rate is F_i, consisting entirely of ingested food. Note that birth in not an inflow; the process of birth adds no biomass to the population, but only divides the existing biomass into smaller pieces. The steady-state assumption is equivalent to

(eq. 2) $$F_i = D_i$$

for $i = 1, \ldots, N$.



All rate constants, D_i, e_i, p_i, d_i, and F_i are in units of biomass per unit time. The predation from level i is the input to level $i+1$, and so

(eq. 3) $$p_i = F_{i+1}.$$

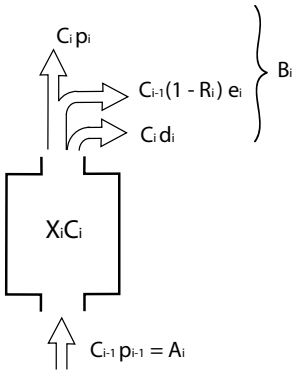
The rate at which ingested food is not metabolized or excreted is $p_i + d_i$, and therefore the incorporation efficiency, E_i, for each population is

(eq. 4) $$E_i = \frac{p_i + d_i}{F_i}.$$

At time $t = 0$, assume a trace substance is added to the environment (e.g., the water or soil in which population 1 grows) so that a constant environmental contamination, characterized by a concentration of C_0, is maintained for all time thereafter. Denote the concentration of pollutant at time t in each population by $C_i(t)$. C_i has units of mass of trace substance per unit biomass. The amount of trace substance in the ith population at time t is $M_i(t) = X_i C_i(t)$.

Define A_i and B_i to be the inflow and outflow of the trace substance (in units of mass per unit time) to each population. The inflow rate, A_i, to the ith population is determined by the predation rate of i on $i-1$ and by the concentration of trace substance in the $(i-1)$ population, C_{i-1}. In particular,

(eq. 5) $$A_i = p_{i-1} C_{i-1}.$$



The rate of loss of the trace substance from the ith population can be expressed as a sum of two terms. The first equals the concentration, C_i, times the rate of outflow (by predation or other forms of death) of trace-substance-containing biomass from that population. The second is proportional to C_{i-1} times the metabolism-plus-excretion rate, with proportionality constant equal to, or a fraction less than, one. (Note that C_{i-1} rather than C_i enters into the loss rate here because excrement has a concentration of trace substance governed by that of the food rather than that of the body as a whole.) That fraction is one minus a retention factor, R_i, where R_i is the fraction of ingested trace substance retained in the body. In equation form, the outflow rate is

(eq. 6) $$B_i + C_i(p_i + d_i) + C_{i-1}(1 - R_i)e_i.$$

Equation 6 states that predation and other types of death $(p_i + d_i)$ lead to a loss of trace substance from the population equal to 100% of whatever is in the bodies of dying organisms; it also states that metabolism and excretion lead to the loss of only a fraction, $(1 - R_i)$, where R_i is the retention factor for the ith population.

Although the X_i (as well as e_i, d_i, p_i) are assumed constant, note that we do not yet suppose that the concentrations of the trace substance are at steady state. So for now those concentrations are possibly non-constant functions of time, $C_i(t)$. (We suppose that $C_0(t)$ is a known function, from which the other concentrations are to be determined.)

From equations (1) through (5) we obtain, for $i = 1, 2, \ldots, N$,

(eq. 7) $$A_i(t) = C_{i-1}(t)\, p_{i-1}$$

(eq. 8) $$B_i(t) = \left[C_i(t)E_i + C_{i-1}(t)(1+R_i)(1-E_i)\right] p_{i-1}.$$

The rate of accumulation of the trace substance in population i, $\dot{M}_i$, is $A_i - B_i$. Thus we have for C_i the differential equation:

(eq. 9.i) $$X_i \dot{C}_i(t) = C_{i-1}(t)\, p_{i-1} - \left[C_i(t)E_i + C_{i-1}(t)(1+R_i)(1-E_i)\right] p_{i-1}.$$

Together, equations 9.1, …, 9.N form a system of linear differential equations for $C_1, \ldots, C_N$. We will see how to solve such systems later. For the moment let us finally assume that these concentrations are also at steady state, hence constant. In this case we have $A_i = B_i$ for each i. Using equations 7 and 8, we obtain after a little algebra that

(eq. 10) $$C_i = \left[1 + R_i \frac{(1-E_i)}{E_i}\right] C_{i-1}.$$

Since E_i is between 0 and 1, the quantity C_i/C_{i-1} is greater than 1. Therefore, the concentration at trophic level i is greater than the concentration at level C_{i-1} for each i. Iterating equation (10) gives

(eq. 11) $$C_i = \left[1 + R_i \frac{(1-E_i)}{E_i}\right] \ldots \left[1 + R_2 \frac{(1-E_2)}{E_2}\right] \left[1 + R_1 \frac{(1-E_1)}{E_1}\right] C_0.$$

As a very crude numerical estimate, suppose that the retention rate R_i and the incorporation efficiency E_i are both 0.5. Then $C_i/C_{i-1} = 1.5$. This means that the concentration increases by 50% from population $i-1$ to population i.

4.9. **First Order Linear ODEs.** In the previous problem we encountered a system of differential equations of the following form, where $x_1, \ldots, x_N$ are unknown functions of t, $f(t)$ is a given (known) function, and a_i, b_i are constants:

$$\begin{aligned} \dot{x}_1 &= a_1 x_1 + f(t) \\ \dot{x}_2 &= a_2 x_2 + b_1 x_1 \\ \dot{x}_3 &= a_3 x_3 + b_2 x_2 \\ &\cdots \\ \dot{x}_N &= a_N x_N + b_{N-1} x_{N-1}. \end{aligned}$$

If we know how to solve a differential equation of the form $\dot{x} = ax + f(t)$ we can solve this system inductively: first solve the first equation, then substitute the solution, x_1, into the second and solve for x_2, etc.

Let us consider the more general equation with specified initial value (for initial time t_0)

(eq. 1) $$\begin{aligned} \dot{x} &= a(t)x + f(t) \\ x(t_0) &= x_0, \end{aligned}$$

where $a(t)$ is also allowed to depend on t.

The key remark for solving this initial value problems is that the function of t defined by

$$I(t) = \exp\left(-\int_{t_0}^{t} a(s)ds\right)$$

satisfies the equation

$$\dot{I} = -a(t)I(t),$$

as you can show by an application of the chain rule. We call I the *integrating factor* for the differential equation.

To solve the initial value problem, first multiply both sides of the equation $\dot{x} - a(t)x = f(t)$ by $I(t)$ and observe that

$$\begin{aligned} fI &= \dot{x}I - xaI \\ &= \dot{x}I + x\dot{I} \\ &= \frac{d}{dt}(xI). \end{aligned}$$

Note that, by the fundamental theorem of calculus and the fact that $I(t_0) = 1$, we have

$$\begin{aligned} \int_{t_0}^{t} (x(s)I(s))'ds &= x(t)I(t) - x(t_0)I(t_0) \\ &= x(t)I(t) - x(t_0). \end{aligned}$$

Therefore, integrating both sides of the equation $fI = (xI)'$ from t_0 to t gives the solution

$$x(t) = I(t)^{-1}\left(x(t_0) + \int_{t_0}^{t} f(s)I(s)ds\right).$$

Suppose that $t_0 = 0$ and that a is a constant. Then

$$I(t) = e^{-at}$$

and the above general solution reduces to

$$x(t) = x(0)e^{at} + e^{at}\int_0^t e^{-sa} f(s)ds.$$

Exercise 16. *Use the above general expression to solve the initial value problem:*

$$\begin{aligned} \dot{x} &= cx + a + b_1 e^{-k_1 t} + \cdots + b_n e^{-k_n t} \\ x(0) &= A, \end{aligned}$$

where a, b_i, c, k_i and A are constants, for $i = 1, \ldots, n$. Show that the solution is

$$x(t) = A + \frac{a}{c}\left(1 - e^{-ct}\right) + \sum_{i=1}^{n} \frac{b_i}{c + k_i}\left(1 - e^{-(c+k_i)t}\right).$$

Exercise 17. *Solve the system of differential equations*

$$\begin{aligned} \dot{x}_1 &= a_1 x_1 + b_0; \\ \dot{x}_2 &= a_2 x_2 + b_1 x_1. \end{aligned}$$

The initial conditions are $x_1(0) = x_2(0) = 0$.

5. Stoichiometry and Linear Algebra

A very nice reference for the material of this section is *Elementary Chemical Reactor Analysis*, by Rutherford Aris. If occasionally a phrase below sounds especially well crafted and clear, it is likely that it was lifted straight out of that excellent book. (The more rigid and formalistic paragraphs are probably my own, though.)

The strict meaning of the word "stoichiometry" is measurement of the elements, but it is commonly used to refer to all sorts of calculations regarding the composition of a chemical system. Stoichiometry is essentially the bookkeeping of the material components of the chemical system.

Changes in the composition of chemical system are subject to certain restrictions. To see the nature of these restrictions let us consider again the reaction

$$2H_2 + O_2 = 2H_2O.$$

An equation such as this can have two meanings. It may be a kinetic description of the reaction and imply that two molecules of hydrogen combine with one of oxygen to form two molecules of water. In this particular case, as suggested in an earlier discussion, the equation is not true as a kinetic description. On the other hand, it may be a stoichiometric description of the reaction, and the equation will then mean that the numbers of hydrogen and oxygen molecules combining to form water are in the ratio $2:1$. The important restriction in this regard is that the total number of atoms of each type be the same on both sides of the equation. In other words, atoms are not created or destroyed.

It is clear that if a reaction expression is true in a kinetic sense it is also true in a stoichiometric sense, but the converse statement is false. For the moment, the kinetic meaning of a chemical equation will only concern us in passing.

From a chemical engineer's point of view, the important thing about a reaction like the one above is that three chemical species (here hydrogen, oxygen, and water) are observed to react in certain proportions. If the engineer can account for the rate at which this takes place in terms of concentrations of the observed chemical species, he has all he needs to know about the reactions so far as making numerical predictions about the outcome of the process is concerned. If a chemical equation cannot account for the detailed kinetics of the process, it is the chemist's job to provide a kinetic description of what is taking place, and to devise a *mechanism* for a reaction. This amounts to looking for elementary reaction steps and their reaction rates. In doing so the chemist may be led to hypothesize certain intermediate species which are present in only trace quantities.

In order to describe the stoichiometric restrictions in general (whether or not it corresponds to a true kinetic description of a process), we first need to set some convenient notation. We start with a set of s chemical species and denote them $X_1, X_2, \ldots, X_l$. Thus if $X_1 = H_2O$, $X_2 = H_2$, and $X_3 = O_2$, the reaction shown above would be $2X_2 + X_3 = 2X_1$. It is convenient to write all the chemical species on one side of the equation and to give a positive sign to the species which are regarded as the products of the reaction:

$$2X_1 - 2X_2 - X_3 = 0.$$

This convention should not be given undue importance, but it is usually possible to observe it, at least for the main reaction.

The numbers 2, -2 and -1 in the previous equation are called the *stoichiometric coefficients*. A natural way to write the general equation is

$$\sum_{i=1}^{l} \alpha_j X_j = 0.$$

Thus X_i denotes the ith chemical species and α_i is its stoichiometric coefficient in the given reaction. Subject to the above convention it is convenient to call the species with positive stoichiometric coefficients the *products* of the reaction and those with negative coefficients the *reactants*. The products are formed from the reactants by the *forward reaction*, while the reverse reaction converting the products into the reactants will be called the *back reaction*.

Both forward and back reactions are usually going on simultaneously and equilibrium is reached when they go at equal and opposite rates. It is sometimes useful to include an inert chemical species in the set of X_i, and since it does not take part in the reaction it is given stoichiometric coefficient 0. An *entire* reaction is thus one whose behavior can be fully described in terms of the concentrations of the species $X_1,\dots,X_l$.

The important thing about the stoichiometric coefficient of a reaction is their ratio rather than their absolute magnitude. Thus the reaction for the formation of water could just as well be written $H_2O - H_2 - \frac{1}{2}O_2 = 0$. We can say therefore that the stoichiometric coefficients of a reaction are given up to a constant multiplier, for the equations $\sum \lambda\alpha_i X_i = 0$ has exactly the same meaning as $\sum \alpha_i X_i = 0$.

A remark about notation: although expressions such as $2H_2 + O_2 = 2H_2O$, which represent reactions as equations, are widely used in the chemistry literature, it will be for us a possible source of confusion and will be often (but not always) avoided. A reaction will more often be expressed by the "harpoon" notation, $2H_2 + O_2 \rightharpoonup 2H_2O$, or by $2H_2 + O_2 \rightleftharpoons 2H_2O$, when the direction of reaction is not of immediate concern or when we need to refer explicitly to the forward and back reactions. Mathematically, all that really matters to characterize a reaction is the expression $2H_2O - 2H_2 - O_2$, which we regard as a vector. This vector is sometimes described by its coordinates, $(2,-2,-1)$. Viewing the reaction expression as a vector, the equation $2H_2 + O_2 = 2H_2O$ is simply a wrong identity.

One feature that distinguishes chemical processes from the more general types mentioned earlier is that the substances undergoing a chemical transformation are composed of more elementary units, or 'atoms,' and the quantities of individual atoms must be conserved over the course of the transformation. We wish now to formalize this idea. The stoichiometric description of a process involves:

(1) a list of participating substances, or *molecules*, (or species) of the process:

$$X_1, X_2, \cdots, X_l;$$

(2) a list of elements, or *atoms*:

$$E_1, E_2, \cdots, E_N.$$

The atomic composition of substance X_j can be written as

$$X_j = E_1^{a_{1j}} E_2^{a_{2j}} \cdots E_N^{a_{Nj}},$$

where the numbers a_{ij} are non-negative (possibly 0) integers.

Example. Suppose that the process we wish to study is the partial combustion of methane. In standard chemical notation this is described as follows: CH_4 (methane) combines with O_2 (oxigen), producing CO (carbon monoxide), H_2O (water), and H_2 (hydrogen). In the above notation, there are 6 molecules:

$$X_1 = CH_4,\ X_2 = O_2,\ X_3 = CO_2,\ X_4 = CO,\ X_5 = H_2O,\ X_6 = H_2;$$

These are composed of 3 different types of atoms:

$$E_1 = C, E_2 = O, E_3 = H.$$

In this notation, the methane molecule for example is $E_1^1 E_2^0 E_3^4$, which we often simplify as $E_1 E_3^4$.

By a *mixture* we will mean a linear combinations of the substances with non-negative coefficients:

$$x_1 X_1 + x_2 X_2 + \cdots + x_l X_l .$$

How should the expression $x_1 X_1 + \cdots + x_l X_l$ be interpreted? The interpretation may vary somewhat with the context, but we will generally think of it as describing the amount or concentration of each substance inside some region in space, which may be a closed or open vessel, (a test tube, say) or simply a small volume in space delimited by an imaginary boundary surface, inside of which molecules will be at close enough distance to interact. In this case the coefficients x_i assume integer values. So, the amount of each species X_i is specified by the coefficient x_i. In all cases, whether the x_i are integer quantities or not, we usually write them in number of moles rather than by their mass.

It is important to note that when we write an expression such as $3.7CH_4 + 2.5O_2 + 1.2H_2$, the symbols CH_4, O_2 and H_2 (methane, oxygen and hydrogen) do not stand for numerical quantities. We treat these symbols the same way we treat expressions like $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ in analytic geometry. Thus the symbols for molecules and atoms will be regarded as unit vectors in some coordinate space. If there are l molecular species, then the coordinate space is the l-dimensional space $\mathbb{R}^l$. This is, by definition, the set of all l-tuples of real numbers. The notations $(x_1, \cdots, x_l)$ and $x_1 X_1 + \cdots, x_l X_l$ represent the same point. Therefore X_i corresponds to the basis vector:

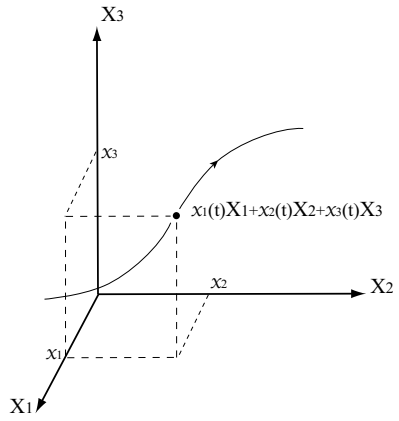
$$X_i = (0, \cdots, 0, 1, 0, \cdots, 0)$$

having 1 at the i-th position and 0 everywhere else.

During a process, the composition of a mixture may change due to chemical reactions or transport across the boundary of the vessel. So the mixture coefficients are generally functions of time, $x_i(t)$. The time evolution of the system thus corresponds to a parametric curve (parametrized by time, t) in the space of mixtures:

$$t \to x_1(t) X_1 + x_2(t) X_2 + \cdots + x_l(t) X_l .$$

This is shown in the next figure for a system involving three species.



5.1. **Linear Independence.** Consider the reactions:

(react. 1) $$CO + 2H_2 \rightleftharpoons CH_3OH$$

(react. 2) $$CO_2 + H_2 \rightleftharpoons H_2O + CO$$

(react. 3) $$CO_2 + 3H_2 \rightleftharpoons CH_3OH + H_2O.$$

Reaction 3 does not tell us any more about the system than is contained in the two other reactions. These reactions are not independent of each other, as the third is the sum of the first two. The precise mathematical description of dependence involves the reaction vectors. Thus the vectors

$$\begin{aligned}\mathbf{R}_1 &= CH_3OH - CO - 2H_2,\\ \mathbf{R}_2 &= H_2O + CO - CO_2 - H_2,\\ \mathbf{R}_3 &= CH_3OH + H_2O - CO_2 - 3H_2,\end{aligned}$$

are *linearly dependent* since

$$\mathbf{R}_3 = \mathbf{R}_1 + \mathbf{R}_2.$$

The general definition of linear independence is this: let $\mathbf{R}_1, \ldots, \mathbf{R}_m$ be a set of vectors in $\mathbb{R}^l$. Then they are said to be linearly independent if the expression

$$c_1\mathbf{R}_1 + \cdots + c_m\mathbf{R}_m = 0$$

can only hold if the numbers $c_1, \cdots, c_m$ are all zero.

As another example, consider the set of reactions that is said to be the kinetic description of the formation of hydrogen bromide. These are:

$$\begin{aligned}Br_2 &\rightharpoonup 2Br,\\ Br + H_2 &\rightharpoonup HBr + H,\\ H + Br_2 &\rightharpoonup HBr + Br,\\ H + HBr &\rightharpoonup H_2 + Br,\\ 2Br &\rightharpoonup Br_2.\end{aligned}$$

How many independent reactions are there in this set? Clearly the first and fifth are dependent (one reaction vector is the negative of the other). The same is true for the second and fourth. Therefore there cannot be more than 3 independent reactions. One way to check that there are exactly three is to use the method of row reduction. We first set the basis vectors:

$$X_1 = Br_2, X_2 = Br, X_3 = H_2, X_4 = H, X_5 = HBr.$$

Then the reaction vectors are:

$$\begin{aligned}\mathbf{R}_1 &= -X_1 + 2X_2,\\ \mathbf{R}_2 &= -X_2 - X_3 + X_4 + X_5,\\ \mathbf{R}_3 &= -X_1 + X_2 - X_4 + X_5,\\ \mathbf{R}_4 &= X_2 + X_3 - X_4 - X_5,\\ \mathbf{R}_5 &= X_1 - 2X_2.\end{aligned}$$

It is convenient to detach the stoichiometric coefficients and write them in matrix form with the coefficients of each reaction vector as a row:

$$\begin{array}{rrrrr} -1 & 2 & 0 & 0 & 0 \\ 0 & -1 & -1 & 1 & 1 \\ -1 & 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & -1 & -1 \\ 1 & -2 & 0 & 0 & 0 \end{array}$$

If we permute the rows, multiply a row by a nonzero number, or add to a row a linear combination of the other rows, the number of linearly independent equations is not changed. We can use these operations to simplify the above matrix until it takes a form that makes the answer (the number of independent equations) clear.

By replacing (1) the fifth row with the sum of the fifth and first rows, and (2) the fourth row with the sum of the fourth and second rows, we obtain:

$$\begin{array}{rrrrr} -1 & 2 & 0 & 0 & 0 \\ 0 & -1 & -1 & 1 & 1 \\ -1 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}$$

We now (3) replace the first row with the negative of itself, and (4) replace the third row with the sum of the third and the new first row:

$$\begin{array}{rrrrr} 1 & -2 & 0 & 0 & 0 \\ 0 & -1 & -1 & 1 & 1 \\ 0 & -1 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}$$

Next, we (5) replace the second row with its negative, and (6) replace the third row with the sum of the new second with the third:

$$\begin{array}{rrrrr} 1 & -2 & 0 & 0 & 0 \\ 0 & 1 & 1 & -1 & -1 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}$$

Ignoring the two trivial last rows, we obtained a matrix which has ones on the diagonal and zeros below the diagonal. Going back to the respective vectors:

$$\begin{aligned} \mathbf{S}_1 &= X_1 - 2X_2 \\ \mathbf{S}_2 &= X_2 + X_3 - X_4 - X_5 \\ \mathbf{S}_3 &= X_3 - 2X_4. \end{aligned}$$

Exercise 18. *Show that the vectors* $\mathbf{S}_1, \mathbf{S}_2$ *and* $\mathbf{S}_3$ *are linearly independent by using the definition. That is, show that the equation*

$$c_1\mathbf{S}_1 + c_2\mathbf{S}_2 + c_3\mathbf{S}_3 = 0$$

only holds if $c_1 = c_2 = c_3 = 0$.

We could have simplified the last matrix further by subtracting the third row from the second and making the result the new second row, and then replacing the first row with the sum of itself plus twice the (new) second row. The result is:

$$\begin{array}{rrrrr} 1 & 0 & 0 & 2 & -2 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}$$

The first three rows correspond, respectively, to the following three reactions (now obviously independent since no two of them contain all the substances involved in the remaining reaction):

$$2HBr \rightharpoonup Br_2 + 2H$$
$$HBr \rightharpoonup Br + H$$
$$2H \rightharpoonup H_2.$$

Exercise 19. *How many of the following* 6 *reactions are independent?*

$$2C_2H_4 + O_2 \rightleftharpoons 2C_2H_4O$$
$$C_2H_4 + 3O_2 \rightleftharpoons 2CO_2 + 2H_2O$$
$$2C_2H_4O + 5O_2 \rightleftharpoons 4CO_2 + 4H_2O.$$

Exercise 20. *How many of the following* 12 *reactions are independent?*

$$4NH_3 + 5O_2 \rightleftharpoons 4NO + 6H_2O$$
$$4NH_3 + 3O_2 \rightleftharpoons 2N_2 + 6H_2O$$
$$4NH_3 + 6NO \rightleftharpoons 5N_2 + 6H_2O$$
$$2NO + O_2 \rightleftharpoons 2NO_2$$
$$2NO \rightleftharpoons N_2 + O_2.$$

Exercise 21. *How many of the following* 6 *reactions are independent?*

$$2NaCl + H_2SO_4 \rightleftharpoons Na_2SO_4 + 2HCl$$
$$Na_2SO_4 + 4C \rightleftharpoons Na_2S + 4CO$$
$$Na_2S + CaCO_3 \rightleftharpoons Na_2CO_3 + CaS.$$

5.2. **Dimension.** Given a set of vectors $\mathbf{R}_1, \cdots, \mathbf{R}_m$ in $\mathbb{R}^l$, the collection of all the other vectors that can be obtained from these by linear combinations constitutes a linear subspace of $\mathbb{R}^l$. The number of linearly independent vectors among the $\mathbf{R}_i$ is the *dimension* of that subspace.

In the example given above, $\mathbf{R}_1, \cdots, \mathbf{R}_5$ span a 3-dimensional subspace of $\mathbb{R}^5$. This is the same subspace spanned by the vectors $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3$ obtained from the $\mathbf{R}_i$ by using row reduction.

We sometimes refer to this dimension as the *rank* of the stoichiometric matrix (the matrix of the stoichiometric coefficients) derived from the vectors $\mathbf{R}_i$.

If a chemical process involves l different substances which are known to be involved in k reactions, $\mathbf{R}_1, \cdots, \mathbf{R}_k$, then we will call the linear subspace of $\mathbb{R}^l$ spanned by these reactions the *reaction space* of the process.

Exercise 22. *Calculate the dimension of the reaction space for the reaction mechanisms defined in the previous three exercises.*

Exercise 23. *Find the dimension of the reaction space involving substances X_1, X_2, X_3, for the system of (isomerization) reactions:*

$$X_1 \rightleftharpoons X_2$$
$$X_2 \rightleftharpoons X_3$$
$$X_3 \rightleftharpoons X_1.$$

Exercise 24. *Suppose that a chemical process involves the two reactions*

$$3H_2 + N_2 \rightharpoonup 2NH_3$$
$$H_2 + CO_2 \rightharpoonup CO + H_2O.$$

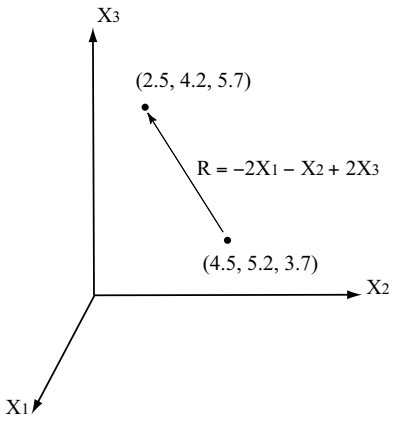
A mixture of 68.4% H_2, 22.6% N_2, *and* 9% CO_2 *reacts until* 15% NH_3 *and* 5% H_2O *are formed. (These are mole percentages.) What are then the percentages of H_2 and N_2?*

5.3. **Degree of Advancement of a Reaction.** Let $\mathbf{x}_{\text{initial}} \in \mathbb{R}^l$ be a point with non-negative coordinates representing an initial composition (in moles) of a mixture of l substances. Suppose that the mixture undergoes a reaction represented by the vector $\mathbf{R}$. Then at any given moment the new mixture composition, $\mathbf{x}_{\text{new}}$, must differ from $\mathbf{x}_{\text{initial}}$ by a multiple of $\mathbf{R}$, that is

$$\mathbf{x}_{\text{new}} = \mathbf{x}_{\text{initial}} + s\mathbf{R},$$

where s is some (possibly negative) number. The number s is called the *molar extent*, or *degree of advancement* of the reaction.

Say, for example, that the reaction is $2X_1 + X_2 \rightharpoonup 2X_3$, where $X_1 = H_2$, $X_2 = O_2$, and $X_3 = H_2O$. Let the initial composition of a mixture of oxygen (gas), hydrogen (gas), and water (liquid) be $\mathbf{x}_{\text{initial}} = (4.5, 5.2, 3.7)$; that is, the mixture contains 4.5 moles of hydrogen, 5.2 of oxygen, and 3.7 of water. Suppose that the reaction proceeds to the extent that 1 mole of oxygen is consumed. Then of necessity 2 moles of hydrogen will be consumed and 2 of water will be produced. This means that $\mathbf{x}_{\text{new}} = (4.5, 5.2, 3.7) + (-2, -1, 2) = (2.5, 4.2, 5.7)$. In this case, the molar extent of the reaction is 1.



If at a given moment the amount of hydrogen is, say, 6.1 moles, then of necessity the composition of the other substances, X_2, X_3, can be obtained from the equation

$$(6.1, x_2, x_3) = (4.5, 5.2, 3.7) + s(-2, -1, 2),$$

for some s. From the amount of hydrogen we obtain $6.1 = 4.5 - 2s$, that is, $s = -0.8$. Therefore the reaction went in reverse, by the extent 0.8 moles. The new amounts of oxygen and water are $x_2 = 5.2 - 0.8 \times (-1) = 6$ and $x_3 = 3.7 - 0.8 \times 2 = 2.1$ moles, respectively.

We need not limit ourselves to only one reaction. If there are k reaction involved, $\mathbf{R}_1, \cdots, \mathbf{R}_k$, we can introduce k advancement parameters, $s_1, \cdots, s_k$, and represent any new composition as

$$\mathbf{x}_{\text{new}} = \mathbf{x}_{\text{initial}} + s_1\mathbf{R}_1 + \cdots + s_k\mathbf{R}_k.$$

Notice, however, that unless $\mathbf{R}_1, \cdots, \mathbf{R}_k$ are independent reactions, the numbers $s_1, \cdots, s_k$ are not uniquely determined from the initial and new values of the mixture composition.

5.4. **Atomic Composition.** A chemical process that takes place in a closed reactor (that is, without exchange of substances between the reactor and the environment) is constrained by the property that the molar amount (or, equivalently, the number of atoms) of each participating atomic species does not change during the process. If you go back to any example of a chemical reaction given so far in these notes, you will see that the number of atoms of each element is the same on both the left and the right sides of a reaction $\alpha_1 X_1 + \cdots + \alpha_l X_l \rightharpoonup \beta_1 X_1 + \cdots + \beta_l X_l$. We explain now one way in which this remark can be used in a systematic way to sharpen our characterization of the set of points in $\mathbb{R}^l$ that are accessible to the evolving process.

Suppose that there are l types of molecules and N types of atoms. We have introduced above the space of mixtures $\mathbb{R}^l$. If we wish to regard only the amounts of each kind of atom in a mixture, it is natural to consider as well the space $\mathbb{R}^N$, with basis $E_1, \cdots, E_N$. Define the *atomic composition* function $\mathcal{E} : \mathbb{R}^l \to \mathbb{R}^N$ as follows: on each molecule, $X_j = E_1^{a_{1j}} E_2^{a_{2j}} \cdots E_N^{a_{Nj}}$,

$$\mathcal{E}(X_j) = a_{1j}E_1 + a_{2j}E_2 + \cdots + a_{Nj}E_N.$$

We extend this definition to an arbitrary mixture by linearity. This means:

$$\mathcal{E}(x_1X_1 + \cdots + x_lX_l) = x_1\mathcal{E}(X_1) + x_2\mathcal{E}(X_2) + \cdots + x_l\mathcal{E}(X_l).$$

Thus defined, $\mathcal{E}$ is a linear map from $\mathbb{R}^l$ to $\mathbb{R}^N$. If we represent points in $\mathbb{R}^l$ and $\mathbb{R}^N$ as column vectors, $\mathcal{E}$ takes the matrix form

$$\mathcal{E}(X_j) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1l} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2l} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{Nj} & \cdots & a_{Nl} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{Nj} \end{pmatrix} = a_{1j}E_1 + \cdots + a_{Nj}E_N.$$

We can now express the idea of conservation of molar amounts of atomic species in the following way.

Principle 25 (Atomic Balance)**.** *In a closed system, any changes in the composition of a mixture should not alter its atomic composition. In other words, if $A, B \in \mathbb{R}^l$ are the vectors representing the mixture composition at two different moments, then $\mathcal{E}(A) = \mathcal{E}(B)$. Equivalently, reaction vectors are contained in the kernel of the linear map $\mathcal{E}$. (I.e., the reaction space is contained in the stoichiometric space.)*

This means that if at a given moment in time the mixture has composition specified by the vector $A_0 = (x_1, \ldots, x_l)$, then no matter what chemical transformations take place, as long as we are dealing with a closed system, the curve described above must remain inside the set

$$\{A \in \mathbb{R}^l \mid \mathcal{E}(A) = A_0\}.$$

We call this set the *stoichiometric space* through A_0.

Example. Suppose that a closed system contains a mixture of three substances (gases): $X_1 = CO_2$, $X_2 = CO$, and $X_3 = O_2$. There are only two types of atoms: $E_1 = C$ and $E_2 = O$. Thus the atomic composition map $\mathcal{E} : \mathbb{R}^3 \to \mathbb{R}^2$ is obtained as follows, where we use that $\mathcal{E}(CO_2) = C + 2O$, $\mathcal{E}(CO) = C + O$, and $\mathcal{E}(O_2) = 2O$:

$$\begin{aligned}\mathcal{E}(x_1X_1 + x_2X_2 + x_3X_3) &= x_1\mathcal{E}(X_1) + x_2\mathcal{E}(X_2) + x_3\mathcal{E}(X_3)\\ &= x_1(E_1 + 2E_2) + x_2(E_1 + E_2) + x_3(2E_2)\\ &= (x_1 + x_2)E_1 + (2x_1 + x_2 + 2x_3)E_2.\end{aligned}$$

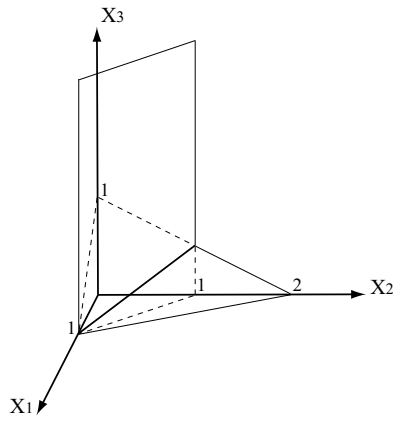
In matrix form,

$$\mathcal{E} = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & 2 \end{pmatrix}.$$

If the initial mixture contains an amount a_1 of atoms of type E_1, and a_2 of type E_2, then the atomic balance principle says that the composition of the mixture must remain in the space defined by the two equations:

$$\begin{aligned} x_1 + x_2 &= a_1 \\ 2x_1 + x_2 + 2x_3 &= a_2. \end{aligned}$$

This system represents the intersection of two planes in $\mathbb{R}^3$. Let us say that $a_1 = 1$ and $a_2 = 2$ (in moles, say). Then you can check that the planes and line of intersection are as in the next figure.



Therefore, any chemical process involving only the substances CO_2, CO and O_2 with the amounts of C and O specified is described by a motion along this line of intersection. The motion won't be a simple uniform translation, of course. The precise function of time (parametrization) of the curve that describes it will depend on the reaction mechanism governing the process, and this mechanism has not been specified yet.

Exercise 26. *Show that, up to a scalar multiple (possibly negative) the only possible reaction for this example is given by*

$$2CO + O_2 \rightharpoonup 2CO_2.$$

Example. For a somewhat more complicated example consider the following substances, which are involved in the process of partial combustion of methane (CH_4):

$$X_1 = CH_4,\ X_2 = O_2,\ X_3 = H_2O,\ X_4 = CO,\ X_5 = H_2.$$

There are three different types of atoms:

$$E_1 = C,\ E_2 = H,\ E_3 = O.$$

Thus the atomic composition map has domain $\mathbb{R}^5$ and target space $\mathbb{R}^3$. We expect the dimension of the stoichiometric space to be $5-3=2$. To see if this is indeed the case, we first need to determine $\mathcal{E}$.

For this example, we have:

$$\begin{aligned}
\mathcal{E}(X_1) &= E_1 + 4E_2\\
\mathcal{E}(X_2) &= 2E_3\\
\mathcal{E}(X_3) &= 2E_2 + E_3\\
\mathcal{E}(X_4) &= E_1 + E_3\\
\mathcal{E}(X_5) &= 2E_2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathcal{E}(x_1X_1 + \cdots + x_5X_5) &= x_1\mathcal{E}(X_1) + \cdots + x_5\mathcal{E}(X_5)\\
&= x_1(E_1 + 4E_2) + x_2(2E_3) + x_3(2E_2 + E_3) + x_4(E_1 + E_3) + x_5(2E_2)\\
&= (x_1 + x_4)E_1 + (4x_1 + 2x_3 + 2x_5)E_2 + (2x_2 + x_3 + x_4)E_3.
\end{aligned}$$

In matrix form,

$$\mathcal{E} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0\\ 4 & 0 & 2 & 0 & 2\\ 0 & 2 & 1 & 1 & 0 \end{pmatrix}.$$

Denoting by a_1, a_2, a_3 the amounts of E_1, E_2, E_3 (in moles), we obtain the linear system

$$\begin{aligned}
x_1 + x_4 &= a_1\\
4x_1 + 2x_3 + 2x_5 &= a_2\\
2x_2 + x_3 + x_4 &= a_3.
\end{aligned}$$

These are 3 linear equations in 5 unknowns, and we should expect it to describe a 2-dimensional subspace of $\mathbb{R}^5$. To obtain a parametric representation of this plane, we can solve for x_1, x_2, x_3 in terms of the a_i and x_4, x_5. The result is:

$$\begin{aligned}
x_1 &= a_1 - x_4\\
x_2 &= \frac{1}{4}(4a_1 - a_2 + 2a_3) - \frac{3}{2}x_4 + \frac{1}{2}x_5\\
x_3 &= \frac{1}{2}(a_2 - 4a_1) + 2x_4 - x_5.
\end{aligned}$$

The stoichiometric plane can thus be regarded as the (x_4, x_5)-plane.

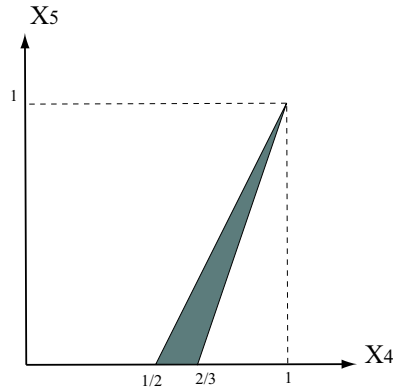
Exercise 27. *Show that any reaction involving $X_1, \cdots, X_5$ as in the above example is a linear combination of*

$$\begin{aligned}
2X_1 + 3X_2 &\rightharpoonup 4X_3 + 2X_4\\
2X_3 &\rightharpoonup X_2 + 2X_5.
\end{aligned}$$

We haven't taken into account the requirement that the x_i be non-negative. Let us see what subset of the stoichiometric plane these conditions give if we have $a_1 = 1$, $a_2 = 2$, $a_3 = 1$. From the inequalites $x_i \geq 0$, $i = 1,2,\ldots,5$, we get the following system of inequalities for x_4, x_5:

$$\begin{aligned} 1 - x_4 &\geq 0 \\ 2 - 3x_4 + x_5 &\geq 0 \\ -1 + 2x_4 - x_5 &\geq 0 \\ x_4 &\geq 0 \\ x_5 &\geq 0. \end{aligned}$$

Exercise 28. *Show that the solution set to this system of inequalities is the triangular region given in the figure below.*



5.5. **Mass Balance.** We have so far specified quantities of substances and elements in moles, but we might also want to consider their masses. Suppose that the various elements E_i have masses m_i (say, in grams per mole). Therefore the total mass of a mixture having atomic composition $a_1E_1 + \cdots + a_NE_N$ is $a_1m_1 + \cdots + a_Nm_N$. Let us introduce a linear function $\mathcal{M} : \mathbb{R}^N \to \mathbb{R}$ by

$$\mathcal{M}(a_1E_1 + \cdots + a_NE_N) = a_1m_1 + \cdots + a_Nm_N.$$

Then the mass of a mixture $K = x_1X_1 + \cdots + x_lX_l$ is the value of $\mathcal{M}$ for the atomic composition vector $\mathcal{E}(K)$. In other words, we obtain the mass of K by evaluating K on the composition function $\mathcal{M} \circ \mathcal{E}$. In particular, if $\mathcal{E}(K(t))$ remains constant as the mixture $K(t)$ changes in time (as it does for a closed system by the principle of atomic balance), then the total mass $\mathcal{M}(\mathcal{E}(K(t)))$ also remains constant in t.

6. Reaction Mechanisms

We have so far used reaction equations to indicate changes in the composition of mixtures of substances without being concerned whether the equation referred to some basic physical process involving binding and dissociation of individual molecules, or simply a description of the overall change in the mixture composition. The overall reaction may be the result of a collection of simpler and more fundamental processes. These processes are represented by the elementary reaction steps

and their reaction rates. Put together, these reaction steps and rates specify a particular reaction mechanism for the process, which can then be used to determine the process's time evolution.

This section explains these things in more detail.

6.1. **Elementary Reaction Steps.** Before being more specific, let us stop for a moment to consider on an intuitive level what we would like to mean by the idea of an elementary reaction of a reaction mechanism.

Let a general reaction, involving substances, or molecules, $X_1, \cdots, X_l$, be given by

$$a_1X_1 + \cdots + a_lX_l \rightharpoonup b_1X_1 + \cdots + b_lX_l$$

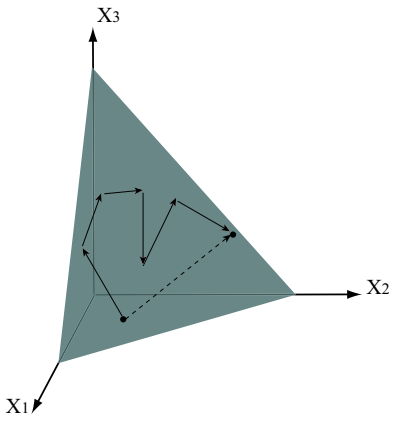
where the coefficients a_i, b_i are non-negative integers. We call the number $a_1 + \cdots + a_l$ the *length* of the reaction. The reaction will be said to be of first order, second order, etc, if its length is less than or equal to 1, 2, etc.

Now, what should we mean by a 'reaction mechanism'? Consider the overall reaction of oxidation of hydrogen:

$$2H_2 + O_2 \rightleftharpoons 2H_2O.$$

A detailed mechanism for this process is imagined to involve a large number of simpler reactions, such as $H_2 + O_2 \rightleftharpoons 2HO$, $OH + H_2 \rightleftharpoons H_2O + H$, among many others. Notice that these reactions are more elementary that the first one in one important respect: for $2H_2 + O_2 \rightharpoonup 2H_2O$ to occur in one step, three molecules must come together in close proximity (collide), with sufficient energy and at a brief enough interval of time in order to interact. In other words, it is a reaction of length 3. The other two reactions, on the other hand, have length 2, so their chances of actually taking place would seem greater. A reaction mechanism is an enumeration of what is believed to be the physically elementary steps through which the overall reaction will proceed.

The next figure represents a sequence of elementary reactions as given by their reaction vectors. There are only three substances in this case. We think of the elementary steps as individual "reaction events" that add up to the overall reaction represented by the broken line vector. (The gray triangle represents stoichiometric space, that is, the kernel of the linear map $\mathcal{E}$.)



Formally, a reaction mechanism will mean for us the specification of the following ingredients:

(1) A list of participating substances: $X_1, \cdots, X_l$;

(2) A list of elementary reactions, $\mathbf{R}_1, \cdots, \mathbf{R}_k$, which will often be directly represented by their reaction vectors or the corresponding stoichiometric coefficients. (The coefficients $a_1, \cdots, a_l, b_1, \cdots b_l$ are unambiguously specified by the reaction vector only if we suppose that $a_i b_i = 0$ for each i; that is, if either a_i or b_i is zero for each i so that no molecular species shows up on both sides of the reaction.)
(3) A list of reaction rates: $r_1, \cdots, r_k$, one for each reaction. (This is explained later.)

The actual determination of a reaction mechanism is typically a very complicated detective job involving theory and experiment. One can propose different hypothetical mechanisms for a given process and then select the one which best accounts for the empirical data obtained in the lab. For doing this selection it is helpful to be able to mathematically derive as much theoretical information as possible from each hypothetical mechanism.

Typically, an elementary reaction has the property of being itself and its reverse reaction both of order two. Reactions of order 3 may be consider elementary if one of the reactants is a catalyst. More on this later. For now, a catalytic reaction will typically take the form $X + X' + M \rightharpoonup Y + Y' + M$ where M, which appears on both sides of the reaction arrow, is the catalyst.

The reaction rates for elementary reactions steps are also expected to take relatively simple analytic expressions. We will often assume that these rates are given by the so-called *mass-action law*. More on this later.

One way in which mathematics can help to specify the set of elementary reactions is in providing ways to answer the following general question:

Question 29. *Suppose it is given a list of substances that we believe to take part in a chemical process under study. How can we find* **all** *possible reactions, subject to the atomic balance requirement, involving molecules from that list?*

We illustrate this point with a few examples. Suppose we are given the following list of substances; see the example on partial combustion of methane. (We are now adding CO_2 to the list.)

$$X_1 = CH_4,\ X_2 = O_2,\ X_3 = H_2O,\ X_4 = CO,\ X_5 = H_2,\ X_6 = CO_2.$$

As in the earlier example, we write $E_1 = C,\ E_2 = H,\ E_3 = O$. We wish to find the possible reactions, and in particular, candidates for elementary reactions, that can be written using this list.

We write a reaction vector in the form $\mathbf{R} = z_1X_1 + \cdots + z_6X_6$, where $z_i = b_i - a_i$. We know that $\mathbf{R}$ should lie in the kernel of $\mathcal{E}$. To determine this kernel, observe that

$$\mathcal{E}(z_1X_1 + \cdots + z_6X_6) = (z_1 + z_4 + z_6)E_1 + (4z_1 + 2z_3 + 2z_5)E_2 + (2z_2 + z_3 + z_4 + 2z_6)E_3.$$

Therefore, the z_i must solve the linear system:

$$\begin{aligned} z_1 + z_4 + z_6 &= 0 \\ 4z_1 + 2z_3 + 2z_5 &= 0 \\ 2z_2 + z_3 + z_4 + 2z_6 &= 0. \end{aligned}$$

The solution set is a linear space of dimension 3. Choosing z_4, z_5, z_6 as the independent variables, we solve for z_1, z_2, z_3:

$$\begin{aligned} z_1 &= -z_4 - z_6 \\ z_2 &= -\frac{3}{2}z_4 + \frac{1}{2}z_5 - 2z_6 \\ z_3 &= 2z_4 - z_5 + 2z_6. \end{aligned}$$

Taking for (z_4, z_5, z_6) the values $(1,0,0)$, $(0,1,0)$, $(0,0,1)$, we obtain (after eliminating denominators) the following linearly independent solution vectors:

$$\mathbf{u}_1 = (-2,-3,4,2,0,0)$$
$$\mathbf{u}_2 = (0,1,-2,0,2,0)$$
$$\mathbf{u}_3 = (-1,-2,2,0,0,1).$$

These three vectors constitute a basis for the kernel of $\mathcal{E}$. Therefore an arbitrary reaction that satisfies the atomic balance condition must have reaction vector

$$\mathbf{R} = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3,$$

where the coefficients are arbitrary (except for the condition that we want the reaction coefficients to be integers). After substitution, we obtain:

$$\mathbf{R} = (-2\alpha_1 - \alpha_3)X_1 + (-3\alpha_1 + \alpha_2 - 2\alpha_3)X_2 + (4\alpha_1 - 2\alpha_2 + 2\alpha_3)X_3 + 2\alpha_1 X_4 + 2\alpha_2 X_5 + \alpha_3 X_6.$$

Any reaction is obtained from this vector by substituting for α_i rational numbers (not necessarily non-negative), eliminating denominators, and then separating positive and negative terms to form the right and left sides of the reaction, respectively.

Each choice of $(\alpha_1, \alpha_2, \alpha_3)$ yields a reaction equation. A somewhat arbitrary choice of values, $(0,0,1)$, $(-1,0,2)$, and $(0,1,1)$, gives respectively,

$$CH_4 + 2O_2 \rightharpoonup CO_2 + 2H_2O$$
$$2CO_2 \rightharpoonup O_2 + 2CO$$
$$CH_4 + O_2 \rightharpoonup CO_2 + 2H_2.$$

Other reactions can be obtained from these by linear operations. We may think of these three reactions as forming a basis for the space of all reactions involving these 6 substances.

Exercise 30. *Is there any elementary reaction that can be generated by linear combinations of the above three? (For the sake of this exercise,* $\mathbf{R}$ *is elementary if both* $\mathbf{R}$ *and* $-\mathbf{R}$ *have order at most* 2*.)*

We wish in the next exercise to describe the possible reactions that can occur in a polymerization process. The monomer is represented by A and we wish to consider all reactions involving the substances $X_1, X_2, \cdots, X_l$, where $X_i = A^i$, a polymer containing i monomers.

Exercise 31. *How many elementary reaction can be written only involving polymers* A^i*, for* $i = 1, \cdots, l$*?*

For the purposes of this exercise we say that a reaction with vector $(b_1 - a_1)X_1 + \cdots + (b_l - a_l)X_l$ is elementary if:

(1) a_i and b_i are non-negative for each i;
(2) either a_i or b_i is zero, for each i;
(3) $a_1 + \cdots + a_l \leq 2$;
(4) $b_1 + \cdots + b_l \leq 2$.

Since there is only one type of "atom" (monomer), there is only one equation that defines the stoichiometric space. The equation is

$$(b_1 - a_1) + 2(b_2 - a_2) + \cdots + l(b_l - a_l) = 0.$$

Exercise 32. *Show that the elementary reactions involving polymers of length at most l and a single type of monomer are: For $l = 2$:*

$$2X_1 \rightleftharpoons X_2$$

For $l = 3$:

$$X_1 + X_3 \rightleftharpoons 2X_2$$
$$2X_1 \rightleftharpoons X_2.$$

What are the reactions for $l = 4$? Can you find the number of reactions for arbitrary l?

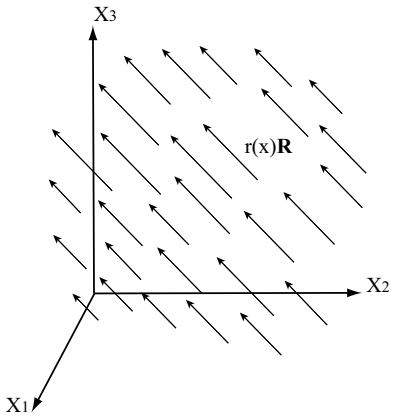
Exercise 33. *Find all the elementary reactions involving the substances: H_2, O_2, H_2O, OH, O, H, H_2O_2, HO_2.*

6.2. **Reaction Rates.** Recall the meaning of the reaction vector $\mathbf{R}$ for a reaction involving l substances: assuming no fluxes of matter in and out of the system (a closed system) and no other reactions taking part in the process, then a mixture which at any given moment has a composition specified by $\mathbf{x} = (x_1, \cdots, x_l) \in \mathbb{R}^l$ (for x_i non-negative), will at any other moment have a composition $\mathbf{x} + s\mathbf{R}$, for some number $s \geq 0$. (If the reverse reaction, $-\mathbf{R}$, also occurs, the molar extent s can be negative as well.)

Here and later we denote by $\mathbb{R}^l_+$ the subset of $\mathbb{R}^l$ consisting of all $\mathbf{x}$ with non-negative coordinates.

The vector $\mathbf{R}$ gives the direction of change in the mixture composition, but it says nothing about the speed at which that change is taking place. We make now the theoretical assumption that it is possible, at least in principle, to determine for a given reaction a *rate function* r that specifies that speed. Then r is to be a function of $\mathbb{R}^l_+$ and possibly of other parameters, such as temperature and pressure; it will be written, typically, as $r(\mathbf{x})$, or as $r(x_1, \cdots, x_l)$, or $r(\mathbf{x}, T, P)$, or in some other similar form. We ignore for now any dependence on parameters,.

By definition, r is the rate of change of the molar extent in time, when the mixture composition is $\mathbf{x}$: $r(\mathbf{x}) = \frac{ds}{dt}$. If r is known, then the change in $\mathbf{x}$ due to the reaction is completely determined by the vector field $r\mathbf{R}$ on $\mathbb{R}^l_+$. This vector field is shown in the next figure.



Another way to interpret the vector field $r\mathbf{R}$ is by saying that over a small interval of time of length $\Delta t = t_{\text{final}} - t_{\text{initial}}$, the change in $\mathbf{x}$ is given by

$$\mathbf{x}_{\text{final}} = \mathbf{x}_{\text{initial}} + r(\mathbf{x}_{\text{initial}})\Delta t\mathbf{R} + o(\Delta t).$$

The notation $o(\Delta t)$ represents a small error that goes 0 with Δt and is small compared to Δt. More precisely, it represents an unspecified function whose only relevant property is that $o(\Delta t)/\Delta t \to 0$ as $\Delta t \to 0$.

More generally, suppose that k reactions participate in the process, whose reaction vectors are $\mathbf{R}_1, \cdots, \mathbf{R}_k$. Our fundamental mathematical assumption concerning the kinetic properties will be that it is possible, at least in principle, to define rate functions $r_1, \cdots, r_k$, one for each reaction, in such a way that the change of $\mathbf{x}$ during a short interval of time of length Δt is the sum of the $\Delta s\mathbf{R}$ contributed by each reaction, up to a small error of the order $o(\Delta t)$:

$$\mathbf{x}_{\text{final}} = \mathbf{x}_{\text{initial}} + r_1(\mathbf{x}_{\text{initial}})\Delta t\mathbf{R}_1 + \cdots + r_k(\mathbf{x}_{\text{initial}})\Delta t\mathbf{R}_k + o(\Delta t).$$

Dividing this expression by Δt and passing to the limit as $\Delta t \to 0$ gives the differential equation (in vector form):

$$\dot{\mathbf{x}} = \sum_{i=1}^{k} r_i(\mathbf{x})\mathbf{R}_i.$$

Let us look at one example. Suppose that the process involves substances X_1, X_2, X_3, and a pair of reactions

$$X_1 + X_2 \rightleftharpoons X_3.$$

whose rate functions are: $r_1 = ax_1x_2$ for the forward reaction and bx_3 for the reverse reaction, where a and b are constants.

First note that the reaction vector for the forward reaction is $\mathbf{R} = -X_1 - X_2 + X_3 = (-1,-1,1)$. (Recall that in this expression the X_i represent standard basis vectors.) For the reverse reaction the vector is $-\mathbf{R}$. Therefore

$$\begin{aligned}(\dot{x}_1, \dot{x}_2, \dot{x}_3) &= \dot{\mathbf{x}} \\ &= r_1\mathbf{R} + r_2(-\mathbf{R}) \\ &= (ax_1x_2 - bx_3)(-1,-1,1) \\ &= (-ax_1x_2 + bx_3, -ax_1x_2 + bx_3, ax_1x_2 - bx_3)\end{aligned}$$

Written as a system of ordinary differential equations, this is:

$$\begin{aligned}\dot{x}_1 &= -ax_1x_2 + bx_3 \\ \dot{x}_2 &= -ax_1x_2 + bx_3 \\ \dot{x}_3 &= \quad ax_1x_2 - bx_3.\end{aligned}$$

To this system it may be added initial values: $x_1(0) = q_1, x_2(0) = q_2, x_3(0) = q_3$.

Exercise 34. *Find an expression for the equilibrium concentrations, that is, the values of x_1, x_2, x_3 when $\dot{\mathbf{x}} = 0$, for the system of the previous example.*

Exercise 35. *Solve the above initial value problem. Here are a few suggested steps for doing it.*

(1) *Notice first that $\dot{x}_1 = \dot{x}_2 = -\dot{x}_3$. Therefore $x_1 - x_2$ and $x_1 + x_3$ are constant in time. It follows that $x_2(t) = x_1(t) + q_2 - q_1$. A similar expression for x_3 in terms of x_1 and the initial values holds.*

(2) *Replacing the expressions for x_2 and x_3 obtained above into the first differential equation gives a differential equation in x_1 only, which is of the form*

$$\dot{x}_1 = Ax_1^2 + Bx_1 + C.$$

Determine the precise form of the constants A, B, C, in terms of a, b, q_1, q_2, q_3.

(3) *Suppose that* $a = b = 1, q_1 = 4, q_2 = 3, q_3 = 0$. *Solve for* $x_1(t), x_2(t), x_3(t)$.

You should get for x_1*:*

$$x_1(t) = \frac{6e^{4t} + 2}{3e^{4t} - 1}.$$

6.3. **The Law of Mass Action.** Before we play with other examples of writing the differential equations from a given reaction mechanism, let us stop to consider one type of rate function that will show up very frequently. This is the so-called *mass action law*.

Suppose an elementary reaction of the general form

$$a_1 X_1 + \cdots + a_l X_l \rightharpoonup b_1 X_1 + \cdots + b_l X_l.$$

(For elementary reactions the sum $a_1 + \cdots + a_l$ rarely exceeds 2.) Then the mass action law rate is defined by the expression

$$r(\mathbf{x}) = k x_1^{a_1} x_2^{a_2} \cdots x_l^{a_l}.$$

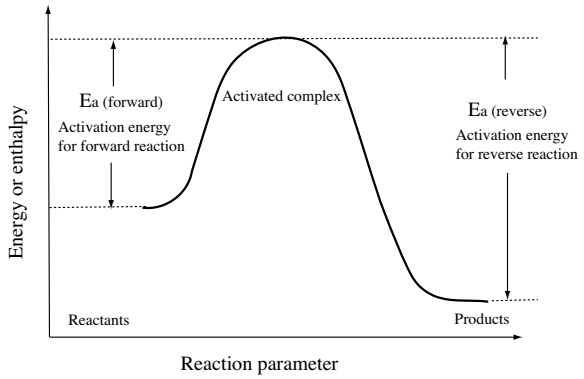
The constant k will typically depend on parameters such as temperature and pressure, but not on the concentrations $x_1, \cdots, x_l$. The dependence of k on the temperature is often adequately described by the form proposed by Arrhenius in the 19 century:

$$k = A \exp\left(-\frac{E_a}{RT}\right).$$

Here E_a is called the *activation energy*, R is the so-called *gas constant*, and A (appropriately enough) is the *pre-exponential factor*.

We will later discuss this expression in the context of probability theory. For now, we may understand it as follows. Consider a reaction of the form $X_1 + X_2 \rightharpoonup$ products. The expression Ax_1x_2 describes the number of collisions between a molecules of X_1 and one of X_2. Not all collisions effectively lead to a reaction. The fraction of molecules with enough energy for the reaction to succeed is proportional to the exponential term $\exp\left(-\frac{E_a}{RT}\right)$. That minimum energy (per mole of X_1 or X_2) is the activation energy E_a.

The precise form that the rate function can take for a specific reaction depends on many complicated factors having to do with the microscopic properties of the interacting molecules, such as their electronic structure and geometry. The task of relating these microscopic properties of the reaction with the macroscopic properties expressed in the rate function is the subject of so-called *collision theory*. For the most part this theory is better established for reactions involving gases and to some extent dilute liquid solutions. The mass action law is generally accepted as the appropriate rate function in these cases.



I will often use the rate function given by the mass action law uncritically. I hope to return to a more detailed discussion of it, and of collision theory in general, later in the course.

Sometimes empirical deviation from the mass action law may signal that a reaction is not elementary, and that the precise reaction mechanism involves intermediate substances and additional reaction steps.

A classical example is the formation of phosgene, $X_1 = COCl_2$, from carbon monoxide, $X_2 = CO$, and chlorine, $X_3 = Cl_2$. (Phosgene is a highly toxic, irritating and corrosive gas that can cause fatal respiratory damage when inhaled. It is used in organic synthesis, in manufacture of dyes, pharmaceuticals, herbicides, insecticides, synthetic foams, resins, and polymers. It first came into prominence during World War I, when it was used against troops, as a chemical weapon.)

If the reaction were elementary, of the form $X_2 + X_3 \rightharpoonup X_1$, one would expect the rate function to be $r = kx_2x_3$, according to the mass action law, where x_2 and x_3 are, respectively, the concentrations of carbon monoxide and chlorine. In reality, the reaction rate far from equilibrium is found to be

$$r = x_2 x_3^{3/2}.$$

To explain this it has been suggested that besides the three species X_1, X_2, X_3 there are also present chlorine atoms, $X_4 = Cl$, and an intermediate $X_5 = COCl$, and that the reaction takes place in three steps that satisfy the mass action law:

(1) $$Cl_2 \rightleftharpoons 2Cl$$

(2) $$Cl + CO \rightleftharpoons COCl$$

(3) $$COCl + Cl_2 \rightleftharpoons COCl_2 + Cl.$$

It is further supposed that reactions (1) and (2) are so fast compared to (3) that they can always be held to be at equilibrium.

Let us try to determine under these assumptions what reaction rate would obtain for the overall reaction $X_2 + X_3 \rightharpoonup X_1$.

The differential equation associated to this mechanism is given by

$$\dot{\mathbf{x}} = (r_1^+ - r_1^-)\mathbf{R}_1 + (r_2^+ - r_2^-)\mathbf{R}_2 + (r_3^+ - r_3^-)\mathbf{R}_3,$$

where $\mathbf{R}_1 = -X_3 + 2X_4$, $\mathbf{R}_2 = -X_2 - X_4 + X_5$, $\mathbf{R}_3 = X_1 - X_3 + X_4 - X_5$, and the rate functions are given as follows (we denote by r_i^+ and r_i^- the rate functions for the reactions associated to $\mathbf{R}_i$ and

$-\mathbf{R}_i$, respectively):

$$\begin{aligned} r_1^+ &= k_1^+ x_3 \\ r_1^- &= k_1^- x_4^2 \\ r_2^+ &= k_2^+ x_2 x_4 \\ r_2^- &= k_2^- x_5 \\ r_3^+ &= k_3^+ x_3 x_5 \\ r_3^- &= k_3^- x_1 x_4. \end{aligned}$$

The assumption that the x_i are equilibrium concentrations for the first two reactions amounts to imposing $r_1^+ = r_1^-$ and $r_2^+ = r_2^-$. That is,

$$\begin{aligned} x_4^2 &= \frac{k_1^+}{k_1^-} x_3 \\ x_5 &= \frac{k_2^+}{k_2^-} x_2 x_4. \end{aligned}$$

With a little algebra now, the differential equation reduces to:

$$\dot{\mathbf{x}} = (a x_2 x_3^{3/2} - b x_1 x_3^{1/2}) \mathbf{R}_3,$$

where the constants a and b are easily obtained from the k_i^+ and k_i^-.

If initially the concentration of $COCl_2$ is very small, the rate of formation of phosgene is $a x_2 x_3^{3/2}$.

Exercise 36. *Show that the map $\mathcal{E}$ associated to the substances $X_1, \cdots, X_5$ is given by*

$$\mathcal{E} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 2 & 0 & 2 & 1 & 1 \end{pmatrix},$$

and that the kernel of $\mathcal{E}$ has a basis consisting of vectors: $\mathbf{S}_1 = -X_1 + X_4 + X_5$, $\mathbf{S}_2 = -X_2 - X_4 + X_5$ and $\mathbf{S}_3 = X_3 - 2X_4$. This means that any reaction involving $X_1, \cdots, X_5$ that respects atomic balance can be linearly generated from the (possibly only fictitious) reactions:

$$\begin{aligned} Cl + COCl &\rightharpoonup COCl_2 \\ CO + Cl &\rightharpoonup COCl \\ 2Cl &\rightharpoonup Cl_2. \end{aligned}$$

As given above the process is described by a system of 5 differential equations, one for each x_i. There are, however, only 3 independent concentrations, so we should be able to reduce those 5 equations to only 3. That can be done by parametrizing the reaction space through the initial condition $\mathbf{x}(0) = \mathbf{q}$ using the reactions $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_1$ defined in the previous exercise. Thus, write

$$\mathbf{x} = \mathbf{q} + s_1 \mathbf{S}_1 + s_2 \mathbf{S}_2 + s_3 \mathbf{S}_3.$$

In coordinates, this means that

$$\begin{aligned} x_1 &= q_1 - s_1; \\ x_2 &= q_2 - s_2; \\ x_3 &= q_3 + s_3; \\ x_4 &= q_4 + s_1 - s_2 - 2s_3; \\ x_5 &= q_5 + s_1 + s_2. \end{aligned}$$

Furthermore, the can express the reaction vectors $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ in the basis $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3$ as

$$\begin{aligned} \mathbf{R}_1 &= -\mathbf{S}_3; \\ \mathbf{R}_2 &= \mathbf{S}_2; \\ \mathbf{R}_3 &= -\mathbf{S}_1 + \mathbf{S}_3. \end{aligned}$$

We can now write

$$\begin{aligned} \dot{s}_1\mathbf{S}_1 + \dot{s}_2\mathbf{S}_2 + \dot{s}_3\mathbf{S}_3 = \dot{\mathbf{x}} & \\ &= (r_1^+ - r_1^-)\mathbf{R}_1 + (r_2^+ - r_2^-)\mathbf{R}_2 + (r_3^+ - r_3^-)\mathbf{R}_3 \\ &= -(r_1^+ - r_1^-)\mathbf{S}_3 + (r_2^+ - r_2^-)\mathbf{S}_2 + (r_3^+ - r_3^-)(-\mathbf{S}_1 + \mathbf{S}_3) \\ &= -(r_3^+ - r_3^-)\mathbf{S}_1 + (r_2^+ - r_2^-)\mathbf{S}_2 + (r_3^+ - r_3^- - r_1^+ + r_1^-)\mathbf{S}_3. \end{aligned}$$

Since the vectors $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3$ are linearly independent we can equate the respective coefficients and then express the rate functions in terms of s_1, s_2, s_3:

$$\begin{aligned} \dot{s}_1 &= -k_3^+(q_3 + s_3)(q_5 + s_1 + s_2) + k_3^-(q_1 - s_1)(q_4 + s_1 - s_2 - 2s_3); \\ \dot{s}_2 &= k_2^+(q_2 - s_2)(q_4 + s_1 - s_2 - 2s_3) - k_2^-(q_5 + s_1 + s_2); \\ \dot{s}_3 &= k_3^+(q_3 + s_3)(q_5 + s_1 + s_2) - k_3^-(q_1 - s_1)(q_4 + s_1 - s_2 - 2s_3) - k_1^+(q_3 + s_3) + \\ &\quad k_1^-(q_4 + s_1 - s_2 - 2s_3)^2. \end{aligned}$$

The initial conditions are $s_i(0) = 0$ or $i = 1, 2, 3$. This is an initial value problem for which we have little hope of finding explicit solutions. Later, when talking about the theory of dynamical systems, we will consider how to try to extract interesting information from the equations without having to fully solve them analytically.

7. Reaction Mechanisms and Differential Equations

We have already seen how to translate a given reaction mechanism into a system of differential equations from which we can hope to obtain the time dependence of the concentrations, $x_i(t)$. Here we would like to start a more systematic study of the differential equations that arise.

First, it may be worthwhile to write the general form that such equations take assuming the mass action law, since this will be the case most used here. Suppose that there are k elementary reactions involving the l molecular species $X_1, \cdots, X_l$. We denote the reaction vectors by $\mathbf{R}_1, \cdots, \mathbf{R}_k$. It is convenient to write the elementary reaction with vector $\mathbf{R}_i$, for each i, in the form

$$\alpha_{i1}X_1 + \cdots + \alpha_{li}X_l \rightharpoonup \beta_{1i}X_1 + \cdots + \beta_{li}X_l.$$

(We will typically, but not always, assume that the α_{ij}, β_{ij} are small positive integers, such that the sums $\sum_j \alpha_{ij}, \sum_j \beta_{ij}$ are at most 3 for each i, and that $\alpha_{ij}\beta_{ij} = 0$ for all i and j. This latter condition means that a substance that appears on one side of the reaction does not appear on the second.)

A useful notation to use sometimes is to write $\mathbf{a}_i = (a_{i1}, \cdots, a_{i,l})$, $\mathbf{b}_i = (b_{i1}, \cdots, b_{i,l})$, for each reaction and denote $\mathbf{x}^{\mathbf{a}_i} = x_1^{a_{i1}} \cdots x_l^{a_{il}}$. We can now write the system of ordinary differential equations associated to this mass action reaction mechanism in vector form as follows, using that $\mathbf{R}_i = \mathbf{b}_i - \mathbf{a}_i$:

$$\dot{\mathbf{x}} = \sum_{i=1}^{k} k_i \mathbf{x}^{\mathbf{a}_i} (\mathbf{b}_i - \mathbf{a}_i).$$

We will refer to the right-hand side of this equation as the *mass action vector field* associated to the given (mass action) reaction mechanism.

7.1. **Single Pair of Forward-Back Reactions.** Consider a pair of reactions

$$a_1 X_1 + \cdots + a_l X_l \rightleftharpoons b_1 X_1 + \cdots + b_l X_l.$$

Let the rate functions be

$$r_+(\mathbf{x}) = k_+ x_1^{a_1} \cdots x_l^{a_l}$$
$$r_-(\mathbf{x}) = k_- x_1^{b_1} \cdots x_l^{b_l}$$

for the forward and backward reactions, respectively. The reaction vector for the forward reaction is the constant vector

$$\mathbf{R} = \sum_{i=1}^{l} (b_i - a_i) X_i = (b_1 - a_1, \cdots, b_l - a_l).$$

Then $\mathbf{x}(t)$ satisfies the differential equation:

$$\dot{\mathbf{x}} = (r_+(\mathbf{x}) - r_-(\mathbf{x}))\mathbf{R}$$

with initial condition $\mathbf{x}(0) = \mathbf{q}$, where $\mathbf{q} = (q_1, \cdots, q_l)$ is some arbitrary vector of initial concentrations. The reaction space in this case is one-dimensional, so we know that the solution must take the form

$$\mathbf{x}(t) = \mathbf{q} + s(t)\mathbf{R},$$

where $s(t)$ is the molar extent of the reaction at time t. Therefore, rather than write a system of differential equations, one for each x_i, we may try to solve for the single function $s(t)$, from which the values of $x_i(t)$ will follow. The differential equation for $s(t)$ can be obtained by noting that, on the one hand

$$\dot{\mathbf{x}} = \dot{s}\mathbf{R}$$

and on the other

$$\dot{\mathbf{x}} = (r_+(\mathbf{q} + s(t)\mathbf{R}) - r_-(\mathbf{q} + s(t)\mathbf{R}))\mathbf{R}.$$

Therefore

$$\dot{s} = r_+(\mathbf{q} + s(t)\mathbf{R}) - r_-(\mathbf{q} + s(t)\mathbf{R}).$$

This is an initial value problem of the form

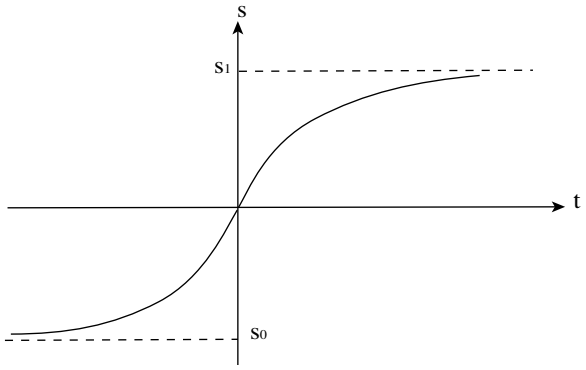
$$\dot{s} = f(s)$$
$$s(0) = 0.$$

This initial value problem can in principle be solved explicitly by noting that the inverse function, $t(s)$, satisfies:

$$t = \int_0^s \frac{du}{f(u)}.$$

If s_0 and s_1 are two zeros of $f(s)$ and no other zero exists on the interval $s_0 < s < s_1$, then $\int_0^s \frac{du}{f(u)}$ is a monotone function of s (either increasing, if $f(s)$ is positive on the interval, or decreasing if

negative). For concreteness let us say that $f(s)$ is positive and that it approaches 0 at the endpoints s_2, s_1 with order $1/s^d$ for $d \geq 1$. Then the integral diverges, that is, $t \to +\infty$ as $s \to s_1$ and $t \to -\infty$ as $s \to s_0$. Therefore the qualitative picture of the solution starting at $s = 0$ is as sketched below.



Example. The decomposition of hydrogen iodide is a reversible reaction of second order given by

$$2HI \underset{k_-}{\overset{k_+}{\rightleftharpoons}} H_2 + I_2.$$

A 200 liter flask was filled with pure HI at 1.24 atm. and 683 K. The decomposition was followed photometrically by measuring the absorption of light by the iodine produced in the reaction. The optical density given is proportional to the iodine concentration. Immediately after the last reading was taken, the flask was chilled and was found to contain 1.17 grams of iodine (atomic weight 127). The following data were obtained:

Time t (min)	42	118	230	397	680	770	940
Optical density	0.81	2.13	3.66	5.04	6.00	6.18	6.21

We wish to see how the data obtained agree with the reaction scheme and use it to estimate the values of the reaction constants k_+ and k_-.

Let us begin by finding the initial concentrations. We set $X_1 = HI$, $X_2 = H_2$, $X_3 = I_2$, and the respective concentrations x_1, x_2, x_3. Initially the flask is filled with pure HI, so $x_2(0) = 0$ and $x_3(0) = 0$. In order to determine $x_1(0)$ we will assume that the perfect gas equation is valid for the HI:

$$PV = nRT$$

where n is the number of moles of HI to be determined, $P = 1.24$ atm., $V = 200$ liters, $T = 683$K, and R is the gas constant:

$$R = 0.0820 \text{ (liter atmosphere) / (degree mole).}$$

Then the molar concentration is

$$x_1(0) = \frac{n}{V} = \frac{P}{RT} = \frac{1.24\text{atm}}{683\text{deg.} \times 0.0820 \text{ (l. atm.) / (deg. mole)}} = 0.022 \text{ moles per liter.}$$

So in units of moles per liter we have (notice that $\mathbf{R} = (-2,1,1)$):

$$x_1(t) = 0.022 - 2s(t)$$
$$x_2(t) = s(t)$$
$$x_3(t) = s(t).$$

Let us now look for the molar extent $s(t)$. The differential equation in vector form is

$$\dot{\mathbf{x}} = (k_+ x_1^2 - k_- x_2 x_3)\mathbf{R}.$$

Writing the equation in terms of s gives:

$$\dot{s} = k_+(0.022 - 2s)^2 - k_- s^2.$$

This is an equation of the form $\dot{s} = f(s)$, where $f(s)$ is a quadratic function in s. With a little algebra it is possible to write

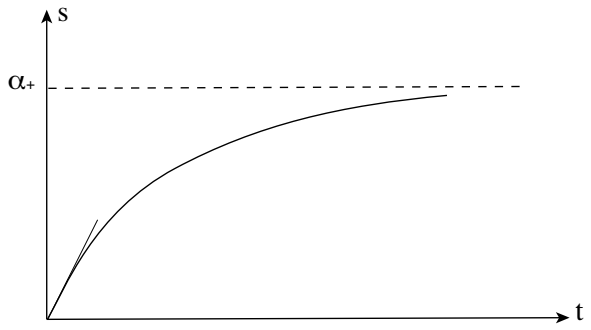
$$f(s) = A(s - \alpha_-)(s - \alpha_+)$$

where $A = 4k_+ - k_-$ and $\alpha_\pm = 0.022/(2 \pm \sqrt{k_-/k_+})$.

Before finding an explicit solution for the above differential equation it is useful to try to anticipate some of the qualitative features the solution should display. From the expression of $f(s)$ we find that $\dot{s} = 0$ if $s = \alpha_-$ or $s = \alpha_+$. At these values we say that the reaction is at equilibrium. At $s = 0$ the rate $\dot{s}$ is given by

$$(4k_+ - k_-)\frac{0.022}{2 + \sqrt{\frac{k_-}{k_+}}}\frac{0.022}{2 - \sqrt{\frac{k_-}{k_+}}} = (0.022)^2 k_+ > 0.$$

Consequently, at and near $s = 0$ the function $s(t)$ grows. Moreover, $\dot{s}$ can only change sign as it crosses $\alpha_\pm$. Notice that $0 < \alpha_+ < |\alpha_-|$. Whether α_- is positive or negative, we have that $\dot{s} > 0$ on the interval $0 \le s < \alpha_+$ and negative on the immediate right of α_+. Therefore $s(t)$ grows monotonically starting at 0 with rate $\dot{s}(0) = (0.022)^2 k_+$, and toward the value $\alpha_+ = 0.022/(2 + \sqrt{k_-/k_+})$. A sketch of the graph of $s(t)$ is shown below.



Given our rough understanding of the solution, let us try to estimate the values of k_- and k_+. Let us next plot the data so that it can be compared with the above graph. Some work is involved in writing the data in terms of s rather than the optical density.

The problem states that the optical density is proportional to the iodine concentration. Call the optical density ρ. Then $x_3 = a\rho$, where a is a proportionality constant. To determine a we use

that at time 940 minutes the flask was found to contain 1.17 grams of iodine, which amounts to $1.17/127 = 9.2\times 10^{-3}$ moles of I. The concentration of I_2 is then

$$x_3(940 \text{ minutes}) = \frac{1}{2}\times \frac{9.2\times 10^{-3}\text{moles}}{200 \text{ liters}} = 2.3\times 10^{-5}\text{moles per liter.}$$

The constant a is then

$$a = \frac{2.3\times 10^{-5}}{6.21} = 0.4\times 10^{-5} \text{ moles per liter.}$$

We can now rewrite the table in terms of x_3. (Recall that $x_3 = s$.)

Time t (min)	42	118	230	397	680	770	940
x_3 (10^{-5} moles per liter)	0.32	0.85	1.46	2.02	2.40	2.47	2.48

A crude estimation of k_- and k_+ can be obtained by using the values of the asymptote α_+, about 2.5×10^{-5} liters per mole, and slope $\dot{s}(0) = (0.022)^2 k_+$ calculated using by the first data point . This gives:

$$k_+ = \frac{0.32\times 10^{-5}}{(0.022)^2\times 42} = 1.6\times 10^{-4} \text{ liters per mole per minute;}$$

$$k_- = \left(\frac{0.022}{2.5\times 10^{-5}} - 2\right)^2 k_+ = 0.9\times 10^3 \times k_+ = 0.15 \text{ liters per mole per minute.}$$

Let us finally look for an explicit solution. Integrating $\int A dt = \int f(s)^{-1} ds$ gives:

$$\begin{aligned} At &= \int_0^s \frac{du}{(u-\alpha_-)(u-\alpha_+)} \\ &= \frac{1}{\alpha_- - \alpha_+}\left(\int_0^s \frac{du}{u-\alpha_-} - \int_0^s \frac{du}{u-\alpha_+}\right) \\ &= \frac{1}{\alpha_- - \alpha_+}\left(\ln\left|\frac{s-\alpha_-}{\alpha_-}\right| - \ln\left|\frac{s-\alpha_-}{\alpha_-}\right|\right) \\ &= \frac{1}{\alpha_- - \alpha_+}\ln\left|\frac{(s-\alpha_-)\alpha_+}{(s-\alpha_+)\alpha_-}\right| \end{aligned}$$

Notice that $A(\alpha_- - \alpha_+) = 0.044 k_+\sqrt{k_-/k_+}$ is positive and that for $0 \le s < \alpha_+$

$$\frac{(s-\alpha_-)\alpha_+}{(s-\alpha_+)\alpha_-} > 0.$$

Therefore,

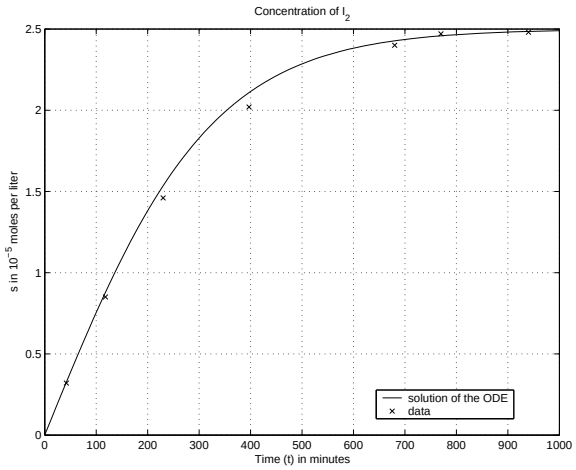
$$A(\alpha_- - \alpha_+)t = \ln\frac{(s-\alpha_-)\alpha_+}{(s-\alpha_+)\alpha_-}$$

It is now easy to solve for s. A little algebra gives:

$$s(t) = \alpha_+\left(e^{A(\alpha_- - \alpha_+)t} - 1\right)\left(e^{A(\alpha_- - \alpha_+)t} - \frac{\alpha_+}{\alpha_-}\right)^{-1}.$$

With our estimated values for the constants we obtain:

$$s(t) = 2.5\times 10^{-5}\frac{e^{6.2\times 10^{-3}t} - 1}{e^{6.2\times 10^{-3}t} + 0.99}.$$



The above graph was produced by Matlab using the commands:

```
x=[42, 118, 230, 239, 680, 770, 940];
y=[0.32, 0.85, 1.46, 2.02, 2.40, 2.47, 2.48];
plot(x,y,'x')
xlabel('Time (t) in minutes')
ylabel('s in 10^-5 moles per liter')
title('Concentration of I2')
hold on
fplot('2.5*(exp(6.2*10^(-3)*x)-1)/(exp(6.2*10^(-3)*x)+0.99)',[0,1000])
```

7.2. **Equilibrium Concentrations.** Suppose that a chemical process involves the molecular species $X_1, \cdots, X_l$ and a reaction mechanism that consists of elementary reactions (indicated by their reaction vectors) $\pm\mathbf{R}_1, \cdots, \pm\mathbf{R}_k$ and respective reaction rates $r_1^{\pm}(\mathbf{x}), \ldots, r_k^{\pm}(\mathbf{x})$. If all changes in $\mathbf{x}$ are due to the reactions only (i.e., if there are no fluxes and diffusion can be neglected) then the rate of change of $\mathbf{x}$ is given by $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, where

$$\mathbf{f}(\mathbf{x}) = (r_1^+(\mathbf{x}) - r_1^-(\mathbf{x}))\mathbf{R}_1 + \cdots + (r_k^+(\mathbf{x}) - r_k^-(\mathbf{x}))\mathbf{R}_k.$$

We say that $\mathbf{x} \in \mathbb{R}_+^l$ is a *critical* point (or a *zero*) of the system of differential equations if $\mathbf{f}(\mathbf{x}) = 0$. (You should be aware that chemists seem to make finer distinctions between the terms *equilibrium* and *stationary* points, whereas the term *critical* point seems not to be much used by them in the sense we are using it. I will observe here the terminology used by mathematicians. We reserve the term "equilibrium" to mean "detailed equilibrium of closed systems," which is explained below.)

Notice that if the system reaches an equilibrium state, at which concentrations do not change in time, then that point must be a critical point of the vector field $\mathbf{f}$. Knowing the zeros of $\mathbf{f}$ is thus a first step in understanding how the process evolves in time.

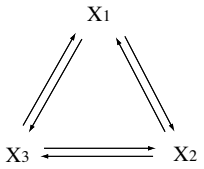
There is a subtlety here that I need to point out. If a point $\mathbf{x}$ is such that $r_i^+(\mathbf{x}) = r_i^-(\mathbf{x})$ for each i, then clearly it is a critical point. But if the reaction vectors are not linearly independent, the converse is not necessarily true.

It is generally assumed, however, based on statistical physics reasoning of a sort that I do not understand, that if a system has attained thermodynamic equilibrium each pair of forward and back

reaction will separately be at equilibrium; that is, $r_i^+(\mathbf{x}) = r_i^-(\mathbf{x})$ for each i . We say in this case that *detailed balance*, or *detailed equilibrium* holds. If I say that a vector $\mathbf{x}$ represents equilibrium concentrations I will generally mean detailed equilibrium. (If the system is not closed, but admits in and out flows at constant rates, then time independence of of $\mathbf{x}$ is often referred to as a "stationary" situation.)

Example. Let us look at a simple example to clarify this point. Consider a process that involves three isomers X_1, X_2, X_3, and the elementary reactions and respective rate functions given by

$$\pm\mathbf{R}_1 = \pm(X_2 - X_1), \;\; r_1^{\pm};$$
$$\pm\mathbf{R}_2 = \pm(X_3 - X_2), \;\; r_2^{\pm};$$
$$\pm\mathbf{R}_3 = \pm(X_1 - X_3), \;\; r_3^{\pm}.$$



Notice that $\mathbf{R}_1 + \mathbf{R}_2 + \mathbf{R}_3 = 0$. Defining $J_i = r_i^+ - r_i^-$ for $i = 1, 2, 3$, we can write the vector field associated with this mechanism as

$$\mathbf{f}(\mathbf{x}) = J_1\mathbf{R}_1 + J_2\mathbf{R}_2 + J_3\mathbf{R}_3 = (J_1 - J_3)\mathbf{R}_1 + (J_2 - J_3)\mathbf{R}_2.$$

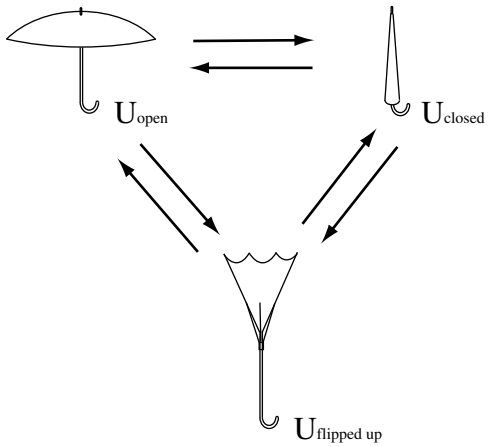
Thus we see that $\mathbf{x}$ is a critical point if

$$J_i(\mathbf{x}) = J_j(\mathbf{x}) \text{ for all } i \text{ and } j,$$

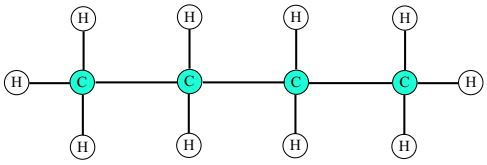
but it is a point of detailed equilibrium if

$$J_i(\mathbf{x}) = 0 \text{ for all } i.$$

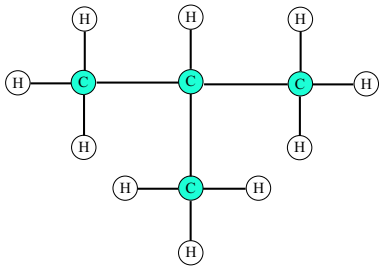
An isomerization process is illustrated by the three pairs of reactions pictured below. Notice that the molecules, U_{open}, U_{closed}, $U_{\text{flipped up}}$ have the same mass and atomic composition. Nevertheless these are very different molecules. A vessel filled with one mole of molecule U_{open} would for example occupy a much greater volume than one mole of U_{closed}.



A more "chemical" example of this notion is provided by the normal-butane and isobutane, two isomers of C_4H_{10}. The next figure describes their molecular graphs.



normal-butane



isobutane

It is interesting to note that by assuming the mass action kinetics and detailed balance the reactions constants must be algebraically related in a certain way. More precisely, since

$$J_1 = k_1^+ x_1 - k_1^- x_2;$$
$$J_2 = k_2^+ x_2 - k_2^- x_3;$$
$$J_3 = k_3^+ x_3 - k_3^- x_1,$$

(the constants $k_i^{\pm}$ are all positive) then by defining the constants $K_i = k_i^+ / k_i^-$ for $i = 1,2,3$, we obtain the relations:

$$x_2 = K_1 x_1, \;\; x_3 = K_2 x_2, \;\; x_1 = K_3 x_3.$$

In particular, it follows that these constants must satisfy

$$K_1 K_2 K_3 = 1.$$

Let us now see what can be said about critical points for this example. We will do this through the following exercises. It is not assumed that $K_1K_2K_3 = 1$ necessarily.

Exercise 37. *Show that* $\mathbf{x}$ *satisfies the equation:*

$$\dot{\mathbf{x}} = (J_3 - J_1)X_1 + (J_1 - J_2)X_2 + (J_2 - J_3)X_3.$$

Conclude that $x_1 + x_2 + x_3$ *is constant by showing that its time derivative is* 0. *Explain why this is true heuristically.*

As the system is essentially two dimensional (it lives in the two dimensional subspace of $\mathbb{R}^3$ spanned by $\mathbf{R}_1$ and $\mathbf{R}_2$), it is possible to express it as a system of two differential equations in two variables. One way to do it to introduce new coordinates, s_1, s_2 by the expression

$$\mathbf{x} = \mathbf{x}(0) + s_1\mathbf{R}_1 + s_2\mathbf{R}_2.$$

(It can be assumed for simplicity that $x_1+x_2+x_3=1$. There is no loss of generality in this since we can use $x_i/(x_1+x_2+x_3)$ as new coordinates. Notice that this is a linear change of coordinates since the denominator is a constant.)

Exercise 38. *Show that in terms of s_1, s_2 the differential equations take the form*

$$\begin{pmatrix}\dot{s}_1\\ \dot{s}_2\end{pmatrix}=\begin{pmatrix}-(k_1^-+k_1^++k_3^-) & k_1^--k_3^+\\ k_2^+-k_3^- & -(k_2^-+k_2^++k_3^+)\end{pmatrix}\begin{pmatrix}s_1\\ s_2\end{pmatrix}+\begin{pmatrix}b_1\\ b_2\end{pmatrix}.$$

for constants b_1 and b_2. These constants are given by

$$\begin{pmatrix}b_1\\ b_2\end{pmatrix}=\begin{pmatrix}k_1^++k_3^- & -k_1^- & -k_3^+\\ k_3^- & k_2^+ & -(k_2^-+k_3^+)\end{pmatrix}\begin{pmatrix}x_1(0)\\ x_2(0)\\ x_3(0)\end{pmatrix}.$$

Therefore the system can be written in the form $\dot{\mathbf{s}}=\mathbf{A}\mathbf{s}+\mathbf{b}$, where $\mathbf{A}$ is a two-by-two matrix and $\mathbf{s}$ and $\mathbf{b}$ are column vectors, all defined in the previous exercise.

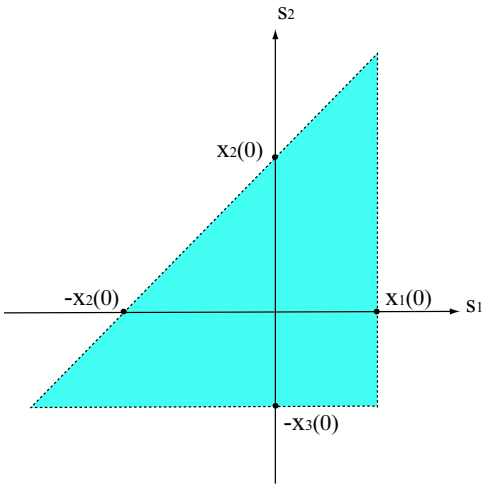
The main qualitative property satisfied by the solutions is stated in the following exercise. This will be discussed further in a later section.

Exercise 39. *Show that $\mathbf{A}$ has positive determinant and negative trace. Conclude that the eigenvalues of $\mathbf{A}$ have negative real part. Explain why the following claim holds: the system has a unique critical point which is attractive. (See below.) Also show that $K_1K_2K_3=1$ if and only if the unique critical point satisfies the detailed balance property.*

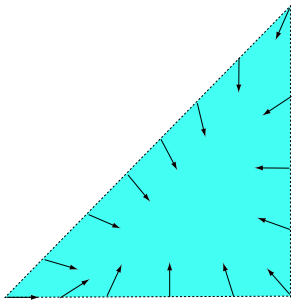
The claim of the exercise can be made more precise as follows. We should first determine the range of values of s_1 and s_2. From the equation $\mathbf{x}=\mathbf{x}(0)+s_1\mathbf{R}_1+s_2\mathbf{R}_2$ we obtain: $x_1=x_1(0)-s_1$, $x_2=x_2(0)+s_1-s_2$, and $x_3=x_3(0)+s_2$. Then by noting that x_1,x_2,x_3 are non-negative we get:

$$\begin{aligned}x_1(0)-s_1&\geq 0;\\ x_2(0)+s_1-s_2&\geq 0;\\ x_3(0)+s_2&\geq 0.\end{aligned}$$

The intersection set for these inequalities is the triangular region represented in the next figure.



Is is a simple (but slightly tedious) exercise to check that the vector field $\mathbf{f}(\mathbf{s}) = \mathbf{A}\mathbf{s} + \mathbf{b}$ always points inward from the boundary of the triangle.



From this simple remark, without knowing the explicit solution of the system of ODEs, it is possible to derive a number of general conclusions:

(1) For all initial conditions $\mathbf{s}(0) = (s_1(0), s_2(0))$ located on the triangular region, the solution $\mathbf{s}(t)$ is always in that region for all $t \geq 0$.
(2) In the interior of the region (that is, in the region but not over the boundary) there exists a critical point, $\mathbf{s}_*$. This is a fixed point in that the solution with initial condition $\mathbf{s}_*$ is the constant function equal to $\mathbf{s}_*$ for all t.
(3) The fixed point is unique and any solution $\mathbf{s}(t)$ converges to $\mathbf{s}_*$ as $t \to \infty$.

To explain these conclusions we need to recall the definition of the flow associated to a vector field and the existence and uniqueness theorem of ODEs. (There is also a nifty fact called *Brower's fixed point theorem* that will bear on these claims.) We will return to these topics in the next subsection where the above claims will be explained.

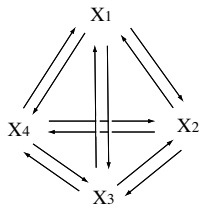
Exercise 40. *Give a general graphical description of reaction mechanisms with mass action kinetics that are associated to linear ODEs. (Suppose that the reversed reaction of each elementary reaction step is also an elementary reaction of the mechanism.)*

Exercise 41. *Suppose that the three reactions proposed earlier for the formation of phosgene are in detailed balance. As before, write $X_1 = COCl_2$, $X_2 = CO$, $X_3 = Cl_2$, $X_4 = Cl$ and $X_5 = COCl$. Show that (by assigning the symbols appropriately) the concentrations must satisfy:*

$$K_1 = x_3^{-1} x_4^2;$$
$$K_2 = x_2^{-1} x_4^{-1} x_5;$$
$$K_3 = x_1 x_3^{-1} x_4 x_5^{-1}.$$

Exercise 42. *Suppose that a reaction mechanism has reaction vectors $\pm\mathbf{R}_s = \mathbf{b}_s - \mathbf{a}_s$, $s = 1, \cdots, k$, and mass action kinetics, so that the reaction rates are $r_s^+(\mathbf{x}) = k_s^+ \mathbf{x}^{\mathbf{a}_s}$, $r_s^-(\mathbf{x}) = k_s^- \mathbf{x}^{\mathbf{b}_s}$. Let $\mathbf{x}^*$ denote a point of detailed equilibrium such that $x_i^* > 0$ for all i, and define $\mathbf{l} = (\ln x_1^*, \cdots, \ln x_l^*)$. Show that for any set of numbers c_s that satisfy $\sum_s c_s \mathbf{R}_s \cdot \mathbf{l} = 0$, we must have $\prod_s K_s^{c_s} = 1$.*

Exercise 43. *Consider the system of first order reactions $X_i \rightleftharpoons X_j$ represented by the figure, where i and j range from 1 to 4, and reaction rates $k_{ij} x_i$. Define $K_{ij} = k_{ij}/k_{ji}$.*



Suppose that the system admits a point **x** *of detailed balance. Find as many relations among the* K_{ij} *as you can.*

8. Temperature and Chemical Equilibrium

9. A Few Application of Chemical Equilibrium

9.1. **Natural Acidity of Biological Processes.** Many organisms are able to change ammonium nitrogen (NH_4^+) to nitrate nitrogen (NO_3^-) by the overall reaction

$$NH_4^+ + 2O_2 \rightharpoonup NO_3^- + 2H^+ + H_2O. \tag{nitr.}$$

This process, called nitrification, is carried out by bacteria that live in the soil. Common sources of ammonium in the soil result from decaying plants and organic matter, or ammonium can come from the application of manure or nitrogen fertilizers. The nitrification process is a source of acidity.

Problem 44. *In typical freshwater lakes, the nitrification rate averages about* 10^{-4} *moles of N per liter per year. This produces an influx of* H^+ *in the lake that increases the acidity of the lake water. Assume the residence time of the water in the lake is half a year. For comparison, suppose that the same acidity results from the lake receiving inflow of water with high pH due, say, to acid rain. What pH should the rain water have that will cause the lake to have the same acidity as produced by the nitrification process?*

It will be necessary to explain first some basic facts about acids. An acid may be defined as a hydrogen-containing substance that dissociates in water to produce hydrogen ions H^+ (actually hydronium ions, H_3O^+). Pure water contains hydrogen ions in concentration 1×10^{-7} moles per liter, and hydroxide ions, OH^- in the same concentration. These ions are formed by dissociation of water:

$$H_2O \rightleftharpoons H^+ + OH^-.$$

When a small amount of acid is added to pure water the concentration of hydrogen ion is increased. The concentration of hydroxide ion then decreases. Acidic solutions contain hydrogen ion in large concentration and hydroxide ion in very small concentration.

Instead of saying that the concentration of hydrogen ion in pure water is 1.0×10^{-7} M (moles per liter), we say that the *p*H of pure water is 7. In general, the *p*H of a solution is defined as the number $-\log[H^+]$, where $[H^+]$ denotes the concentration of hydrogen ion (in moles per liter) and log is the base 10 logarithm. In other words,

$$[H^+] = 10^{-p\mathrm{H}}.$$

The following table gives the approximate *p*H level of a few substances.

pH scale	Substances	acid/base
0		
1	battery acid	
2.4	lemon juice	
3	vinegar, mayonnaise	
3.4	California Chardonnay	⇑ increasing acidity
4.5		
5.5	bread	
6	milk	
7		neutral
7.8	egg whites	
8		
9	baking soda, sea water	
10	soap	
10.5	milk of magnesia	
11		⇓ increasing alkalinity
12		
12.5	ammonia	
13		
13.5	lye	
14		

We can now get back to the problem. From the nitrification reaction (nitr.) we see that for every mole of NH_4^+ that is nitrified to NO_3^-, two mole of H^+ are produced. Hence, if 10^{-4} moles of N per liter are nitrified each year, then 2×10^{-4} moles of H^+ per liter will also be produced each year. Let V denote the volume of the lake, in liters. The product

$$2\times 10^{-4}\times V \text{ moles per year}$$

is the inflow of H^+ due to nitrification.

Suppose now that instead of nitrification the cause of acidity is the inflow of acidic water, with a pH of a, to be determined. This means that $[H^+]=10^{-a}$ moles per liter of the inflow water. This inflow takes place at a rate $F_W = V/T$ (liters per year), where T is the residence time of water in the lake, assume to be 0.5 year. The inflow of H^+ thus happens at a rate of

$$V[H^+]/T = V\times 10^{-a}/0.5.$$

We wish to find a so that this value equals $2\times 10^{-4}\times V$. Therefore

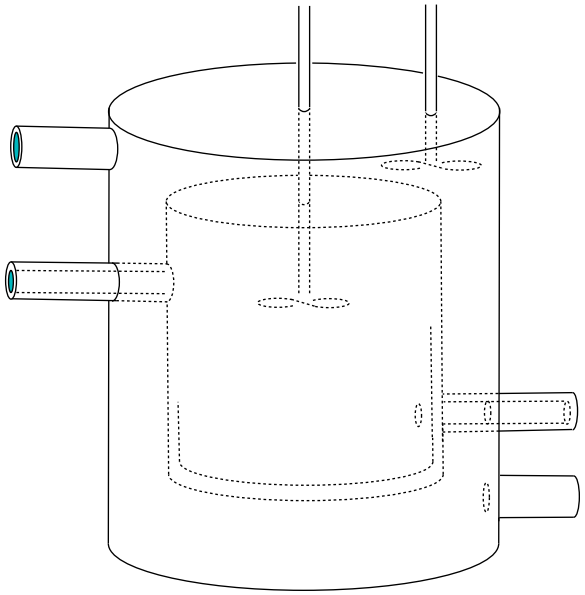
$$2\times 10^{-4}V = V\times 10^{-a}/0.5$$

This shows that $a=4$, that is, a pH of 4.

9.2. **The pH of Pristine Precipitation.**

9.3. **Will the Seas Go Flat?**

10. THE STIRRED TANK REACTOR



11. VOCABULARY

Chemical system: a region of space containing matter of various kinds. It is separated from the outside *ambient* by a *boundary*. Any influence the ambient can have on the system happens through the boundary by the exchange of *matter* and *energy* such as heat, light, etc. The system is *isolated* if there are no exchanges between it and the ambient. The ambient can have no influence on the system if it is isolated. It is a *closed system* if there are no exchanges of matter, although energy transfer is allowed. The system is *open* if both matter and energy can move across the boundary.

State quantities. The measurable quantities used to describe the *state* of the system at a given moment in time, such as pressure, temperature, volume, viscosity, electric resistivity, etc., are called *state quantities*. A state quantity is an *extensive quantity* if it is a number associated to the system as a whole having the property of being *additive*. This means that if a system consists of two non-overlapping sub-systems, the value of that quantity for the whole system is the sum of the values for the two sub-systems. Examples are mass, volume, total electric charge, etc. *Intensive quantities* are quantities which are defined at each point of the system, thus specifying a (possibly discontinuous) function of the space variables, such as concentrations of substances, density of charge, temperature, pressure, etc. (Mathematically, extensive quantities are *measures* and intensive quantities are *densities*.)

The system is *uniform* if all intensive quantities are constant functions. A *phase* is a subset of the system on which all intensive quantities are described by continuous functions. For example, a glass partially filled with water has two phases: water and air. (The density of mass is discontinuous on the transition from water to air.) A subset on which all intensive quantities are constant is a *uniform phase*.

Composition of a uniform phase. The global composition of a chemical system (either monophasic or polyphasic) is specified by the amounts of each of its *chemical species*. (Chemical species are also called *substances*.) These amounts are typically given in *moles*, which is a unit proportional to

the number of molecules of a chemical species present in the system. If there are l substances in amounts $n_1, n_2, \cdots, n_l$ (moles) the numbers $x_i = n_i / \sum_j n_j$ is called the *molar fraction* of the respective substance.

The molar fraction, x_i, of a uniform phase can be interpreted as the probability that a molecule drawn at random from the phase will be of the chemical species i. The *concentration* of substance i in a system of volume V is the number $C_i = n_i/V$.

The total amount n_i of a chemical species is the sum of the amounts $n_{i,\alpha}$ of the same species in each of its phases. The phase is indicated by the extra sub-index 'α.'

One may also consider the *mass* m_i of each chemical species in the system. If m_i is the amount of species i in units of mass, then $w_i = m_i / \sum_j m_j$ is the corresponding *mass fraction* and $\rho_i = m_i/V$ is the corresponding *mass concentration*. If M_i represents the mass of one mole of species i, then $m_i = n_i M_i$.

Mass balance in an open system. Let m_{in} be the total mass that enter the system through its boundary over an interval of time from t_1 to t_2, and m_{out} the total mass that leaves the system through the boundary over the same interval. Under the assumption that total mass is conserved, then $\Delta m = m_{\text{in}} - m_{\text{out}}$ is the amount of mass that the system acquired during that interval. There is *accumulation* of matter if Δm is positive, and *evacuation* if negative. If $\Delta m = 0$ over any interval of time we say that the transport of matter across the boundary occurs in a *permanent mass regime*.

If, besides the total amount of matter in the system, all intensive quantities are time independent functions, we say that the system is in an *stationary state*.

Chemical reactions. The amounts n_i may change in time due to transport of the substance in and out of the system and due to chemical reaction. If the n_i change in a closed system, then some chemical reaction has occurred. The substances that decrease in amount due to a chemical reaction are called the *reactants*; those that increase in amount are the *reaction products*. There may be substances that affect the speed of a reaction but whose amounts do not change, such as *catalysts*.

In a closed or stationary system, the quantities (in moles) of each species of atoms is conserved. The changes in the molar quantities of each chemical species due only to reaction (i.e., all changes that are not due to transport of matter across the boundary) must satisfy a *reaction equation*. This is an expression such as

$$N_2 + 3H_2 = 2NH_3,$$

which is interpreted as saying that changes in the molar amounts of nitrogen gas, hydrogen gas and ammonia in the system due to this reaction involves the disappearance of one mole of nitrogen, 3 of hydrogen, and the formation of 2 moles of ammonia.

Traditionally, reactants are written on the left-hand side of the equation and products on right. The numbers in front of the molecular symbols (1 for N_2, 2 for NH_3, for example) are the *stoichiometric coefficients*. These numbers give no indication of the actual amount of matter involved in the equation; only their ratios are meaningful. In this sense, the previous reaction equation and $2N_2 + 6H_2 = 4NH_3$ have exactly the same meaning. The changes in the actual molar amounts of each constituent species of this reaction are thus related by the equations

$$\frac{\Delta n_{H_2}}{-3} = \frac{\Delta n_{N_2}}{-1} = \frac{\Delta n_{NH_3}}{+2} \quad (>0).$$

12. A General Box Model

(Flows, divergence theorem, continuity equation with reactions, etc.; add energy?)

13. ODEs, Vector Fields and Flows.

The most general system of differential equation we weed to consider is of the following form. Let $\mathcal{D}$ denote a subset in $\mathbb{R}^l$ and $\mathbf{f} : \mathcal{D} \to \mathbb{R}^l$ a function. Then our differential equations will be written

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}).$$

(It turns out that most types of differential equation, whether they are of order one or greater, or depend explicitly on time or not, can be written in such a form after some standard manipulation. The equations of chemical kinetics for the most part already come in this form.)

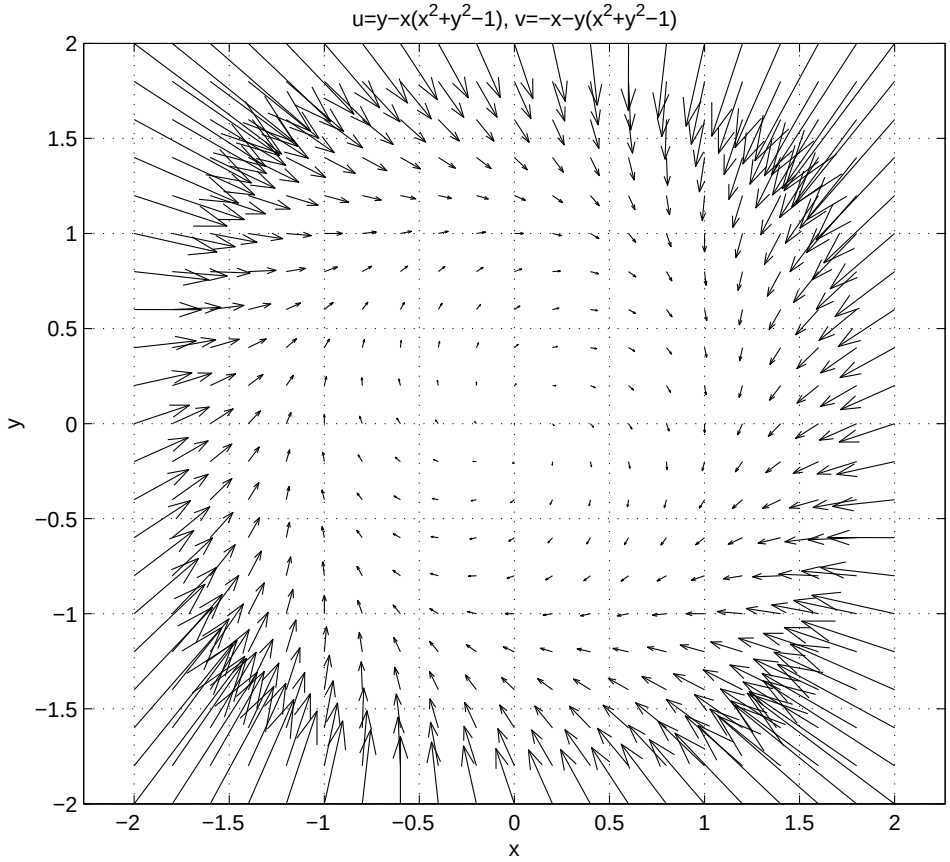
We want to think of $\mathbf{f}$ as a *vector field* on $\mathcal{D}$. For example, consider the two dimensional function defined on the square $\mathcal{D} = \{(x,y) \in \mathbb{R}^2 : |x| \leq 2, |y| \leq 2\}$ given as follows:

$$u = y - x(x^2 + y^2 - 1)$$
$$v = -x - y(x^2 + y^2 - 1).$$

It was used here that $\mathbf{x} = (x,y)$ and $\mathbf{f}(\mathbf{x}) = (u(x,y), v(x,y))$. We interpret $\mathbf{f}$ as a vector field that assigns to each point $(x,y) \in \mathcal{D}$ a vector $(u(x,y), v(x,y))$ with its base point drawn at (x,y) and arrow tip at $(x+u(x,y), y+v(x,y))$.

The vector field can be plotted using Matlab with the commands:

```
% First create a grid of points (r1, r2) on the square of side 4.
%These points are a distance
% 0.2 apart horizontally and vertically.
>> r1=-2:.2:2;
>> r2=-2:.2:2;
>> [x,y]=meshgrid(r1,r2);
% We now define the two coordinates of the vector field:
>> u=y+x -x.*(x.^2+y.^2);
>> v=y-x -y.*(x.^2+y.^2);
% The arrows are produced by the 'quiver' command:
>> quiver(x,y,u,v,4)
% You can change the value '4' to adjust the scale of the
% vector field to your liking.
% By modifying this parameter the end tip of each vector is place not at
% (x+u(x,y), y+v(x,y)) but at (x+au(x,y), y+av(x,y)),
% for an appropriately small 'a'
% that make the picture look less cluttered.
>> grid
>> axis equal
>> title('u=y-x(x^2+y^2-1), v=-x-y(x^2+y^2-1)')
>> xlabel('x')
>> ylabel('y')
% We will shortly draw some flow lines on top of this picture.
% The following command retains the plot just
% obtained so that others produced
% later will be superposed to it.  (To unlock the 'hold on'
% command use 'hold off')
>> hold on
```



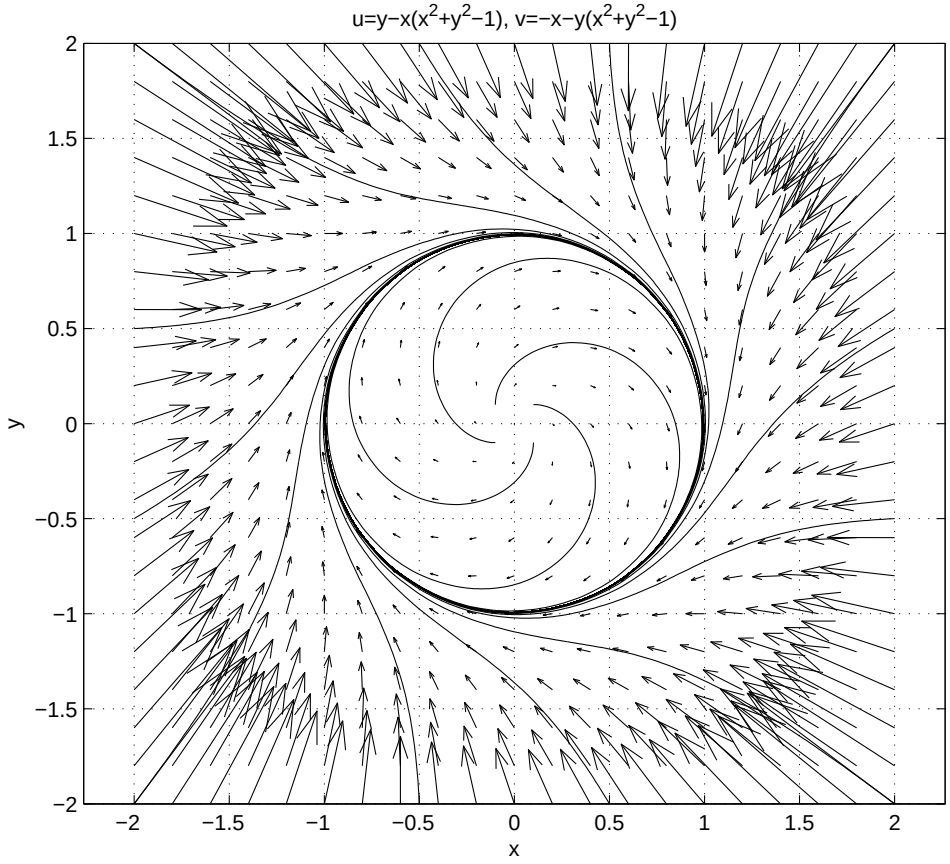
The differential equation $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ can be interpreted as follows. Given an initial point $\mathbf{x}_0$ in $\mathcal{D}$, we look for a curve $\mathbf{x}(t)$ that starts at $\mathbf{x}_0$ when $t = 0$, and has for each t the velocity vector $\mathbf{f}(\mathbf{x}(t))$. In particular, the curve is, at each point, tangent to the vector field at that point.

If the solution curve is at a point $\mathbf{x}(t)$ at a time t, then at a future time $t + \Delta t$ it will find itself at the point

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{f}(\mathbf{x}(t))\Delta t + o(\Delta t).$$

(Recall that $o(a)$ represents any function of a having the property that $o(a)/a$ goes to 0 as a goes to 0.) *Integrating*, or solving, a differential equation of the form $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ then amounts to finding such a curve $\mathbf{x}(t)$. We call this curve an *integral curve* for the vector field $\mathbf{f}$.

We can draw a few solution curves on the previous graph using the Matlab command 'streamline.' The next picture gives the integral lines starting at the points: $(-2,-2)$, $(-2,-0.5)$, $(-2,0.5)$, $(-2,2)$, $(2,-2)$, $(2,-0.5)$, $(2,0.5)$, $(2,2)$, $(-0.5,-2)$, $(0.5,-2)$, $(-0.5,2)$, $(0.5,2)$, $(-0.1,-0.1)$, $(-0.1,0.1)$, $(0.1,-0.1)$, $(0.1,0.1)$.



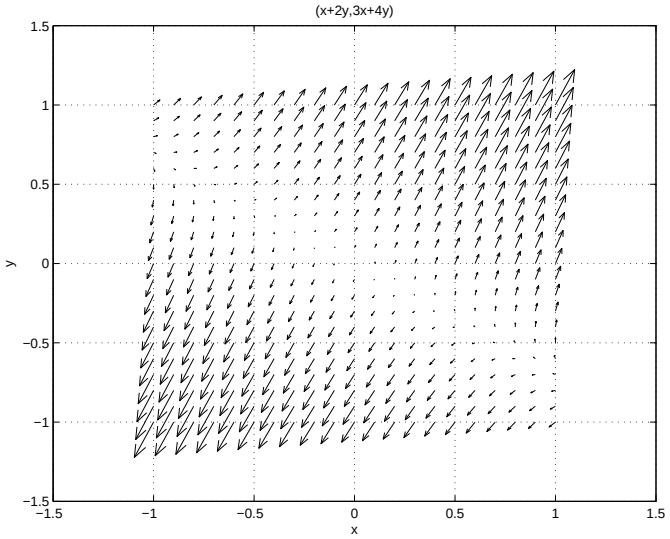
The Matlab commands for the previous figure are:

```
>> x0=[-2 -2 -2 -2 2 2 2 2 -.5 .5 -.5 .5 -.1 -.1 .1 .1];
>> y0=[-2 -.5 .5 2 -2 -.5 .5 2 -2 -2 2 2 -.1 .1 -.1 .1];
>> streamline(x,y,u,v, x0, y0)
>> hold off
```

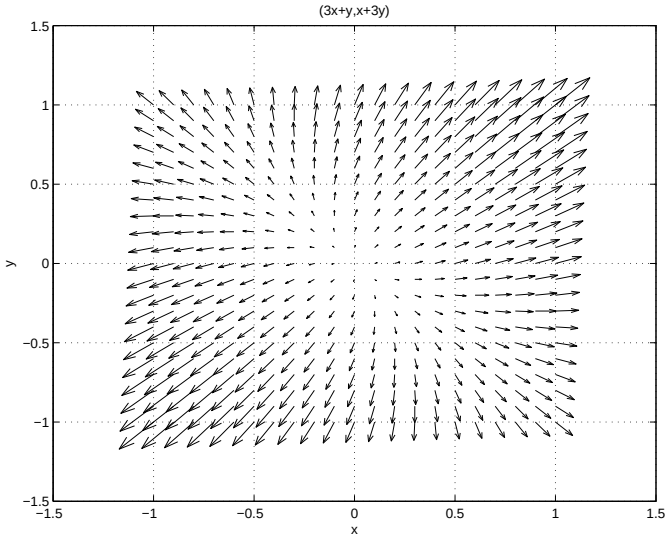
We say that $\mathbf{f}(\mathbf{x})$ is a *linear* vector field if it has the form

$$\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x},$$

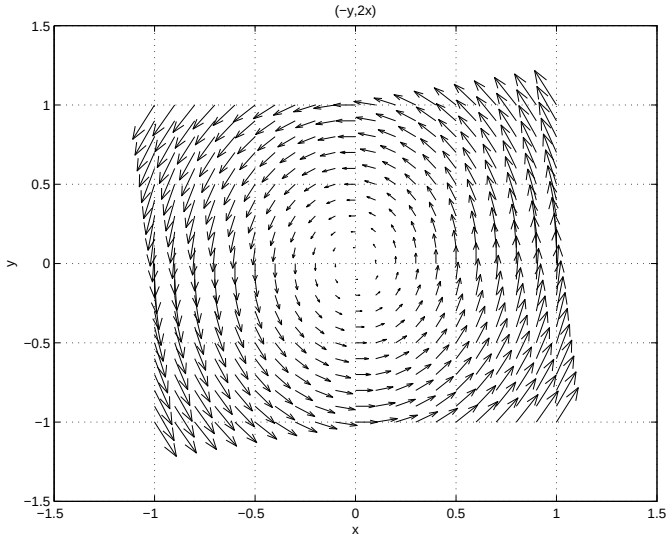
where $\mathbf{A}$ is an l by l real matrix and $\mathbf{x}$ is a column vector. Here are a few examples in dimension 2. The corresponding vector fields are written below the picture.



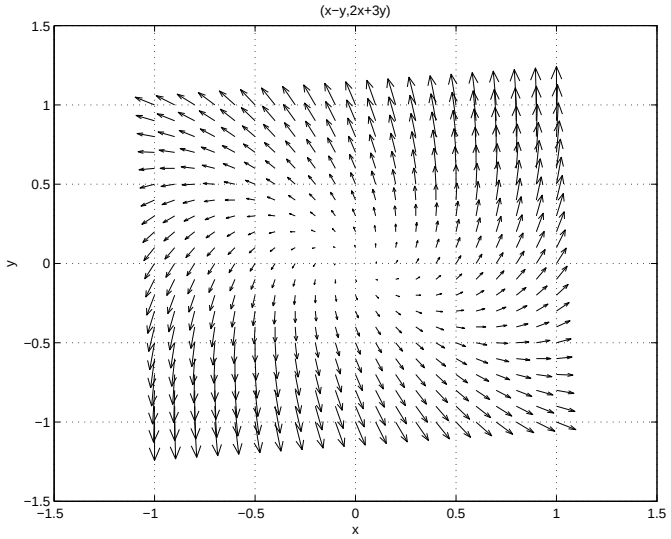
$$\mathbf{f}(\mathbf{x}) = (x + 2y, 3x + 4y)$$



$$\mathbf{f}(\mathbf{x}) = (3x + y, x + 3y)$$



$\mathbf{f}(\mathbf{x}) = (-y, 2x)$



$\mathbf{f}(\mathbf{x}) = (x-y, 2x+3y)$.

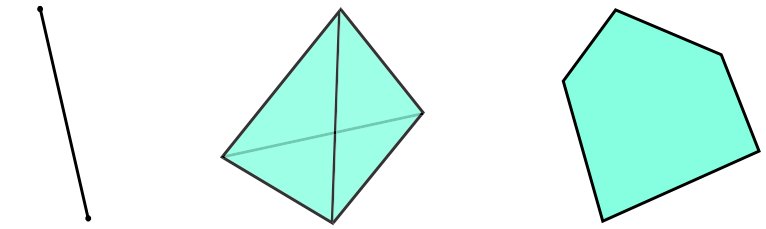
The Matlab commands for the first of the above four figures are:

```
>>clear
>> r1=-1:.1:1;
>> r2=-1:.1:1;
>> [x,y]=meshgrid(r1,r2);
>> u=x+2*y;
```

```
>> v=3*x+4*y;
>> quiver(x,y,u,v,2)
>> grid
>> xlabel('x')
>> ylabel('y')
>> title('(x+2y,3x+4y)')
```

13.1. **The Fundamental Existence-Uniqueness Theorem.** We state the fundamental theorem in a somewhat more restricted form than it is normally described, which will be nevertheless sufficient for our study.

The set $\mathcal{D}$ will be assumed to have the following properties: it is a *compact* set in $\mathbf{R}^l$ whose boundary is a finite union of differentiable surfaces of dimension $l-1$. Compactness can be defined as saying that $\mathcal{D}$ is both *bounded* (that is, $\mathcal{D}$ does not extend to infinity) and *closed* (that is, $\mathcal{D}$ contains all of its boundary points). Whatever the exact mathematical characterization of $\mathcal{D}$, the main examples to keep in mind are the bounded sets defined by a finite collection of inequalities of the form $\mathbf{n}\cdot\mathbf{x} \leq a$. In other words, our main examples are bounded sets defined as the intersection of half spaces. A few of such sets (in $\mathbf{R}^3$) are depicted below.



compact sets which are intersections of half-spaces and affine subspaces

not bounded not closed

As for the vector field, we will suppose that $\mathbf{f}$ is a differentiable function and that on every boundary point $\mathbf{x}$ of $\mathcal{D}$ the vector $\mathbf{f}(\mathbf{x})$ does not point outward.

If $\mathbf{f}$ and $\mathcal{D}$ have these properties we say that the vector field is *nice*.

Theorem 45. *Let $\mathbf{f} : \mathcal{D} \to \mathbb{R}^l$ be a nice vector field. Then the following properties hold:*

(1) *For any $\mathbf{x}_0 \in \mathcal{D}$ there exists a differentiable curve $\mathbf{x}(t) \in \mathcal{D}$, $t \geq 0$, which is the unique integral curve of $\mathbf{f}$ starting at $\mathbf{x}_0$. Denote this curve by $\Phi_t(\mathbf{x}_0)$.*

(2) *For each $t \geq 0$, the function $\Phi_t : \mathcal{D} \to \mathcal{D}$ is differentiable. If $\mathbf{f}$ is differentiable to all orders, so is Φ_t for each $t \geq 0$.*

(3) Φ_t *is a* one-parameter semigroup of transformations *of* $\mathcal{D}$. *This means that (i)* Φ_0 *is the identity function and for all non-negative* s,t, *it holds that:*

$$\Phi_t \circ \Phi_s = \Phi_{t+s}.$$

We will also consider at times vector fields which are not *nice* in the above formal sense, but which are quite good in that a simple existence-uniqueness theorem as the above one also holds. The next theorem is about vector fields whose domain $\mathcal{D}$ is all of $\mathbb{R}^l$. We will say that such a vector field is *super* if it is either a linear vector field, i.e., $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ for an l by l matrix $\mathbf{A}$ (here $\mathbf{x}$ is regarded as a column matrix), or if $\mathbf{f}$ is bounded. The latter means that there exists a constant M such that $\|f(\mathbf{x})\| \leq M$ for all $\mathbf{x}$ in $\mathbb{R}^l$. Here the norm of a vector $\mathbf{u}$ is defined as $\|\mathbf{u}\| = \sqrt{\mathbf{u}\cdot\mathbf{u}}$. You will notice from the conclusion of the next theorem that being 'super' is better than being 'nice.'

Theorem 46. *Let* $\mathbf{f} : \mathbb{R}^l \to \mathbb{R}^l$ *be a vector field which we assume to be super. Then the following properties hold:*

(1) *For any* $\mathbf{x}_0 \in \mathbb{R}^l$ *there exists a differentiable curve* $\mathbf{x}(t) \in \mathcal{D}$, $-\infty < t < \infty$, *which is the unique integral curve of* $\mathbf{f}$ *starting at* $\mathbf{x}_0$. *As before, denote this curve by* $\Phi_t(\mathbf{x}_0)$. *Notice that this is now defined for all* t.

(2) *For each* t, *the function* $\Phi_t : \mathbb{R}^l \to \mathbb{R}^l$ *is differentiable and invertible. If* $\mathbf{f}$ *is differentiable to all orders, so is* Φ_t *for each* $t \geq 0$.

(3) Φ_t *is a* one-parameter group of transformations *of* $\mathbb{R}^l$. *This means that (i)* Φ_0 *is the identity function and for all* s,t, *it holds that:*

$$\Phi_t \circ \Phi_s = \Phi_{t+s}.$$

In particular, for each t

$$\Phi_{-t} = \Phi_t^{-1}.$$

The function Φ_t *will be referred to as the* flow *of the vector field* $\mathbf{f}$.

The technical name for a vector field that satisfies the conclusions of the last theorem is *complete*. Therefore if a vector field is super then it is complete. On the other hand it is possible for a vector field to be complete without being either linear or bounded.

One example will illustrate the main conclusions of the theorem. We write $\mathbf{x} = (x,y)$ and define

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

Notice that this is neither *nice* nor *super*, but it will be seen to be a complete vector field nevertheless. To find the flow Φ_t for this example we need to solve the system:

$$\dot{x} = -x$$
$$\dot{y} = 2y + x^2$$

and initial conditions $x(0) = x_0$, $y(0) = y_0$. The first equation is easy to solve: $x(t) = x_0 e^{-t}$. Substituting $x(t)$ in the second equation gives:

$$\dot{y} = 2y + x_0^2 e^{-2t}.$$

This is a first order linear equation (non-homogeneous now). We have solved such equations earlier in the course. The solution is:

$$y(t) = y_0 e^{2t} + \frac{x_0^2}{2}\sinh(2t).$$

Therefore the flow in this case is:

$$\Phi_t(\mathbf{x}) = \left(xe^{-t}, ye^{2t} + \frac{x^2}{2}\sinh(2t) \right).$$

Exercise 47. *Show by a direct calculation that Φ_t obtained above actually satisfies the property $\Phi_{t+s} = \Phi_t \circ \Phi_s$.*

Exercise 48. *Obtain the flow of the one-dimensional vector field defined by the differential equation*

$$\dot{x} = \frac{x^2}{1+x^2}.$$

In particular, check that this is a complete vector field.

13.2. **Systems of Linear ODEs with Constant Coefficients.** We would like to take some time here to discuss the general properties of equations of the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{h}(t),\ \mathbf{x}(0) = \mathbf{x}_0,$$

where $\mathbf{A}$ is an n-by-n matrix and $\mathbf{x}$ and $\mathbf{h}(t)$ are column vectors.

Recall the solution of the corresponding equation in dimension 1, $\dot{x} = Ax + h(t)$, $x(0) = x_0$:

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-s)}h(s)ds.$$

It turns out the same holds in general, once we know how to make sense of matrix exponentials. In other words, we will see that the solution of the above initial value problem is given by

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}_0 + \int_0^t e^{\mathbf{A}(t-s)}\mathbf{h}(s)ds$$

If all this is to make any sense, the exponential $e^{\mathbf{A}t}$ should be a matrix valued function of t having the following property: For each column vector $\mathbf{x}_0$, the function $\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}_0$ is the unique solution curve of the differential equation $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ with initial value $\mathbf{x}(0) = \mathbf{x}_0$. In other words, the exponential of $\mathbf{A}$ is nothing other than the flow of this differential equation. We take this as our definition of exponential. It remains to show that thus defined the exponential has all the properties we expect of a bona fide exponential function. We will see also how we should go about calculating it explicitly.

First we should note a few general properties. If $\mathbf{A}$ is an n by n matrix, then $e^{\mathbf{A}t}$ is also. The issue here is linearity. More precisely, if $\mathbf{x}_0$ and $\mathbf{y}_0$ are two initial conditions, then the claim is that

$$e^{\mathbf{A}t}(\mathbf{x}_0 + \mathbf{y}_0) = e^{\mathbf{A}t}\mathbf{x}_0 + e^{\mathbf{A}t}\mathbf{y}_0.$$

But this can be seen to follow immediately from the uniqueness part of the main theorem. (Call $\mathbf{x}(t) = e^{\mathbf{A}t}(\mathbf{x}_0 + \mathbf{y}_0)$ and $\mathbf{z}(t) = e^{\mathbf{A}t}\mathbf{x}_0 + e^{\mathbf{A}t}\mathbf{y}_0$, and show that these two functions satisfy the same initial value problem. At a key point you will need to use that $\mathbf{A}$ defines a linear map.)

The next fact is quite useful.

Theorem 49. *Two n by n matrices $\mathbf{A}$ and $\mathbf{B}$ commute, that is, $\mathbf{AB} = \mathbf{BA}$, if and only if for all t*

$$e^{\mathbf{A}t+\mathbf{B}t} = e^{\mathbf{A}t}e^{\mathbf{B}t}.$$

Proof. Let us first show that if the equality of the exponentials holds, so does commutativity of $\mathbf{A}$ and $\mathbf{B}$. For that notice that taking the second derivative in t on both sides of the identity gives (after some calculation which I leave to you as an exercise):

$$(\mathbf{A}+\mathbf{B})^2 = \mathbf{A}^2 + 2\mathbf{AB} + \mathbf{B}^2.$$

Now develop the square on the left hand side (keeping in mind that we do not yet know that the matrices commute!). After some cancelations we arrive at $\mathbf{AB} = \mathbf{BA}$.

For the converse, let us first establish that if $\mathbf{A}$ and $\mathbf{B}$ commute then, for all t, $\mathbf{A}e^{\mathbf{B}t} = e^{\mathbf{B}t}\mathbf{A}$. Define $\mathbf{x}(t) = \mathbf{A}e^{\mathbf{B}t}\mathbf{x}_0$ and $\mathbf{z}(t) = e^{\mathbf{B}t}\mathbf{A}\mathbf{x}_0$. We wish to show that these two curves coincide no matter which $\mathbf{x}_0$ we choose. This is shown by noticing (after a calculation which, again, I leave for you to do) that both $\mathbf{x}$ and $\mathbf{z}$ satisfy the same initial value problem. By uniqueness they must be equal. (By symmetry the same holds with the roles of $\mathbf{A}$ and $\mathbf{B}$ reversed.)

The proof of the main claim is now easy, given what has been proved so far. Again, it involves showing that the functions $e^{\mathbf{A}t+\mathbf{B}t}$ and $e^{\mathbf{A}t}e^{\mathbf{B}t}$ satisfy the same initial value problem. Once more. I leave the details as an exercise. □

The previous theorem generalizes to arbitrary vector fields once we know what it should mean to say that two vector fields commute. It turns out that the right definition is as follows. Let $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \cdots, f_n(\mathbf{x}))$ and $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \cdots, g_n(\mathbf{x}))$ be two vector fields. We say that $\mathbf{f}$ and $\mathbf{g}$ commute if for each $i = 1, \cdot, n$,

$$\sum_{j=1}^{n} \left(g_j \frac{\partial f_i}{\partial x_j} - f_j \frac{\partial g_i}{\partial x_j} \right) = 0.$$

With this definition in place, the following result holds. (To save time I will not prove it here.)

Theorem 50. *Let* $\mathbf{f}$ *and* $\mathbf{g}$ *two commuting vector fields. Then their flows,* Φ_t *and* Ψ_t *also commute. In other words,*

$$\Phi_t \circ \Psi_s = \Psi_s \circ \Phi_t$$

for all s and t.

Let us now calculate a few matrix exponentials explicitly. The most trivial case is of a 1 by 1 matrix, A. Then e^{At} is, by definition, the unique solution of $\dot{x} = Ax$ with initial condition $x(0) = 1$. In other words, it is the ordinary exponential of a number.

If $\mathbf{A}$ is diagonal, say

$$\mathbf{A} = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

then, for each column vector $\mathbf{z}$, $\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{z}$ is the vector valued function of t that satisfies $\cdot x = \mathbf{A}\mathbf{x}$, with initial condition $\mathbf{x}(0) = \mathbf{z}$. This amounts to the system of equations:

$$\begin{aligned} \dot{x}_1 &= \lambda_1 x_1 \\ \dot{x}_2 &= \lambda_2 x_2 \\ &\vdots \\ \dot{x}_n &= \lambda_n x_n, \end{aligned}$$

and initial conditions $x_i(0) = z_i$, for each i. By the first example we have $x_i(t) = e^{\lambda_i t} z_i$. Therefore the matrix $e^{\mathbf{A}t}$ is also diagonal:

$$e^{\mathbf{A}t} = \begin{pmatrix} e^{\lambda_1 t} & 0 & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & e^{\lambda_n t} \end{pmatrix},$$

Knowing how to exponentiate diagonal matrices, suppose now that we have a *diagonalizable* matrix: $\mathbf{A} = \mathbf{C}^{-1}\mathbf{D}\mathbf{C}$, where $\mathbf{D}$ is diagonal and $\mathbf{C}$ is an n by n invertible matrix. Then I claim that

$$\exp(\mathbf{C}^{-1}\mathbf{D}\mathbf{C}t) = \mathbf{C}^{-1}\exp(\mathbf{D}t)\mathbf{C}.$$

This is easily seen to hold by the following remark: Consider the initial value problem: $\dot{\mathbf{y}} = \mathbf{D}\mathbf{y}$, $\mathbf{y}(0) = \mathbf{C}\mathbf{z}$. We know that its unique solution is $\mathbf{y}(t) = e^{\mathbf{D}t}\mathbf{C}\mathbf{z}$. So $\mathbf{x}(t) = \mathbf{C}^{-1}\mathbf{y}(t) = \mathbf{C}^{-1}e^{\mathbf{D}t}\mathbf{C}\mathbf{z}$. As $\mathbf{z}$ is arbitrary we obtain the claimed identity.

We will recall how to go about diagonalizing matrices, when it is possible, or putting them in normal form in general. But for now, let us do one example using Matlab. First some terminology. We say that a column vector $\mathbf{z}$ is an *eigenvector* of $\mathbf{A}$ associated to an *eigenvalue* λ if $\mathbf{z}$ in non-zero and satisfies the equation

$$\mathbf{A}\mathbf{z} = \lambda\mathbf{z}.$$

Notice that if $\mathbf{A} = \mathbf{C}^{-1}\mathbf{D}\mathbf{C}$ for a diagonal matrix $\mathbf{D}$, then the columns of $\mathbf{C}^{-1}$ are eigenvectors of $\mathbf{A}$ and the diagonal entries of $\mathbf{D}$ are the respective eigenvalues.

Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & 4 & 5 & 6 & 7 & 1 \\ 3 & 4 & 5 & 6 & 7 & 1 & 2 \\ 4 & 5 & 6 & 7 & 1 & 2 & 3 \\ 5 & 6 & 7 & 1 & 2 & 3 & 4 \\ 6 & 7 & 1 & 2 & 3 & 4 & 5 \\ 7 & 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix}.$$

To enter such a matrix into Matlab, we type

```
>> A=[1 2 3 4 5 6 7;
2 3 4 5 6 7 1;
3 4 5 6 7 1 2;
4 5 6 7 1 2 3;
5 6 7 1 2 3 4;
6 7 1 2 3 4 5;
7 1 2 3 4 5 6];
```

We are going to call on Matlab's eigenvalue-eigenvector command. Type

```
>>[Cinverse,D]=eig(A)
```

This produces two matrices named `Cinverse` and `D`. The former is our $\mathbf{C}^{-1}$ and the latter is our $\mathbf{D}$. Matlab's output is:

$$\mathbf{C}^1 = \begin{pmatrix} -0.4526 & -0.5045 & -0.5312 & 0.0598 & 0.1765 & 0.2844 & 0.3780 \\ -0.5045 & -0.0598 & 0.4526 & -0.2844 & -0.5312 & -0.1765 & 0.3780 \\ -0.1765 & 0.5312 & -0.2844 & 0.4526 & 0.0598 & -0.5045 & 0.3780 \\ 0.2844 & -0.1765 & 0.0598 & -0.5312 & 0.5045 & -0.4526 & 0.3780 \\ 0.5312 & -0.4526 & 0.1765 & 0.5045 & -0.2844 & -0.0598 & 0.3780 \\ 0.3780 & 0.3780 & -0.3780 & -0.3780 & -0.3780 & 0.3780 & 0.3780 \\ -0.0598 & 0.2844 & 0.5045 & 0.1765 & 0.4526 & 0.5312 & 0.3780 \end{pmatrix}$$

$$\mathbf{D} = \begin{pmatrix} -8.0667 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -4.4767 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3.5900 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3.5900 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4.4767 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8.0667 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 28.0000 \end{pmatrix}$$

We can also obtain $\mathbf{C}$ by taking the inverse matrix of $\mathbf{C}^{-1}$. Matlab's command for inverse is

```
>> Cinverse^(-1)
```

It just so happens (we will see why later on) that in this case the inverse of $\mathbf{C}$ (or of its inverse) is just the transpose matrix. (In Matlab the transpose of a matrix `A` is written `A'`.)

Thus we obtain

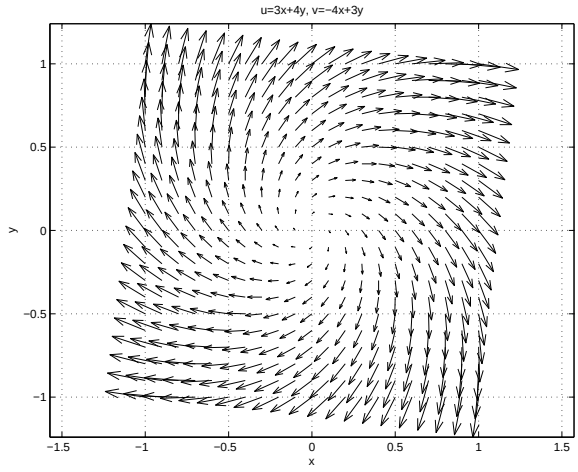
$$e^{\mathbf{A}t} = \mathbf{C}^{-1} \begin{pmatrix} e^{-8.0667t} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{-4.4767t} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{-3.59t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{3.59t} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{4.4767t} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{8.0667t} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & e^{28} \end{pmatrix} \mathbf{C}.$$

We could multiply it out, if we wanted, but the factorized form is probably more useful.

Let us look at another example. Consider

$$\mathbf{A} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

If you calculate its eigenvalues you will see that they are $a \pm ib$. The method used above also works when the matrices are complex as we will see in a moment. It may be instructive, however, to first find the exponential of$\mathbf{A}$ by a different method that does not require going into complex numbers and then comparing the result with what would be obtained by diagonalization. First, let us examine the portrait of the linear vector field defined by $\mathbf{A}$ to see if it offers some geometric clue.



One feature of the graph that can be seen after some experimentation to independent of the choice of a and b is that it seems to be invariant under rotations about the origin. With this in mind let us change the differential equation to polar coordinates before attempting to solve it.

Thus we take $x = r\cos\theta$ and $y = r\sin\theta$. Observe that

$$\dot{x} = \dot{r}\cos\theta - r\dot{\theta} = r(a\cos\theta + b\sin\theta)$$
$$\dot{y} = \dot{r}\sin\theta + r\dot{\theta}\cos\theta = r(-b\cos\theta + a\sin\theta).$$

The pair of equalities on the right-hand side can be written in matrix form as

$$\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\dot{r}\\ r\dot{\theta}\end{pmatrix} = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}ar\\ -br\end{pmatrix}.$$

We can cancel the (invertible) matrix on each side of the equality. Therefore

$$\dot{r} = ar$$
$$\dot{\theta} = -b.$$

This is a very easy system to integrate. The solution is

$$r(t) = r(0)e^{at}$$
$$\theta(t) = \theta(0) - bt.$$

We can finally write

$$x(t) = r_0 e^{at}\cos(\theta_0 - bt)$$
$$y(t) = r_0 e^{at}\sin(\theta_0 - bt).$$

Exercise 51. *Show that the above functions can also be written in the form*

$$\begin{pmatrix}x(t)\\ y(t)\end{pmatrix} = \begin{pmatrix}\cos(bt) & \sin(bt)\\ -\sin(bt) & \cos(bt)\end{pmatrix}\begin{pmatrix}x_0\\ y_0\end{pmatrix}.$$

The next theorem contains yet another useful trick.

Theorem 52. *Suppose that for some non-negative integer k, $\mathbf{A}^{k+1} = \mathbf{0}$, where $\mathbf{A}$ is an n by n matrix and $\mathbf{0}$ is the zero matrix. Then*

$$e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{1}{2!}(\mathbf{A}t)^2 + \cdots + \frac{1}{k!}(\mathbf{A}t)^k.$$

Proof. From the equation $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ (and since $\mathbf{A}$ does not depend on t) we obtain $\ddot{\mathbf{x}} = \mathbf{A}\dot{\mathbf{x}} = \mathbf{A}^2\mathbf{x}$. Iterating this procedure gives for the lth derivative in t: $\mathbf{x}^{(l)} = \mathbf{A}^l\mathbf{x}$. As $\mathbf{A}^l$ is $\mathbf{0}$ as soon as $l = k+1$, this shows that $\mathbf{x}$ is a polynomial in t of degree at most k. The coefficients of this polynomial are the Taylor coefficients $\mathbf{x}^{(l)}(0)/l!$. Thus the theorem amounts to writing the Taylor sum of $\mathbf{x}(t)$ to order k and observing that no error term remains. □

As an example, let us calculate the exponential of

$$\mathbf{A} = \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}.$$

We write $\mathbf{A} = \lambda\mathbf{I} + \mathbf{N}$ where $\mathbf{I}$ is the identity matrix of order 4:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and

$$\mathbf{N} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

It is then easy to compute:

$$\mathbf{N}^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{N}^3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{N}^4 = \mathbf{0}.$$

It is clear that $\mathbf{N}$ commutes with $\lambda\mathbf{I}$ (since the identity matrix $\mathbf{I}$ commutes with any square matrix of the same order). Therefore

$$\begin{aligned} \exp(\mathbf{A}t) &= \exp(\lambda t\mathbf{I})\exp(\mathbf{N}t) \\ &= e^{\lambda t}(\mathbf{I} + \mathbf{N}t + \frac{1}{2}\mathbf{N}^2t^2 + \frac{1}{6}\mathbf{N}^3t^3) \\ &= e^{\lambda t}\begin{pmatrix} 1 & t & t^2/2 & t^3/6 \\ 0 & 1 & t & t^2/2 \\ 0 & 0 & 1 & t \end{pmatrix}. \end{aligned}$$

It turns out the the examples computed so far are the building blocks for exponentiating an arbitrary square matrix known as the Jordan Theorem. This is due to the following fundamental fact about matrices. Any n by n matrix can be written as a sum $\mathbf{S} + \mathbf{N}$ where $\mathbf{S}$ is diagonalizable and $\mathbf{N}$ is a *nilpotent* matrix that commutes with $\mathbf{S}$. Being nilpotent means that $\mathbf{N}^l = \mathbf{0}$ for all l greater than

some non-negative integer k. We can now obtain the exponential of the given matrix as the product of the exponentials of $\mathbf{S}$ (obtained by first writing it in the form $\mathbf{C}^{-1}\mathbf{D}\mathbf{C}$) and of $\mathbf{N}$:

$$\exp(\mathbf{S}+\mathbf{N}) = \mathbf{C}^{-1}e^{\mathbf{D}}\mathbf{C}(I+\mathbf{N}t+(\mathbf{N}t)^2/2!+\cdots+(\mathbf{N}t)^k/k!).$$

13.3. **More on Diagonalization.** We should go back to the idea of matrix diagonalization and illustrate the procedure with one example. Take again the matrix

$$\mathbf{A} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Recall that the eigenvalues and eigenvectors of $\mathbf{A}$ are obtained by solving for λ and $\mathbf{u}$ in the equation $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$, which is equivalent to $(\mathbf{A}-\lambda\mathbf{I})\mathbf{u} = 0$. A non-zero $\mathbf{u}$ satisfying the equation exists if and only if the matrix $\mathbf{A}-\lambda\mathbf{I}$ is not invertible, which is equivalent to

$$\det(\mathbf{A}-\lambda\mathbf{I}) = 0.$$

For the example, this is the equation $(\lambda-a)^2+b^2=0$, whose zeroes are $\lambda_1 = a+ib$ and $\lambda_2 = a-ib$. We can now find the eigenvectors by solving for $\mathbf{u}_1$ and $\mathbf{u}_2$ the linear systems $(\mathbf{A}-\lambda_i\mathbf{I})\mathbf{u}_i = 0$, for $i = 1,2$. Notice that $\mathbf{u}_i$ is unique only up to a multiplicative constant. We easily get

$$\mathbf{u}_1 = \begin{pmatrix} i \\ -1 \end{pmatrix}, \ \mathbf{u}_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}.$$

We can now write:

$$\mathbf{C}^{-1} = \begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix}, \ \mathbf{C} = \frac{1}{2i}\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}.$$

Therefore $\mathbf{A}$ has the factorization

$$\mathbf{A} = \frac{1}{2i}\begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix}\begin{pmatrix} a+ib & 0 \\ 0 & a-ib \end{pmatrix}\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}.$$

The exponential function therefore is

$$e^{\mathbf{A}t} = \frac{1}{2i}\begin{pmatrix} i & i \\ -1 & 1 \end{pmatrix}\begin{pmatrix} e^{(a+ib)t} & 0 \\ 0 & e^{(a-ib)t} \end{pmatrix}\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}.$$

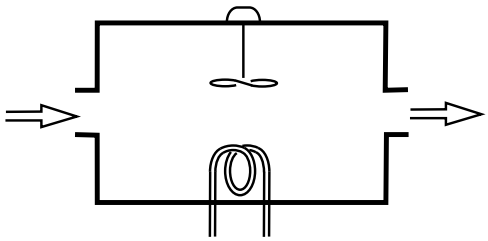
Multiplying out and using Euler's identities:

$$\cos\theta = \frac{e^{i\theta}+e^{i\theta}}{2}, \ \sin\theta = \frac{e^{i\theta}-e^{i\theta}}{2i},$$

we obtain as before

$$e^{\mathbf{A}t} = e^{at}\begin{pmatrix} \cos(bt) & \sin(bt) \\ -\sin(bt) & \cos(bt) \end{pmatrix}.$$

13.4. **The General Linear Chemical System with Mass Action Kinetics.** The figure represents an open, well-mixed system operating isothermically. Suppose that substances $X_1, X_2, \ldots, X_l$ enter the system at constant rate and a mixture of them exit at a constant rate in such a way that the net mass remains constant, while a set of linear reactions takes place inside. We would like to find the differential equations that model the process and try to find some information about the solutions. We consider the closed system as a special case in which all flow rates are 0.



Linear reactions have the following form, in which we represent the forward and back reactions separately, with reaction rates given, respectively, by $r_{ij}(\mathbf{x}) = k_{ij}x_i$ and $r_{ji}(\mathbf{x}) = k_{ji}x_j$:

$$X_i \xrightarrow{k_{ij}} X_j$$

$$X_j \xrightarrow{k_{ji}} X_i.$$

It is convenient sometimes to write the fictitious reactions:

$$O \xrightarrow{k_{0i}} X_i$$

$$X_i \xrightarrow{k_{i0}} O.$$

They are meant to represent creation and destruction of X_i in the system due to inflow and outflow. Notice that assuming the validity of mass action law for these reactions amounts to assuming that inflow of X_i takes place at a constant rate k_{0i} while the outflow of X_i happens at a rate $k_{i0}x_i$. (This makes sense for a well-stirred reactor. In this case, k_{i0} has the same value for all i.)

The associated system of differential equations is given by

$$\dot{\mathbf{x}} = \sum_{i=1}^{l}\sum_{j=1}^{l} k_{ij}x_i(X_j - X_i) + \sum_{i=1}^{l}(k_{0i} - k_{i0}x_i)X_i.$$

If the system is closed, the second term on the right-hand side of the equation is 0. In the general case, the first term can be written as

$$\sum_{i<j}(k_{ij}x_i - k_{ji}x_j)(X_j - X_i).$$

A little manipulation shows that the differential equation has the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{k},$$

where

$$\mathbf{A} = \begin{pmatrix} -\sum_{i\neq 1} k_{1i} & k_{21} & k_{31} & \cdots & k_{l1} \\ k_{12} & -\sum_{i\neq 2} k_{2i} & k_{32} & \cdots & k_{l2} \\ k_{13} & k_{23} & -\sum_{i\neq 3} k_{3i} & \cdots & k_{l3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ k_{1l} & k_{2l} & k_{3l} & \cdots & -\sum_{i\neq l} k_{li} \end{pmatrix}$$

(the sums run from 0 to l skipping over the indicated index) and $\mathbf{k}$ is the column vector with components $(k_{01}, k_{02}, \cdots, k_{0l})$.

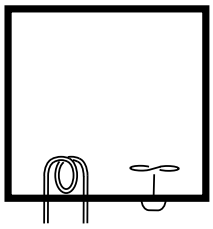
For the next exercise, keep in mind that $\mathbf{A}$ and its exponential commute and that the derivative of $e^{\mathbf{A}t}$ in t is $\mathbf{A}e^{\mathbf{A}t}$.

Exercise 53. *Show that the differential equation* $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{k}$, *with initial value* $\mathbf{x}(0) = \mathbf{x}_0$, *has the solution*

$$\mathbf{x}(t) = e^{t\mathbf{A}}\mathbf{x}_0 + \mathbf{A}^{-1}(e^{t\mathbf{A}} - \mathbf{k})\mathbf{k}.$$

This solution is valid even if $\mathbf{A}^{-1}$ *is not invertible. How should we make sense of it then? (Interpret the exponentials as power series in* $\mathbf{A}$*.)*

If the system is closed (this is represented in the next figure) $\mathbf{k} = 0$ and the sums on the diagonal entries begin with index 1.



isothermal, well-stirred, constant volume closed reactor

Mass conservation forces the values of concentrations to remain bounded (i.e. they do not grow to infinity). This can be seen as follows. Let μ_i denote the mass per mole of substance X_i and V the volume of the container, which we assume constant. The reactions are all of the form $X_i \rightharpoonup X_j$, so unless $k_{ij} = 0$ we are forced to assume that $\mu_i = \mu_j$. (For simplicity it will be supposed that all k_{ij} are non-zero but possibly very small, although this is not an essential assumption.) Therefore we can drop the sub-index from the μ_i The total mass inside the container can now be written

$$M = \mu V \sum_{i=1}^{l} x_i.$$

In particular the sum of the molar concentrations is constant:

$$\sum_{i=1}^{l} x_i(t) = \sum_{i=1}^{l} x_i(0) = C.$$

We can now define a compact set (i.e., a closed and bounded subset of $\mathbb{R}^l$) where the differential equation $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{k}$ is defined by:

$$\mathcal{R} = \left\{ \mathbf{x} \in \mathbb{R}^l : x_i \geq 0 \text{ for all i, and } \sum_i x_i = C \right\}.$$

In the remaining of this section we will show the following result:

Proposition 54. *The set* $\mathcal{R}$ *is forward invariant by the flow of* $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{k}$ *and contains a critical point of the differential equation, i.e., a point* $\mathbf{x}^*$ *such that* $\mathbf{A}\mathbf{x}^* + \mathbf{k} = \mathbf{0}$. *In particular, the associated flow has a fixed point. If all* k_{0i} *are different from* 0, *then there is a unique critical point,* $\mathbf{x}^*$, *which lies in the interior of* $\mathbf{R}$ *(i.e., not on the boundary of* $\mathcal{R}$*.) This fixed point is stable in the following sense:* $\lim_{t\to\infty} \Phi_t(\mathbf{x}) = \mathbf{x}^*$, *for every* $\mathbf{x} \in \mathcal{R}$.

Proof. That $\mathcal{R}$ is invariant under the flow for positive times and the existence of fixed points (or critical points of the vector field) will be shown in the next subsection under more general assumptions. Let us show that if the k_{0i} are positive, then the critical point is unique and lies in the interior of $\mathcal{R}$. A boundary point $\mathbf{z}$ has the property that at least one of its coordinates, say z_i, is 0. We have

$z_j \geq 0$ for all j and for each $j \neq i$ we have $a_{ij} = k_{ji} \geq 0$. But for a critical point $\mathbf{z}^*$ the equation $\mathbf{A}\mathbf{z}^* = -\mathbf{k}$ holds. If $\mathbf{z}^*$ were to be found on the boundary, the assumption that $z_i = 0$ would imply $0 \leq \sum_{j\neq i} a_{ij} z_j = -k_i < 0$. But this is a contradiction.

Let us check now that the critical point is unique. If we had two, say $\mathbf{z}_1$ and $\mathbf{z}_2$, then for each s, the point $\mathbf{z}(s) = s\mathbf{z}_1 + (1-s)\mathbf{z}_2$ would also be a critical point since

$$\begin{aligned}\mathbf{A}\mathbf{z}(s) + \mathbf{k} &= s\mathbf{A}\mathbf{z}_1 + (1-s)\mathbf{A}\mathbf{z}_2 + \mathbf{k} \\ &= -s\mathbf{k} - (1-s)\mathbf{k} + \mathbf{k} \\ &= 0.\end{aligned}$$

But $\mathbf{z}(s)$ is a parametric line that must intersect the boundary of $\mathcal{R}$ for some s. This would give a critical point at the boundary, which cannot happen.

Finally, we need to check stability. Write $\mathbf{u} - \mathbf{x} - \mathbf{z}^*$, where $\mathbf{z}^*$ is the critical point. Then the differential equation for $\mathbf{x}$ gives $\dot{\mathbf{u}} = \mathbf{A}\mathbf{u}$. In fact,

$$\begin{aligned}\mathbf{A}\mathbf{u} &= \mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{z}^* \\ &= (\mathbf{A}\mathbf{x} + \mathbf{k}) - (\mathbf{A}\mathbf{z}^* + \mathbf{k}) \\ &= \mathbf{A}\mathbf{x} + \mathbf{k} \\ &= \dot{\mathbf{x}} \\ &= \dot{\mathbf{u}},\end{aligned}$$

where it was used that $\mathbf{A}\mathbf{z}^* + \mathbf{k} = 0$, and $\dot{\mathbf{z}}^* = 0$. Therefore we want to show that for arbitrary $\mathbf{u}_0$ the point $e^{t\mathbf{A}}\mathbf{u}$ approaches 0 as $t \to \infty$. It suffices for this to show that all the eigenvalues of $\mathbf{A}$ have negative real part, which is our final claim.

Let $\lambda = a + ib$ be an eigenvalue of $\mathbf{A}$. If a were positive there would be a non-zero $\mathbf{u}_0$ for which $\|e^{t\mathbf{A}}\mathbf{u} + 0\|$ would grow exponentially at a rate e^{at}. This is not possible since $\mathcal{R} - \mathbf{z}_*$ is a bounded region. Suppose that $a = 0$. If $b = 0$ there would be a nonzero $\mathbf{u}_0$ such that $\mathbf{A}\mathbf{u}_0 = \mathbf{0}$, but this is not possible since $\mathbf{z}^* + \mathbf{u}_*$ would then be a second critical point for the original vector field. It remains to consider the case of a purely imaginary nonzero λ.

In this case we have a two dimensional simplex V contained in $\mathcal{R} - \mathbf{z}^*$ invariant under $e^{t\mathbf{A}}$ whose points rotate about $\mathbf{0}$ with angular rate b. But this is not possible unless $b = 0$, since the vertices of the simplex are a longer distance from $\mathbf{0}$ than their neighboring points and must be fixed points of the rotation. □

Let us consider again the isomerization reaction example (the umbrellas), now with the extra assumption that a point of detailed balance exists. Therefore, there exists $\mathbf{x}^* = (x_1^*, x_2^*, x_3^*)$ such that $K_{ij} = x_i^*/x_j^*$. We suppose that this is an interior point, that is, $x_i^* > 0$ for all i.

Exercise 55. *Consider the umbrella example. Suppose that $\mathbf{x}^*$ is a point of detailed equilibrium and define the change of coordinates: $\mathbf{y} = \mathbf{D}(\mathbf{x} - \mathbf{x}^*)$, where*

$$\mathbf{D} = \begin{pmatrix} \frac{1}{\sqrt{x_1^*}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{x_2^*}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{x_3^*}} \end{pmatrix}.$$

Show that in the new coordinates the equation reads $\dot{\mathbf{y}} = \mathbf{DAD}^{-1}\mathbf{y}$, *where*

$$\mathbf{DAD}^{-1} = \begin{pmatrix} -(k_1^+ + k_3^-) & \sqrt{k_1^- k_1^+} & \sqrt{k_3^- k_3^+} \\ \sqrt{k_1^- k_1^+} & -(k_2^+ + k_1^-) & \sqrt{k_2^- k_2^+} \\ \sqrt{k_3^- k_3^+} & \sqrt{k_2^- k_2^+} & -(k_3^+ + k_2^-) \end{pmatrix}.$$

(Notice that this is a symmetric matrix, that is, it is equal to its transpose. Symmetric matrices are always diagonalizable with real eigenvalues.)

From what was discussed above, the eigenvalues of $\mathbf{DAD}^{-1}$ must have negative real part. It turns out that those eigenvalues are all real, as remarked at the bottom of the exercise. Here is an exercise where you can check that this is true.

Exercise 56. *Show that if* $\mathbf{B}$ *is a symmetric n by n matrix, then* $\mathbf{B}$ *is diagonalizable, its eigenvalues are all real, and its eigenvectors can be chosen to form an orhtonormal basis. Show this using the following argument:*

(1) *Let S denote the set of all* $\mathbf{x} \in \mathbb{R}^n$ *of norm* 1 *and define on S the function* $g(\mathbf{x}) = (\mathbf{B}x) \cdot \mathbf{x}$. *Explain: there exists a point* $\mathbf{x}_1 \in S$ *where g attains a maximum (or minimum) value.*
(2) *Let* $\mathbf{x}_1$ *be a point of maximum for g on S. (In particular,* $\mathbf{x}_1$ *is a critical point.) Explain why* $\mathbf{x}_1$ *must be an eigenvector of* $\mathbf{B}$. *(Hint: let* $\mathbf{u}$ *be any unit vector perpendicular to* $\mathbf{x}_1$, *so that* $c(t) = (\cos t)\mathbf{x}_1 + (\sin t)\mathbf{u}$ *is a parametric curve that passes through* $\mathbf{x}_1$ *for* $t = 0$ *and is tangent to S, with tangent vector* $\mathbf{u}$. *Show that* $(g \circ c)'(0) = 0$ *due to the fact that* $\mathbf{x}_1$ *is a point of maximum. Calculate the derivative using the chain rule and show that it gives* $\mathbf{B}\mathbf{x}_1 \cdot \mathbf{u} = 0$. *Since* $\mathbf{u}$ *is arbitrary among vectors that are perpendicular to* $\mathbf{x}_1$ *it must follow that* $\mathbf{B}\mathbf{x}_1$ *is parallel to* $\mathbf{x}_1$.
(3) *Show that the linear subspace* $\mathbf{x}_1^\perp$ *of* $\mathbb{R}^n$ *of diimension* $n-1$ *consisting of all vectors that are perpendicular to* $\mathbf{x}_1$ *has the property that* $\mathbf{B}\mathbf{x}_1^\perp \subset \mathbf{x}_1^\perp$.
(4) *Now consider the set* S' *of unit vectors in this perpendicular subspace and the restriction of g to* S'. *By the same argument used above show that* S' *contains an eigenvector, call it* $\mathbf{x}_2$, *of* $\mathbf{B}$ *with real eigenvalue. By construction* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *are perpendicular.*
(5) *Now use induction to finish the exercise.*

13.5. **A General Stability Result.** We discuss a few stability results for isotermal homogeneous reactions taking place in a constant volume reactor. Suppose that a chemical mechanism consists of k reactions with reaction vectors $\mathbf{R}_s = \mathbf{b}_s - \mathbf{a}_s$, and mass action rate law, $r_s(\mathbf{x}) = k_s \mathbf{x}^{\mathbf{a}_s}$, for $s = 1, \cdots, k$. (The components of $\mathbf{a}_s$ or $\mathbf{b}_s$ will be written $\mathbf{a}_s(i)$ and $\mathbf{b}_s(i)$.) If we make no explicit mention to the contrary, reactions of the type $O \rightharpoonup X_i$ and $X_i \rightharpoonup O$ are also allowed. (Clearly, these reactions are not present for closed systems.) Therefore the general reaction we study here is of the form

$$\dot{\mathbf{x}} = \sum_{s=1}^{k} k_s \mathbf{x}^{\mathbf{a}_s} (\mathbf{b}_s - \mathbf{a}_s) + \sum_{i=1}^{l} (k_{0i} - k_{i0} x_i) X_i.$$

We do assume, however, that the inflow and outflow of mass exactly match, so that the total mass inside the reactor is a constant, M. Therefore $\sum_{i=1}^{l} \mu_i x_i V = M$, where μ_i is the weight of X_i per mole.

Define the set

$$\mathcal{R} = \left\{ \mathbf{x} \in \mathbb{R}^l : x_i \geq 0 \text{ for all i, and } \sum_{i=1}^{l} \mu_i x_i = M/V \right\}.$$

For language convenience we call differential equations such as the one obove (and their flow) the chemical equation (and respective chemical flow).

Proposition 57. *The set $\mathcal{R}$ is forward invariant under the chemical flow and contains a critical point for the chemical equation.*

Proof. The key point is that **R** is flow invariant. From this the existence of a critical point is a consequence of a general result in topology, the Brower fixed point theorem, which is explained below. (Since Brower's theorem is about fixed points rather than critical points, the following should be noted: for a vector field which does not vanish over a compact set, there is $t > 0$ sufficiently small such that Φ_t does not have a fixed point on that compact set.)

Invariance of $\mathcal{R}$ in turn is a consequence of the following observation: on the subset of $\mathcal{R}$ having $x_i = 0$ for some i, $\mathbf{x}^{\mathbf{a}_s} = 0$ whenever $\mathbf{a}_s(i)$ is not zero. Therefore the differential equation gives

$$\dot{x}_i = \sum_{s=1}^{k} k_s \mathbf{x}^{\mathbf{a}_s} \mathbf{b}_s(i) + k_{0i} \geq 0.$$

This means that at the part of the boundary where $x_i = 0$, this component of **x** cannot decrease further. □

A subset of $\mathbb{R}^n$ is said to be *convex* if given any **x** and **y** in the set, the line segment from **x** to **y** is contained in the set. Another way of saying this is that for any number λ such that $0 \leq \lambda \leq 1$, the point $(1-\lambda)\mathbf{x} + \lambda\mathbf{y}$ is in the set.

Here is a statement of Brower's fixed point theorem in a somewhat specialized form. (Convexity can be replaced with a more general assumption, but the theorem is sufficient for us in this form.)

Theorem 58 (Brower Fixed Point Theorem)**.** *Let $\mathcal{R}$ be a compact and convex subset of $\mathbf{R}^l$ and $\Phi : \mathcal{R} \to \mathcal{R}$ a continuous map. Then there exists a point $\mathbf{x}^* \in \mathcal{R}$ such that $\Phi(\mathbf{x}^*) = \mathbf{x}^*$.*

From here on till the end of the section we assume that the chemical equation admits a critical point which is a point of detailed balance and that the system is closed. Note the following: if we assume that a critical point $\mathbf{x}^*$ is a point of detailed equilibrium, then

$$k_{ij} x_i^* = k_{ji} x_j^*$$

for all ij. Let us suppose that these constants are all non-zero, and define $K_{ij} = k_{ij}/k_{ji}$. It follows from $K_{ij} = x_i^*/x_j^*$ that for all i, j, l,

$$K_{il} = K_{ij} K_{jl}.$$

Conversely, it is not difficult to see that if this condition holds for all i, j, k, then there exists a point $\mathbf{x}^*$ of detailed equilibrium.

We suppose that for any elementary reaction present in the reaction mechanism, the back reaction is also present. The reaction constants will be written either as $k_{\pm s}$ or $k_s^{\pm}$.

Proposition 59. *Suppose the system is closed, isothermal, and constant volume, and that the reaction rates are of mass action type. Also suppose that there exists a positive (i.e., not on the boundary of $\mathcal{R}$) point of detailed equilibrium, $\mathbf{x}^*$. Let Φ_t denote the chemical flow. Then, for any $\mathbf{x} \in \mathcal{R}$, we have that $\Phi_t(\mathbf{x})$ approaches asymptotically (as $t \to \infty$) the set of detailed equilibrium points.*

Proof. Define a function

$$G(\mathbf{x}) = \sum_{i=1}^{l} x_i \left(\ln \frac{x_i}{x_i^*} \right).$$

If $\mathbf{x}^*$ is a point of detailed equilibrium, then for each s,

$$\frac{k_s^+}{k_s^-} = \prod_{i=1}^{l} (x_i^*)^{\mathbf{b}_s(i) - \mathbf{a}_s^s(i)}.$$

We now calculate the time derivative of $G(\mathbf{x}(t))$.

$$\begin{aligned}
\frac{d}{dt}G(\mathbf{x}(t)) &= \sum_{i=1}^{l}\left[\dot{x}_i\left(\ln\frac{x_i}{x_i^*}\right)+\dot{x}_i\right] \\
&= \sum_{i=1}^{l}\dot{x}_i\ln\frac{x_i}{x_i^*} \\
&= \sum_{i=1}^{l}\left(\sum_s k_s\mathbf{x}^{\mathbf{a}_s}(\mathbf{b}_s(i)-\mathbf{a}_s(i))\ln\frac{x_i}{x_i^*}\right) \\
&= \sum_s k_s\mathbf{x}^{\mathbf{a}_s}\ln\frac{\prod_i x_i^{\mathbf{b}_s(i)-\mathbf{a}_s(i)}}{\prod_i x_i^{*\mathbf{b}_s(i)-\mathbf{a}_s(i)}} \\
&= \sum_s k_s\mathbf{x}^{\mathbf{a}_s}\ln\frac{\prod_i x_i^{\mathbf{b}_s(i)-\mathbf{a}_s(i)}}{\prod_i k_s^+/k_s^-} \\
&= \sum_s k_s\mathbf{x}^{\mathbf{a}_s}\ln\frac{k_s^-\prod_i x_i^{\mathbf{b}_s(i)}}{k_s^+\prod_i x_i^{\mathbf{a}_s(i)}} \\
&= \sum_{s=\pm 1}^{\pm k} r_s(\mathbf{x})\ln\frac{r_{-s}(\mathbf{x})}{r_s(\mathbf{x})} \\
&= -\sum_{s=1}^{k}(r_s(\mathbf{x})-r_{-s}(\mathbf{x}))\ln\frac{r_{-s}(\mathbf{x})}{r_s(\mathbf{x})} \\
&= -\sum_{s=1}^{k}(r_s(\mathbf{x})-r_{-s}(\mathbf{x}))\left(\ln r_{-s}(\mathbf{x})\ln r_s(\mathbf{x})\right).
\end{aligned}$$

But the expression $(r_s(\mathbf{x})-r_{-s}(\mathbf{x}))\left(\ln r_{-s}(\mathbf{x})-\ln r_s(\mathbf{x})\right)$ is non-negative since ln is a monotone function, and it is zero if and only if $r_s(\mathbf{x})=r_{-s}(\mathbf{x})$. Therefore $G(\mathbf{x}(t))$ is a decreasing function, and the derivative is zero exactly on the points of detailed equilibrium. □

We would like now to understand what happens in a neighborhood of an isolated point of detailed equilibrium. We write the reaction vector as $\mathbf{R}_s=\mathbf{b}_s-\mathbf{a}_s$, and $\mathbf{b}_{-s}=\mathbf{a}_s$. Let $\mathbf{z}$ be such a point. The equilibrium condition $k_s^+\mathbf{z}^{\mathbf{a}_s}=k_s^-\mathbf{z}^{\mathbf{b}_s}$ is then written as $k_s\mathbf{z}^{\mathbf{a}_s}=k_{-s}\mathbf{z}^{\mathbf{a}_{-s}}$, for all s. Now observe the following (we write below $\mathbf{u}=\mathbf{x}-\mathbf{z}$):

$$\begin{aligned}
\frac{d\mathbf{u}}{dt} &= \frac{d\mathbf{x}}{dt} \\
&= \sum_s k_s\mathbf{x}^{\mathbf{a}}\mathbf{R}_s \\
&= \sum_s k_s(\mathbf{x}^{\mathbf{a}}-\mathbf{z}^{\mathbf{a}})\mathbf{R}_s \\
&= \sum_s (r_s(\mathbf{z}+\mathbf{u})-r_s(\mathbf{z}))\mathbf{R}_s \\
&= \sum_s (\nabla r_s)_{\mathbf{z}}\cdot\mathbf{u}\mathbf{R}_s + o(\|\mathbf{u}\|).
\end{aligned}$$

Also observe that

$$\frac{\partial r_s}{\partial x_i}(z) = \frac{\mathbf{a}_s(i)}{z_i}r_s(\mathbf{z}).$$

It will be convenient to use the notation:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{l} \frac{u_i v_i}{z_i}.$$

Then it follows that

$$\frac{d\mathbf{u}}{dt} = \sum_s \langle \mathbf{a}_s, \mathbf{u} \rangle r_s(\mathbf{z}) \mathbf{R}_s.$$

Since $r_s(\mathbf{z}) = r_{-s}(\mathbf{z})$, we have:

$$\begin{aligned}
\frac{d\mathbf{u}}{dt} &= \sum_{s=\pm 1}^{\pm k} \langle \mathbf{a}_s, \mathbf{u} \rangle r_s(\mathbf{z}) \mathbf{R}_s + o(\|\mathbf{u}\|) \\
&= \sum_{s=1}^{k} \left(\langle \mathbf{a}_s, \mathbf{u} \rangle r_s(\mathbf{z}) \mathbf{R}_s - \langle \mathbf{b}_s, \mathbf{u} \rangle r_s(\mathbf{z}) \mathbf{R}_s \right) + o(\|\mathbf{u}\|) \\
&= -\sum_{s=1}^{k} \langle \mathbf{R}_s, \mathbf{u} \rangle r_s(\mathbf{z}) \mathbf{R}_s + o(\|\mathbf{u}\|) \\
&= \mathbf{A}\mathbf{u} + o(\|\mathbf{u}\|).
\end{aligned}$$

The linear map $\mathbf{A}$ defined by the last line of the above chain of equalities has the property that

$$\langle \mathbf{A}\mathbf{u}, \mathbf{v} \rangle = -\sum_{s=1}^{k} \langle \mathbf{R}_s, \mathbf{u} \rangle \langle \mathbf{R}_s, \mathbf{v} \rangle r_s(\mathbf{z}).$$

Proposition 60. *The linear transformation* $\mathbf{A}$ *is diagonalizable and its eigenvalues are all non-positive.*

The fact that $\mathbf{A}$ is diagonalizable with real eignevalues follow from the same argument used in exercise 56, except that the dot product $\mathbf{u} \cdot \mathbf{v}$ is replaced with the inner product $\langle \mathbf{u}, \mathbf{v} \rangle$. The eigenvectors are orthonormal for this new inner product. To see that the eigenvalues are all negative, notice that if $\mathbf{A}\mathbf{u} = \lambda \mathbf{u}$ then, from the definition of $\mathbf{A}$ we get:

$$\begin{aligned}
\lambda \langle \mathbf{u}, \mathbf{u} \rangle &= \langle \mathbf{A}\mathbf{u}, \mathbf{u} \rangle \\
&= -\sum_{s=1}^{k} \langle \mathbf{R}_s, \mathbf{u} \rangle^2 r_s(\mathbf{z}) \\
&\leq 0.
\end{aligned}$$

Therefore $\lambda \leq 0$.

14. Chemical Oscillations

14.1. **The Poincaré-Bendixson Theorem.** This is a criterion to establish the existence of closed orbits for two-dimensional systems.

Theorem 61. *Suppose that* $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ *is a continuously differentiable vector field on an open set in* $\mathbb{R}^2$ *containing a compact set R and that the following hypothesis hold:*

(1) *R does not contain any fixed points; and*
(2) *there exists a trajectory C that is "confined" in R, that is, that stays in R for all time* $t \geq 0$.

Then either C is a closed orbit, or it spirals toward a closed orbit as $t \to \infty$. *In either case, R contains a closed orbit.*

In view of our earlier remarks about the existence of fixed points, it may appear that hypothesis 1 is too restrictive. However, if all fixed points are unstable (expanding), we can apply the theorem in the way explained in the next example.

Consider the set of reactions

$$X \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} A$$

$$B \overset{k_2}{\rightarrow} Y$$

$$2X + Y \overset{k_3}{\rightarrow} 3X$$

We suppose that A and B are substances whose amounts can be controlled throughout the process and kept constant. We wish to follow the concentrations of X and Y, which we denote by x and y. The (constant) concentrations of A and B are a and b, respectively. Assuming mass action law we arrive at the equations:

$$\dot{x} = k_{-1}a - k_1 x + k_3 x^2 y$$
$$\dot{y} = k_2 b - k_3 x^2 y.$$

Before studying this system it is convenient to change coordinates so as to make the quantities involved dimensionless. Here is a suggested change:

$$\tau = k_1 t$$
$$u = \sqrt{\frac{k_3}{k_1}}x$$
$$v = \sqrt{\frac{k_3}{k_1}}y.$$

Exercise 62. *Show that under the above coordinate changes the system takes the form:*

$$\frac{du}{d\tau} = \alpha - u + u^2 v$$
$$\frac{dv}{d\tau} = \beta - u^2 v,$$

where $\alpha = k_3^{1/2} k_1^{-3/2} k_{-1} a$ *and* $\beta = k_3^{1/2} k_1^{-3/2} k_2 b$.

Therefore, up to a change of coordinates, the original system is the same as the system

$$\dot{x} = a - x + x^2 y$$
$$\dot{y} = b - x^2 y,$$

where a and b are (new) positive constants. We study the system in this form.

Exercise 63. *Show that the vector field* $\mathbf{f}(\mathbf{x}) = (a - x + x^2 y, b - x^2 y)$ *has a single critical point,* $\mathbf{x}^* = (a+b, b/(a+b)^2)$*, and that the jacobian matrix of* $\mathbf{f}(\mathbf{x})$ *at* $\mathbf{x}^*$ *is given by*

$$\begin{pmatrix} -(a-b)/(a+b) & (a+b)^2 \\ -2b/b + a & -(a+b)^2 \end{pmatrix}.$$

Show that the determinant of this matrix is $D = (a+b)^2$*. Argue that the critical point is a repeller if the trace of this matrix is positive, and a sink if negative. (Recall that the eigenvalues satisfy the quadratic equation* $\lambda^2 - T\lambda + D$*, where* T *is the trace.)*

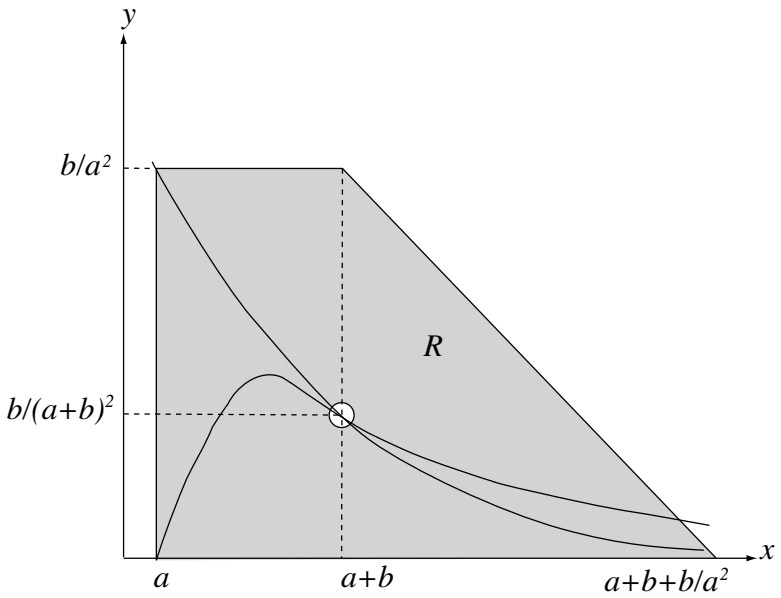
Therefore the condition for the critical point to be a repeller is

$$T = \frac{b-a}{b+a} - (a+b)^2 > 0.$$

The vector field defined by this system is continuously differentiable for all (x,y) in $\mathbb{R}^2$. Therefore in order to obtain invariant limit cycles (if they exist) we need to show that there exists a compact subset $R \subset \mathbb{R}^2$ satisfying conditions 1 and 2 of the Poincaré-Bendixson theorem. (We are interested in such cycles only on the positive quadrant if x and y are to be interpreted as concentrations.)

Exercise 64. *This exercise refers to the next figure.*

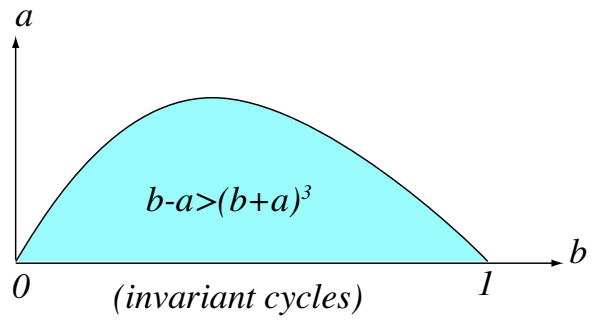
(1) *Verify that the intersecting curves correctly represent the null clines for $\dot{x}$ and $\dot{y}$, and that the point of intersection is the unique critical point of the given vector field.*
(2) *Show that the polygonal region consisting of the union of the shaded region R and the small open disc around the critical point (which was excised from the polygoan in the drawing) is a compact trapping region for the system.*
(3) *If the critical point is a repeller, show that by excising a small enough disc, as in the figure, the resulting compact region R is still trapping (and free of fixed points for the flow associated to the vector field).*
(4) *Conclude that if $T > 0$, R contains at least one invariant cycle. Furthermore, any invariant cycles must surround the deleted disc.*



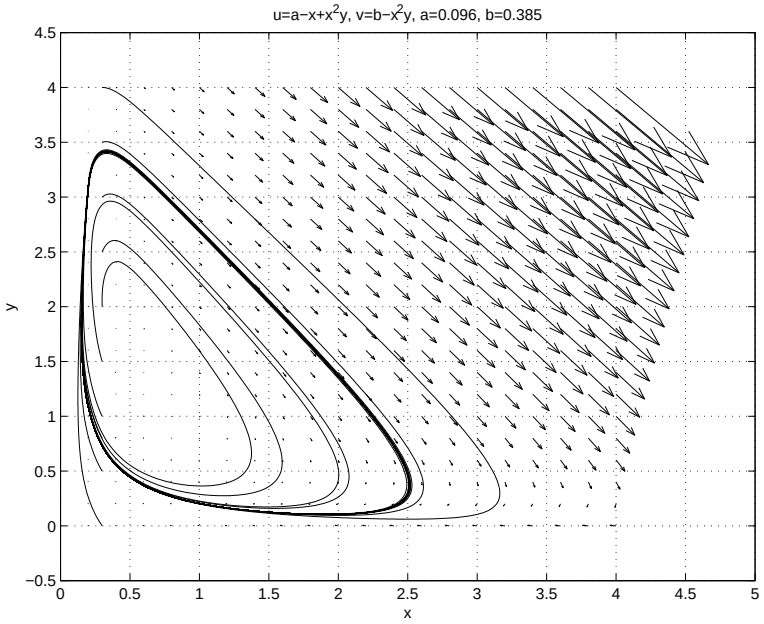
Exercise 65. *Show that the set of values of (a,b) for which there exists an invariant cycle is as represented in the next figure, where the graph of the function $a = h(b)$ is given in parametric form by the curve*

$$b = \frac{s+s^3}{2}, \quad a = \frac{s-s^3}{2},$$

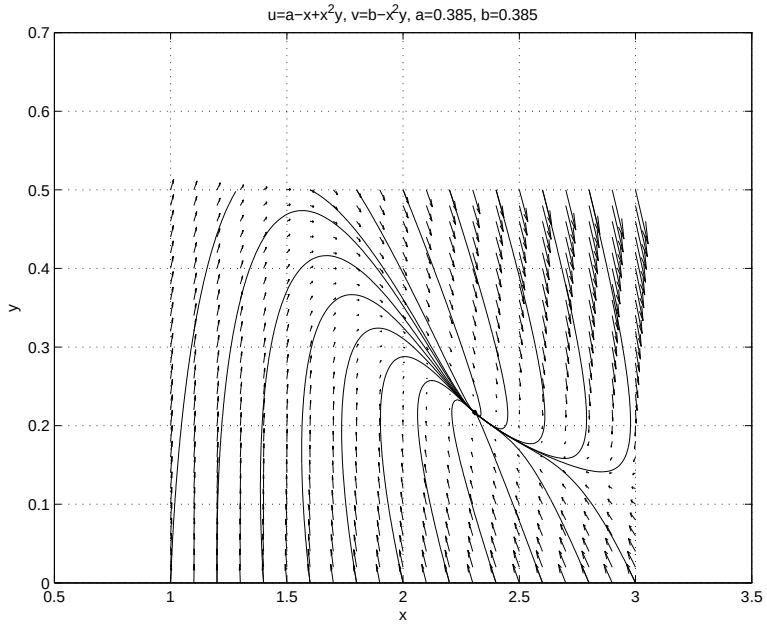
for $0 \leq s \leq 1$.



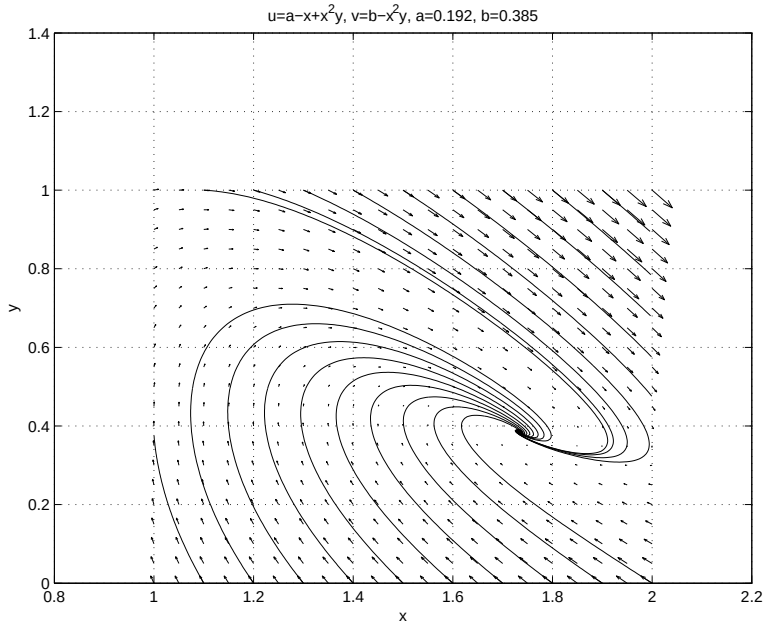
The next figure shows a number of trajectories for the system with $a = 1/6\sqrt{3}$, $b = 2/3\sqrt{3}$. In this case (b,a) lies in the shaded region, of the previous figure.



Next, we choose $a = 2/3\sqrt{3}$, $b = 2/3\sqrt{3}$. In this case (b,a) lies outside the shaded region.



Next, we choose $a = 1/3\sqrt{3}$, $b = 2/3\sqrt{3}$. Now (b,a) lies on the boundary of the shaded region. The eigenvalues of the Jacobian matrix of the vector field are purely imaginary, so that the linear approximation has a center at the critical point.



The same behavior observed in the previous example can be seen in a simpler and more explicit example. Consider the system:

$$\dot{x} = by - x(x^2 + y^2 - a)$$
$$\dot{y} = -bx - y(x^2 + y^2 - a).$$

This is defined for all $(x,y) \in \mathbb{R}^2$.

We will study this system in the next few exercises.

Exercise 66. *Show that the above system can be expressed in polar coordinates as*

$$\dot{r} = -r(r^2 - a)$$
$$\dot{\theta} = -b.$$

These are two independent differential equations, one for r and the other for θ. The equation for θ is easily solved: $\theta(t) = \theta_0 - bt$. Rather than solve the second equation explicitly (which can be done) let us try to understand its behavior qualitatively.

Exercise 67. *Describe qualitatively all possible orbits of $\dot{r} = -r(r^2 - a^2)$, $r \geq 0$, for different values of a. (There are two cases to distinguish: $a = 0$ and $a \neq 0$.) What are the critical points for r? are they stable? unstable? What does this say about the properties of orbits of the two dimensional system? Is there a limit cycle? What happens to trajectories of points not on the limit cycle as $t \to \infty$?*

Exercise 68. *Solve explicitly the linear system system:*

$$\dot{r} = ar$$
$$\dot{\theta} = -b.$$

Notice that this is the linearization of the previous system. Compare the various solutions (for different values of a, b and different initial conditions) with the qualitative solutions of non-linear system.

If $b \neq 0$, then as a varies from a negative value to a positive value a stable critical point (at the origin) turns into an unstable point and a limit cycle is created. This phenomenon is known as a *Hopf bifurcation*.

14.2. **The Belousov-Zhabotinskii Reaction.** See handout.

14.3. **Coupled Non-linear Oscillators.** Suppose that a two-dimensional process is governed by the system of equations which we write in vector form as $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$. Let a new system consist of two identical but independent (uncoupled), systems governed by the same vector field $\mathbf{f}(\mathbf{x})$. Thus a state of the system consists of a vector $(\mathbf{x}, \mathbf{y})$ in $\mathbb{R}^4$ and the process is governed by the system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$$
$$\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}).$$

We wish to create a weak coupling between these two independent systems. This will be done by adding a linear cross-term, as follows:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) - \varepsilon(\mathbf{x} - \mathbf{y})$$
$$\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) + \varepsilon(\mathbf{x} - \mathbf{y}),$$

where ε is a small number.

In this section we wish to remark on a property of the coupled system known as *phase locking*, which can take place when the original two-dimensional system admits limit cycles. This is about the spontaneous synchronization of weakly linked non-linear oscillators.

For concreteness we suppose that the two-dimensional system is similar to the one defined in exercise 68. We write it here in the form:

$$\dot{x} = by - x\left((x^2+y^2)^{k/2} - a\right)$$
$$\dot{y} = -bx - y\left((x^2+y^2)^{k/2} - a\right),$$

In vector form this corresponds to the vector field $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \|\mathbf{x}\|^k\mathbf{x}$ where $\mathbf{A} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$.

Exercise 69. *Let* $\mathbf{x} = (r_1\cos\theta_1, r_1\sin\theta_1)$ *and* $\mathbf{y} = (r_2\cos\theta_2, r_2\sin\theta_2)$ *Show that the system of coupled equations, with linear coupling and the vector field* $\mathbf{f}$ *just defined, can be written in polar coordinates as follows:*

$$\dot{r}_1 = (a-\varepsilon)r_1 - r_1^{k+1} + \varepsilon r_2\cos(\theta_2-\theta_1)$$
$$\dot{r}_2 = (a-\varepsilon)r_2 - r_2^{k+1} + \varepsilon r_1\cos(\theta_2-\theta_1)$$
$$r_1\dot{\theta}_1 = br_1 + \varepsilon r_2\sin(\theta_2-\theta_1)$$
$$r_2\dot{\theta}_2 = br_2 - \varepsilon r_1\sin(\theta_2-\theta_1).$$

We say that the two oscillators become synchronized if the difference $\theta_2-\theta_1$ approach 0 asymptotically as t grows. More generally, we say that phase locking takes place if this difference becomes constant asymptotically. To study synchronization in the above system, let us first write a differential equation for $\psi = \theta_2 - \theta_1$. From the equations

$$\dot{\theta}_1 = b + \varepsilon\frac{r_2}{r_1}\sin(\theta_2-\theta_1)$$
$$\dot{\theta}_2 = b - \varepsilon\frac{r_1}{r_2}\sin(\theta_2-\theta_1)$$

we obtain

$$\dot{\psi} = -\varepsilon\left(\frac{r_2}{r_1} + \frac{r_1}{r_2}\right)\sin\psi.$$

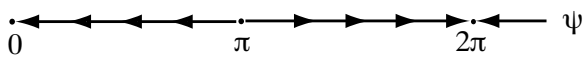
We can understand the qualitative behavior of this equation without solving it explicitly. First note that for small ε (and large t) we should expect r_1 and r_2 to be close to the critical radius of each uncoupled oscillator. This value is $r = a^{1/k}$. Thus the function

$$K(t) = \left(\frac{r_2}{r_1} + \frac{r_1}{r_2}\right),$$

which in principle requires solving the differential equations for $r_1, r_2, \theta_1, \theta_2$ to be determined explicitly, should be close to a constant (in fact, 2). The important point is that for small enough ε and large enough t we can suppose that $K(t)$ is positive.

Therefore

$$\begin{cases} \dot{\psi} < 0 & \text{if } 0 < \psi < \pi \\ \dot{\psi} > 0 & \text{if } \pi < \psi < 2\pi. \end{cases}$$



Notice that ψ naturally evolves in t so as to approach 0 ($= 2\pi$).

14.4. **Phase-locking.** Let us suppose now that the two coupled systems are not quite identical, but differ by a small change in the constans a and b. Thus the new equations become:

$$\begin{aligned}
\dot{r}_1 &= (a_1-\varepsilon)r_1 - r_1^{k+1} + \varepsilon r_2\cos(\theta_2-\theta_1)\\
\dot{r}_2 &= (a_2-\varepsilon)r_2 - r_2^{k+1} + \varepsilon r_1\cos(\theta_2-\theta_1)\\
\dot{\theta}_1 &= b_1 + \varepsilon\frac{r_2}{r_1}\sin(\theta_2-\theta_1)\\
\dot{\theta}_2 &= b_2 - \varepsilon\frac{r_1}{r_2}\sin(\theta_2-\theta_1).
\end{aligned}$$

Taking the difference of the last two equations for the time derivative of $\psi = \theta_2 - \theta_1$:

$$\dot{\psi} = b_2 - b_1 - \varepsilon\left(\frac{r_1}{r_2}+\frac{r_2}{r_1}\right)\sin\psi. \tag{1}$$

Although the system cannot be expected to have a critical point with $\dot{\theta}_i = 0$, it is often observed that that are stable solutions for which ψ is constant. When that happens we say that *phase-locking* occurs. We wish now to understand how the above system can display such type of behavior.

Exercise 70. *Suppose that a solution $(r_(t),\theta_1(t),r_2(t),\theta_2(t))$ of the above system is such that $\psi(t) = \theta_2(t)-\theta_1(t)$ is constant (independent of t). Show that r_1 and r_2 are also constant.*

The claim of the above exercise can be seen as follows: from equation 1 it follows that $\rho+\rho^{-1}$ is constant, where $\rho = r_2/r_1$. It follows (by solving a quadratic equation for ρ) that ρ itself is constant. Taking the time derivative of ρ gives: $\dot{r}_1 r_2 - \dot{r}_2 r_1 = 0$. From this and the equations for $\dot{r}_1$ and $\dot{r}_2$ it follows that

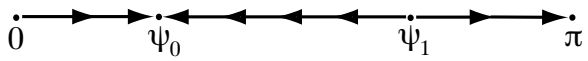
$$\begin{aligned}
0 &= \frac{\dot{r}_1 r_2 - \dot{r}_2 r_1}{r_1 r_2}\\
&= (a_2-a_1) - (r_2^k - r_1^k) - \varepsilon\,()\cos\psi,
\end{aligned}$$

and we obtain that $r_2^k - k_1^k$ is also a constant. But if $k \geq 1$ and both r_2/r_1 and $r_2^k - k_1^k$ are constants, it follows that r_1, r_2 are constants. (Notice that if $a_2 \neq a_1$, then $r_2 \neq r_1$.)

The first pair unperturbed equations have stable (attracting) solutions given by $r_i = a_i^{1/k}$, for $i = 1,2$. A small perturbation (small ε) will will still have a stable solution for r_i, for a possibly different value of constant. The question is why this should be true for ψ since the unperturbed equation for ψ is $\dot{\psi} = b_2 - b_1$, which does not have a constant solution unless $b_2 = b_1$. However, after adding the perturbation, we obtain for sufficiently small $b_2 - b_1$ roots ψ_1, ψ_2 for the equation

$$b_2 - b_1 = \varepsilon\left(\frac{r_1}{r_2}+\frac{r_2}{r_1}\right)\sin\psi = 0$$

such that $\psi_1 > 0$ $\psi_1 < \psi_2 < \pi$. A simple inspection of the sign of $\dot{\psi}$ from equation 1 shows that ψ_0 is attracting and ψ_1 is repelling.



It follows that a unique stable solution exists. For very small Δb, ψ must also be very small, so that $\sin\psi \approx \psi$. It follows that

$$\psi \approx \frac{\Delta b}{\varepsilon\left(\frac{r_1}{r_2}+\frac{r_2}{r_1}\right)}.$$

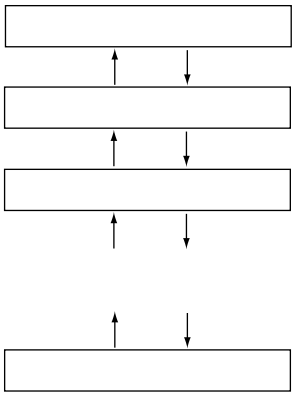
This leads to the following interesting remark. Writing $\psi = \Delta\theta$, we have for the limit as $\Delta b \to 0$ and $\Delta a \to 0$:

$$\frac{d\theta}{db} = \frac{1}{2\varepsilon}.$$

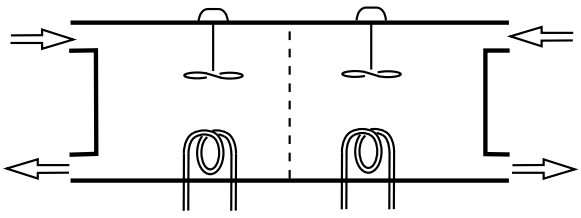
(Both r_1 and r_2 approach in the same value, given by the stable radius when $a_1 = a_1$ and $b_1 = b_2$, so $r_2/r_1 \to 1$.) Therefore

$$\theta(t) = \theta_0(t) + \frac{b}{2\varepsilon}.$$

For example, suppose that our system consists of thin layer reactors arranged vertically, with weak coupling between the layers. The reactions parameters a, b is supposed to vary with the height due to a gradient effect. Then by the last formula we should expect the phase shift in the angle theta to also depend on the height. This would produce vertical (stationary) wavy like pattern. Such patterns are indeed observed, for example, for non-stirred Belousov-Zhabotinsky reactions taking place in a long cylindrical container.



15. A CASE STUDY



16. DIFFUSION, PROBABILITY, AND THERMO-CHEMISTRY

As will be discussed later in detail, the *temperature* of a system can be identified with the average kinetic energy of the molecules comprising the system. If there are d molecules with masses

$m_1, \cdots, m_d$ (possibly all equal) and at a moment in time their velocities are $\mathbf{v}_1, \cdots \mathbf{v}_d$, then the following expression can be regarded as the definition of temperature, T, where k is the so-called Boltzmann constant:

$$\frac{1}{d}\sum_{i=1}^{d}\frac{1}{2}m_i\|\mathbf{v}_i\|^2 = \frac{3}{2}kT.$$

The reason why this definition of temperature corresponds to the empirical notion of temperature derived from the idea of thermal equilibrium is explained later.

Assuming that all the masses are equal to m, it follows that the average root-mean-square velocity is given by

$$\langle\|\mathbf{v}\|^2\rangle^{\frac{1}{2}} = \sqrt{\frac{3kT}{m}},$$

where the angle brackets designate sample average.

In an isotropic medium we should expect the three cartesian components of the velocity to be statistically independent. If this is the case then $\langle v_x^2\rangle = \langle v_y^2\rangle = \langle v_z^2\rangle = \langle\|\mathbf{v}\|^2\rangle/3$, so that the root-mean-square component of the velocity along any particular direction in space is

$$\langle v^2\rangle^{\frac{1}{2}} = \sqrt{\frac{kT}{m}}.$$

The following numerical example is taken from Berg's little book (*Random Walks in Biology*). *Lysozyme* has a molecular weight 1.4×10^4 g. This is the mass of one mole, or 6.0×10^{23} molecules; the mass of one molecule is $m = 2.3\times10^{-20}$ g. The value of kT at 300 K (27° C) is 4.14×10^{-14} g cm^2/sec^2. Therefore,

$$\langle v^2\rangle^{\frac{1}{2}} = 1.3\times10^3 \text{ cm/sec}.$$

This is the speed of thermal agitation of Lysozyme molecules in a solution at temperature 27° C. Due to constant collisions with the molecules of the solvent, it is clear that a Lysozyme molecule will not move about ten meters in a second. The actual displacement in space will be much slower. In fact, we will argue in a moment that this displacement will be proportional to the square root of time, rather than to time itself. (The constant of proportionality is the square root of the so-called *diffusion coefficient*.)

Notice that for different masses, m, M, the speeds are related by

$$\langle v_M^2\rangle^{\frac{1}{2}} = \langle v_m^2\rangle^{\frac{1}{2}}\sqrt{\frac{m}{M}}.$$

Therefore, if M is, say, 100 times m, then the corresponding root-mean-square velocity will be ten times slower than that for m.

Diffusion is the random migration of molecules or small particles arising from motion due to thermal energy. A nice simulation of molecular motion due to random thermal agitation is shown in `http://www.phy.ntnu.edu.tw/java/gas2D/gas2D.html`

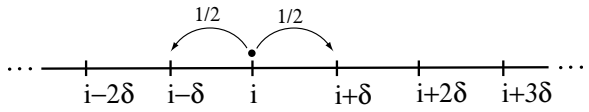
Mathematically, the bridge that connects the microscopic random agitation and the macroscopic diffusion properties is the *central limit theorem* of probability theory, in one of its many forms. We will describe this idea in some detail below.

16.1. **One-Dimensional Random Walk.** *Random walks* are simple mathematical models of the motion of a particle in a medium subject to thermal agitation.

We imagine a point particle which hops along the real line according to the following rules. Choose (small) positive numbers δ, τ and let $S_1, S_2, S_3, \cdots$ denote a sequence of independent identically distributed random variables that can take values 1 or -1 with equal probability ($p = 1/2$).

Each S_i can be thought of as flipping a fair coin the i-th time. If the particle is at position X at a given moment of time t, then at the next moment, $t+\tau$, it will be at $X+\delta S_i$. Notice that we are considering time to be discrete, measured in multiples of τ. More precisely:

(1) the position of the particle at time $t=0$ is $X_0=0$;
(2) the position at time $t=(i+1)\tau$ is $X_{i+1}=X_i+\delta S_{i+1}$, for $i=0,1,2,\cdots$;



A family of random variables $\{X_t\}$ parametrized by either continuous or discrete time is called a *stochastic process*. The fundamental problem is to obtain for our stochastic process information about its probability law $P(X_i\in[a,b]|X_0=0)$, which is the probability of finding the particle in the interval $[a,b]$ at time $i\tau$, given that it started at position 0, say.

To answer such question we need first to describe the mean and variance of the X_i. We denote by $\mu=E[X]$ the expected value of a random variable X. Recall that the variance of X is defined by $\sigma^2=E[(X-\mu)^2]$, where σ is the *standard deviation* of X. Angle brackets will be used to indicate sample averages, such as when averaging over the molecules in a container at a given time.

We will check below that:

(1) $E[X_i]=E[X_0]=0$ for all i;
(2) $E[X_i^2]=2\mathcal{D}t_i$, where $\mathcal{D}=\delta^2/2\tau$ and $t_i=\tau i$.

To check (1), simply notice that

$$E[X_i-X_{i-1}]=E[\delta S_i]=\frac{1}{2}\delta+\frac{1}{2}(-\delta)=0,$$

so that $E[X_i]=E[X_{i-1}]$ for all i.

We now obtain (2). First note that $E[S_i^2]=1$ and that X_i and S_{i+1} are independent random variables. (If they were not, X_i would be seeing into the future! Notice that X_i only involves coin tosses $S_1,\ldots,S_i$, which are independent of, i.e., cannot help predict, S_{i+1}.) Thus,

$$E[X_iS_{i+1}]=E[X_i]E[S_{i+1}]=E[X_i]0=0,$$

and it follows that:

$$\begin{aligned}E[X_{i+1}^2]&=E[(X_i+\delta S_{i+1})^2]\\&=E[X_i^2+2\delta X_iS_{i+1}+\delta^2S_{i+1}^2]\\&=E[X_i^2]+2\delta E[X_iS_{i+1}]+\delta^2E[S_{i+1}^2]\\&=E[X_i^2]+\delta^2.\end{aligned}$$

By induction, we obtain:

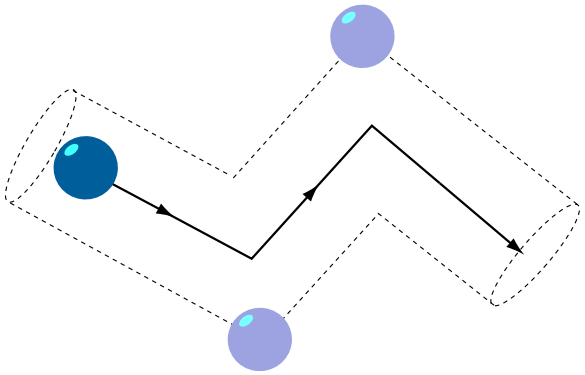
$$E[X_i^2]=\delta^2 i=\frac{\delta^2}{\tau}t_i=2\mathcal{D}t_i.$$

This result justifies the earlier claim that the particle's displacement is proportional to the square root of time.

What kind of limit should we get by letting δ and τ go to 0? The answer really is "It depends!" There are many ways to take such a limit since δ and τ have so far been regarded as independent parameters. Let us stop for a moment to consider how δ and τ should be related on the basis of gas kinetics.

16.2. **Collision Frequency and Mean Free Path.** In order to estimate the interval of time between molecular collisions we need to go back to dimension 3. We use a model for a dilute gas known as the *hard-sphere model*.

We imagine a large number of small billiard balls of diameter d, which collide elastically when their separation (the distance between their centers) equals this diameter. The average number of collisions per particle per unit time is called the *collision frequency*, f. The *mean free path*, λ, is the average distance traveled by a particle between collisions. To estimate these quantities for the hard-sphere model, consider the following picture.



A test particle moves in a zig-zag path, with speed v, in a sea of stationary particles. As it travels, the particle sweeps an imaginary tube of radius d (recall that d is the particle's diameter). The approximate volume of this tube is thus $\pi d^2 l$, where $l = vt$ is the length traveled by the particle during time t. The number of stationary particles within this "cylinder with elbows" equals the number of collisions of the test particle along its trajectory. Assuming that the gas is homogeneous, the number of particles contained within the tube is

$$\begin{aligned} \#_{\text{tube}} &= (\text{number of particles per unit volume}) \times (\text{volume of tube}) \\ &= \frac{N}{V}\pi d^2 vt. \end{aligned}$$

The collision frequency is now:

$$\begin{aligned} f &= \lim_{t\to\infty} \frac{\#_{\text{tube}}}{t} \\ &= \frac{N}{V}\pi d^2 \langle v\rangle. \end{aligned}$$

Of course, we do not have only one particle moving while the others remain fixed. A rigorous explanation would require much more work, but the estimate we get in this pedestrian discussion will suffice. Another liberty we take is to use the root mean square velocity to estimate the value of $\langle v\rangle$. Then

$$f = \text{const.}\frac{N}{V}d^2\sqrt{\frac{kT}{m}},$$

where the constant is approximately 2. (It can be calculated exactly by a more refined argument, but that will not be necessary here.)

The average time between collisions can be estimated by the reciprocal of f:

$$\tau = f^{-1} = \frac{1}{\frac{N}{V} d^2 \sqrt{\frac{kT}{m}}}.$$

The mean free path is now $\lambda = v\tau$.

Observe that

$$\mathcal{D} = \frac{\lambda^2}{2\tau} = \frac{1}{2Cd^2}\sqrt{\frac{kT}{m}},$$

where $C = N/V$ is the gas concentration. This suggests that the correct way to pass to a limit in the random walk process is to assume that $\mathcal{D}$ is constant (assuming temperature, concentration, molecular mass and molecular diameter as fixed).

Exercise 71. *Redo this derivation, assuming now that the test particle is not of the same kind as the molecules of the gas in which it is immersed. This means that the diameters and masses are no longer the same for the two species.*

16.3. **The Diffusion Limit.** The above discussion suggests that the correct way to scale δ and τ for the purpose of taking the limit of the random walk on the line is to regard $\mathcal{D} = \delta^2/2\tau$ as constant. We wish to see now what this assumption leads to.

First recall the *central limit theorem* for a sequence of coin tossings (a *Bernoulli process*). This is proved a little later in the text. If S_i is a sequence of independent, identically distributed random variables with mean 0 and standard deviation 1, and $E[|S_i|^3]$ is finite (this number is 1 in our case), then the probability distribution for the random variable $(S_1 + \cdots + S_n)/\sqrt{n}$ converges to the standard Gaussian distribution:

$$\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Another way of saying the same is that

$$P\left(a \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_i \leq b\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx,$$

where P denotes probability. If S_i has mean μ and variance σ^2, then the probability distribution is easily shown (by a change of coordinates) to be

$$\rho(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}.$$

Now, back to the random walk . We write $X(t_n) = X_n$, where $t_n = n\tau$. Then

$$\begin{aligned} X(t_n) &= \delta S_1 + \cdots + \delta S_n \\ &= \sqrt{\frac{\delta^2}{2\tau}} \sqrt{2\tau n} \frac{S_1 + \cdots + S_n}{\sqrt{n}} \\ &= \sqrt{2\mathcal{D}t_n} \frac{S_1 + \cdots + S_n}{\sqrt{n}}. \end{aligned}$$

Since $\sqrt{2\mathcal{D}t_n} S_i$ has variance $2\mathcal{D}t_n$, we obtain, as $n \to \infty$ and $t_n \to t$:

$$P(a \leq X(t_n) \leq b) \to \frac{1}{\sqrt{\frac{m}{4\pi\mathcal{D}t}}} \int_a^b e^{-x^2/4\mathcal{D}t} dx.$$

16.4. **The Diffusion Equation.** In the previous section we saw how the normal distribution arises from a random walk. We haven't considered the possibility of a boundary condition, for example, that the particle is absorbed at the first moment it hits one end of the interval. Clearly the Gaussian distribution does not represent such a situation since it is not zero over the entire real line.

A way of approaching the problem of adding a boundary process is to try to derive the probability distribution as the solution of a partial differential equation with boundary conditions.

We show now that the partial differential equation that describes a diffusion process in dimension one is

$$\frac{\partial u}{\partial t} = \mathcal{D}\frac{\partial^2 u}{\partial x^2},$$

where $u(x,t)$ is the probability density of finding the particle at x at time t, given that it started at $x = 0$ at time $t = 0$.

The function $u(x,t)$ can also be interpreted as the number of particles per unit length. In this case, rather than considering a single particle with a random position, we imagine that a large number of particles undergo independent random walks on the line, and sample average replaces expected value. We adopt the latter viewpoint now.

First, let us see what the diffusion equation above means. Any differential equation involves a combination of arithmetic operations and the passage to a limit. We can gain a good understanding of the meaning of the equation by discretizing it looking more closely at its arithmetic essence.

We discretize time and space by taking t to be an integer multiple of τ, and x an integer multiple of δ. Now observe that

$$\frac{\partial u}{\partial t}(x,t) = \frac{u(x,t+\tau) - u(x,t)}{\tau} + O(\tau),$$

and

$$\begin{aligned}\frac{\partial^2 u}{\partial x^2}(x,t) &= \frac{u_x(x+\delta/2,t) - u_x(x-\delta/2,t)}{\delta} + O(\delta) \\ &= \frac{1}{\delta}\left(\frac{u(x+\delta,t)-u(x,t)}{\delta} - \frac{u(x,t)-u(x-\delta,t)}{\delta}\right) + O(\delta) \\ &= \frac{u(x+\delta,t)+u(x-\delta,t)-2u(x,t)}{\delta^2} + O(\delta).\end{aligned}$$

Replacing these approximations into the diffusion equation gives, after simplification (notice, in particular, that $\tau\mathcal{D}/\delta^2 = 1/2$)

$$u(x,t+\tau) = \frac{1}{2}(u(x+\delta,t)+u(x-\delta,t)) + O(\delta)\tau + O(\tau)\tau.$$

We now imagine that the interval, say from a to b, is divided in a number of "cells" of length δ. The number of particles in the cell at x at time t is then $N(x,t) = u(x,t)\delta$. Disregarding the small error term, the discretized diffusion equation becomes:

$$N(x,t+\tau) = \frac{1}{2}(N(x+\delta,t)+N(x-\delta,t)).$$

Therefore, the essential content of the diffusion equation can be described in words as follows: *The number of particles at cell x at time $t+\tau$ is the average value of the number of particles occupying the neighboring cells at time t.*

It is now easy to understand why the random walk should give rise to a diffusion equation. The number of particles in cell x at time $t+\tau$ can be calculated as follows: of the $N(x,t)$ particles that were at x at time t, about half will jump to the right and half will jump to the left, no none remains.

On the other hand, half of those that were on the left neighboring cell at time t and half that were on the right neighboring cell will jump into the cell at x. Therefore, the new value at x, $N(x,t+\tau)$ is the average of the values of N at time t over the two neighboring cells.

16.5. **Dimensions 2 and 3.** Everything that was said in the previous section easily extends to dimension 2 or 3. We leave the details as an exercise to you.

The diffusion equation now reads

$$\frac{\partial u}{\partial t}(\mathbf{x},t) = \nabla^2 u(\mathbf{x},t),$$

where ∇^2 denotes the Laplace operator:

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}.$$

Exercise 72. *Suppose that $u(\mathbf{x},t)$ is a solution of the diffusion equation that decreases to 0 as $\mathbf{x} \to \infty$ with a faster rate than $1/\|\mathbf{x}\|^2$, i.e., $o(\|\mathbf{x}\|^{-2})$. Denote by $M(t)$ the integral of $u(\mathbf{x},t)$ over $\mathbf{x} \in \mathbb{R}^3$. Show that $M(t)$ is constant, i.e., the total mass of the solution is constant. In other words, if the solution is normalized for one value of t, then it is normalized for all values. (We say that the solution is* normalized *if its integral over $\mathbf{x} \in \mathbb{R}^3$ is equal to 1.) Hint: use the divergence theorem of calc. III.*

Exercise 73. *Suppose that $u(\mathbf{x},t)$ is a normalized solution of the diffusion equation with diffusivity constant $\mathcal{D}$. Show that the following functions also are normalized solutions of the same equation:*

(1) *$v(\mathbf{x},t) = u(\mathbf{x}+\mathbf{c},t)$, where $\mathbf{c}$ is an arbitrary constant vector;*
(2) *$v(\mathbf{x},t) = u(R(\mathbf{x}),t)$, where R is any orthogonal matrix of order 3;*
(3) *$v(\mathbf{x},t) = \gamma^3 u(\gamma\mathbf{x},\gamma^2 t)$, where γ is any positive number.*

This shows that the group of symmetries *of the diffusion equation contains the translations, rotations, and the time-space scaling by γ defined above, as well as all the transformation of space-time generated by compositions of these three.*

Exercise 74. *Use the above exercise to show that if $u(\mathbf{x},t)$ is a rotationally symmetric normalized solution of the diffusion equation, then so is*

$$v(\mathbf{x},t) = \frac{1}{t^{3/2}} u(\|\mathbf{x}\|^2/t, 1).$$

If $v(\mathbf{x},t) = \frac{1}{t^{3/2}} f(\|\mathbf{x}\|^2/t)$ is a solution of the diffusion equation, find an ordinary differential equation that the single variable function f must satisfy. Show that the second order differential equation for f can be written as a system of two first order equations of the form

$$\begin{aligned} 4\mathcal{D} f' + f &= h \\ 2h' - 3h &= 0. \end{aligned}$$

Solve for the general solution f. Show that the only bounded solution, $f(z)$ for $z \geq 0$, has the form $f(z) = f(0)e^{-z/4\mathcal{D}}$. To what solution $u(\mathbf{x},t)$ does it correspond?

16.6. **Boundary Value Problems I - Finite Dim Approximation.** Imposing initial and boundary conditions to the diffusion equation leads to the following class of problems.

Let $\mathcal{R}$ be a subset of $\mathbb{R}^n$ (for $n = 1, 2$, or 3, for the most part). We denote the boundary of $\mathcal{R}$ by $\partial\mathcal{R}$. Let $h(\mathbf{x})$ a function on $\mathcal{R}$ and $g(\mathbf{x})$ a function on $\partial\mathcal{R}$.

The initial value problem for the diffusion equation with *Dirichlet boundary condition* is to find $u(\mathbf{x},t)$, for $t \in [0,\infty)$, $\mathbf{x} \in \mathcal{R}$, such that

$$\begin{aligned} \frac{\partial u}{\partial t} &= \nabla^2 u \\ u(\mathbf{x},0) &= h(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathcal{R} \\ u(\mathbf{x},t) &= g(\mathbf{x}), \text{ for all } \mathbf{x} \in \partial\mathcal{R} \text{ and } t \geq 0. \end{aligned}$$

Other types of boundary conditions are often considered. The *Neumann boundary condition* specifies the normal component of the gradient (i.e., the normal derivative) of u over the boundary. We can also consider linear combinations of the Dirichlet andn Neumann conditions, as well as mixed conditions imposed on different parts of the boundary.

To develop a feeling for what the boundary value problem amounts to, we discuss here a discretized version of the Dirichlet problem in dimension 1.

We take $\mathcal{R}$ in this case to be the interval $[a,b]$, whose boundary consists of the two endpoints, a, b. In dimension one, the Dirichlet problem reduces to solving:

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{\partial^2 u}{\partial x^2} \\ u(x,0) &= h(x), \text{ for all } x \in [a,b] \\ u(a,t) &= u_{\text{left}}, \text{ for all } t \geq 0 \\ u(b,t) &= u_{\text{right}}, \text{ for all } t \geq 0, \end{aligned}$$

where $u_{\text{left}}, u_{\text{right}}$ are given numbers.

We now partition the interval $[a,b]$ into n subintervals of length $\delta = (b-a)/n$ and end-points $x_0 = a, x_1 = a+\delta, x_2 = a+2\delta, \cdots, x_{n-1} = a+(n-1)\delta, x_n = a+n\delta = b$. It is convenient to write $u(x_i,t) = u_i(t)$. The initial condition is $(u_1(0), \cdots, u_{n-1}(0)) = (h_1, \ldots, h_{n-1})$, and the boundary condition is $u_0(t) = u_{\text{left}}$, $u_n(0) = u_{\text{right}}$. Since the problem is to find the $n-1$ functions, $u_i(t)$, $i = 1, \cdots, n-1$, it will be convenient to define a vector $\mathbf{u}(t)$ that excludes the components u_0 and u_n. Thus we define the column vector

$$\mathbf{u}(t) = (u_1(t), \cdots, u_{n-1}(t))^{\text{transp.}}.$$

Now, the discretized diffusion equation, $u_i(t+\tau) = (u_{i-1}(t) + u_{i+1}(t))/2$, and the boundary condition, can be combined into the following matrix equation:

$$\begin{pmatrix} u_1(t+\tau) \\ u_2(t+\tau) \\ u_3(t+\tau) \\ \vdots \\ u_{n-2}(t+\tau) \\ u_{n-1}(t+\tau) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \\ \vdots \\ u_{n-2}(t) \\ u_{n-1}(t) \end{pmatrix} + \frac{1}{2} \begin{pmatrix} u_{\text{left}} \\ 0 \\ 0 \\ \vdots \\ 0 \\ u_{\text{right}} \end{pmatrix}$$

We write this matrix equation from now on in the form

$$\mathbf{u}(t+\tau) = \mathbf{A}\mathbf{u}(t) + \mathbf{b}.$$

To this recursive equation it should be added the initial value $\mathbf{u}(0) = \mathbf{h}$.

As an example, suppose that $a = 0$, $b = 4$, $\tau = 1$, and $\delta = 1$. For the boundary values we choose $u(0,t) = 1, u(4,t) = 5$, and the initial condition is defined by the vector $\mathbf{h} = (3,2,8)^{\text{transp.}}$. In this case, the matrices are as follows:

$$\mathbf{A} = \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{b} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 5 \end{pmatrix}.$$

A direct calculation gives the first few values:

$$\begin{aligned} \mathbf{u}(\tau)^{\text{transp.}} &= (1.500, 5.500, 3.500) \\ \mathbf{u}(2\tau)^{\text{transp.}} &= (3.250, 2.500, 5.250) \\ \mathbf{u}(3\tau)^{\text{transp.}} &= (1.750, 2.750, 3.750) \\ \mathbf{u}(4\tau)^{\text{transp.}} &= (1.875, 2.750, 3.875) \\ \mathbf{u}(5\tau)^{\text{transp.}} &= (1.937, 2.875, 3.937). \end{aligned}$$

Exercise 75. *Show that the above sequence converges to* $(2, 3.4)$. *Explain on physical grounds why this simple progression should be expected.*

Exercise 76. *In the general case, show that*

$$\mathbf{u}(k\tau) = \mathbf{A}^k\mathbf{h} + (\mathbf{b} + \mathbf{A}\mathbf{b} + \cdots + \mathbf{A}^{k-1}\mathbf{b}).$$

Show that

$$\lim_{k\to\infty} \mathbf{u}(k\tau) = \sum_{i=0}^{\infty} \mathbf{A}^i\mathbf{b} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b}.$$

Use this result to verify that the limit in the example is indeed $(2, 3, 4)$.

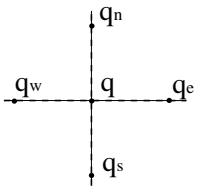
In the continuous setting, the stationary (asymptotic) value of $u(x,t)$ must be the solution of the *Dirichlet problem*:

$$\begin{aligned} &\nabla^2 f = 0 \\ &f(\mathbf{x}) = g(\mathbf{x}) \text{ for all } \mathbf{x} \in \partial\mathcal{R}. \end{aligned}$$

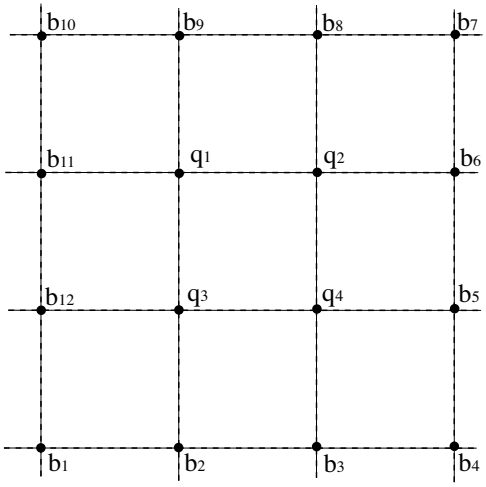
This is a rather easier problem to solve in dimension 1, since a function $f(x)$ that satisfies $f'' = 0$ must be of the form $f(x) = Ax + B$, where a and b are determined by the boundary values at the two ends of the interval $\mathcal{R} = [a,b]$. So let us consider the discretized version of this problem in dimension 2. We take for $\mathcal{R}$ a rectangular grid, in this example of 4×4 points. A solution of $\nabla^2 f = 0$ is now a function of the grid points that satisfies the discrete Laplace's equation:

$$f(q) = \frac{f(q_n) + f(q_w) + f(q_s) + f(q_e)}{4},$$

for each *interior* point q on the grid.



For a sixteen point grid, there are four values of f to determine: $f(q_i)$, $i = 1,2,3,4$. The boundary values, $f(b_i)$ are specified. In this example we assume that f is 1 on $b_1, b_2, b_3, b_4, b_7, b_8, b_9, b_{10}$ and 0 on b_5, b_6, b_{11}, b_{12}.



This leads to the system:

$$f(q_1) = \frac{1}{4}(f(q_2) + 1 + 0 + f(q_3))$$
$$f(q_2) = \frac{1}{4}(f(q_4) + 0 + 1 + f(q_1))$$
$$f(q_3) = \frac{1}{4}(1 + f(q_4) + f(q_1) + 0)$$
$$f(q_4) = \frac{1}{4}(1 + 0 + f(q_2) + f(q_3)).$$

After a little algebra, this can be put In matrix form:

$$\begin{pmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{pmatrix} \begin{pmatrix} f(q_1) \\ f(q_2) \\ f(q_3) \\ f(q_4) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

In short hand, we need to solve for the 4-dimensional vector $\mathbf{f}$ in the equation $\mathbf{Af} = \mathbf{b}$. The solution is simply $\mathbf{f} = \mathbf{A}^{-1}\mathbf{b}$. We can solve this linear system in a number of ways (e.g., row-reduction). I will simply write the result, which you can check as an exercise:

$$f(q_1) = f(q_2) = f(q_3) = f(q_4) = 1/2.$$

Exercise 77. *How could we have guessed that all values are equal by symmetry considerations?*

16.7. **Boundary Value Problems II - Fourier's Method.** Before explaining Fourier's method in the continuous case, let us look at the discrete version of the boundary value problem a little further, using now what we know about diagonalization of matrices.

First observe that the $(n-1) \times (n-1)$ matrix $\mathbf{A}$ is symmetric. As we have seen earlier, this property implies that the eigenvalues $\lambda_1, \cdots, \lambda_{n-1}$ of $\mathbf{A}$ are all real and there exists an orthonormal

basis $\mathbf{w}_1, \cdots, \mathbf{w}_{n-1}$ of $\mathbb{R}^{n-1}$ of eigenvectors. It makes sense to write $\mathbf{b} = l_1\mathbf{w}_1 + \cdots + l_{n-1}\mathbf{w}_{n_1}$, $\mathbf{h} = c_1\mathbf{w}_1 + \cdots + c_{n-1}\mathbf{w}_{n_1}$, and

$$\mathbf{u}(t) = f_1(t)\mathbf{w}_1 + \cdots + f_{n-1}(t)\mathbf{w}_{n_1}.$$

Solving for $\mathbf{u}(t)$, therefore, amounts to solving for the $f_i(t)$, as indicated in the next exercise.

Notice, in particular, that the constants l_i and c_i are quite easy to calculate, since the basis of eigenvectors is orthonormal:

$$l_i = \mathbf{b} \cdot \mathbf{w}_i, \;\; c_i = \mathbf{h} \cdot \mathbf{w}_i, \text{ for all } i.$$

Exercise 78. *Show solving for $\mathbf{u}(t)$ is equivalent to finding $f_i(t)$, $i = 1, \cdots, n-1$, such that, for each i,*

$$f_i(t+\tau) = \lambda_i f_i(t) + l_i$$
$$f_i(0) = c_i.$$

Exercise 79. *Solve the example by the above method, using the eigenvalues and eigenvectors of*

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

The recursive equation for $f_i(t)$ indicated in the previous exercise can be written in a different way, so as to look like the approximation to a differential equation:

$$\frac{f_i(t+\tau) - f_i(t)}{\tau} = \frac{\lambda_i - 1}{\tau} f_i(t) + \frac{l_i}{\tau}.$$

The significance of these algebraic remarks becomes apparent as we move from the discrete to the continuous. Fourier's approach to solving the boundary value problem for the diffusion equation also involves first solving a diagonalization problem, now for the Laplacian instead of the matrix $\mathbf{A}$. (Actually, the discrete version of the Laplacian should be thought of as being $\mathbf{I} - \mathbf{A}$.)

As can be expected, diagonalizing something like ∇^2 involves some technical points that do not arise when diagonalizing a finite dimensional matrix, as we will point out in due course. Nevertheless, we will see now that the main ideas are the same.

A preliminary remark about boundary conditions is needed. Suppose that $f(\mathbf{x})$ is a function on $\mathcal{R}$ that satisfies

$$\nabla^2 f = 0$$
$$f(\mathbf{x}) = g(\mathbf{x}) \text{ for all } \mathbf{x} \in \partial\mathcal{R}.$$

If now $v(\mathbf{x}, t)$ is a solution of the diffusion equation with initial condition $v(\mathbf{x}, 0) = h(\mathbf{x})$ and boundary value 0, then $u(\mathbf{x}, t) = v(\mathbf{x}, t) + f(\mathbf{x})$ is a solution of the diffusion equation with boundary value $g(\mathbf{x})$ and initial condition $u(\mathbf{x}, 0) = h(\mathbf{x}) + f(\mathbf{x})$. Thus solving the diffusion equation for a non-zero boundary condition amounts to finding a harmonic function (a solution of Laplace's equation $\nabla^2 f = 0$) with the given boundary value, and separately solving the diffusion equation with boundary value 0 and a different initial condition. In what follows, we will limit ourselves to the case of zero boundary value.

We look for a family of functions $\phi_m(\mathbf{x})$ such that

$$\nabla\phi_m = \lambda_m\phi_m$$
$$\phi_m(\mathbf{x}) = 0, \text{ for all } \mathbf{x} \in \partial\mathcal{R}.$$

If we can find such $\phi_m(\mathbf{x})$ in sufficient number to generate all other functions by (infinite) linear combinations, then we can look for a solution of the diffusion problem in the form

$$u(\mathbf{x},t) = \sum_m f_m(t)\phi_m(\mathbf{x}).$$

The functions $f_m(t)$ must satisfy the ordinary differential equation

$$f_m' = \mathcal{D}\lambda_m f_m,$$

therefore $f_m(\mathbf{x}) = C_m e^{\mathcal{D}\lambda_m t}$ for some constant C_m which is determined by the initial condition

$$h(\mathbf{x}) = \sum_m C_m\phi_m(\mathbf{x}).$$

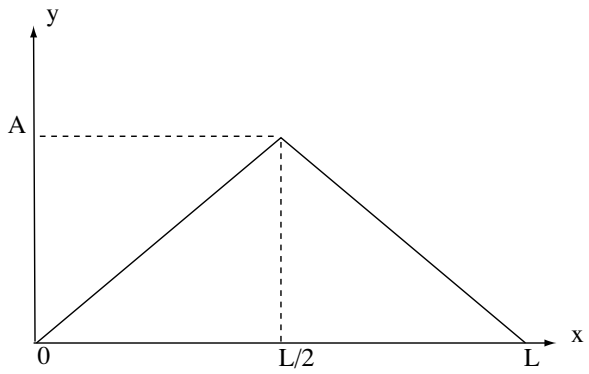
This leads to an expression for the solution $u(\mathbf{x},t)$ in the form:

$$u(\mathbf{x},t) = \sum_m C_m e^{\mathcal{D}\lambda_m t}\phi_m(\mathbf{x}).$$

Let us consider the following example. Here $\mathcal{R}$ is the interval $[0,L]$.

$$\begin{aligned}
\frac{\partial u}{\partial t} &= \mathcal{D}\frac{\partial^2 u}{\partial x^2} \\
u(0,t) &= 0; \\
u(L,t) &= 0; \\
u(x,0) = h(x) &= \begin{cases} 2Ax/L, & \text{for } x \in [0,L/2] \\ 2A(1-x/L), & \text{for } x \in [L/2,L]. \end{cases}
\end{aligned}$$

The graph of $h(x)$ is shown next.



We first wish to find the *eigenfunctions* $\phi(x)$ associated to ∇^2, in this case simply $\frac{\partial^2}{\partial x^2}$, satisfying the boundary conditions $\phi(0) = \phi(L) = 0$. Being an eigenfunction now means that

$$\phi''(x) = \lambda\phi(x)$$

for some λ.

Exercise 80. *Show that if ϕ satisfies $\phi'' = \lambda\phi$, $\phi(0) = \phi(L) = 0$, then, for some integer n,*

$$\lambda_n = -\pi^2 n^2/L^2$$
$$\phi_n(x) = A\sin(n\pi x/L).$$

The constant A is arbitrary.

As already noted, we look for a solution $u(x,t)$ of the form:

$$u(x,t) = \sum_m C_m e^{\mathcal{D}\lambda_m t}\phi_m(x),$$

where the constants C_m should be chosen so that

$$h(x) = \sum_m C_n\phi_m(x).$$

Therefore, it is necessary to find an expansion of the tent function $h(x)$ into a series of sine functions. It is not difficult to find out what C_m should be *if* such expansion is possible. We describe now how to find C_m and leaf for later a discussion of the meaning of writing $h(x)$ as a trigonometric series (a *Fourier sine series*)

$$h(x) = \sum_{m=1}^{\infty} C_m \sin(n\pi x/L).$$

Before determining C_m, recall the following integrals, which I leave for you to calculate as an exercise:

$$\int_0^L \sin(n\pi x/L)\sin(m\pi x/L)dx = \begin{cases} 0 & \text{if } n \neq m; \\ L/2 & \text{if } n = m. \end{cases}$$

Exercise 81. *Using the above integral, show that if $h(x)$ can be written as a sine series, then*

$$C_m = \frac{2}{L}\int_0^L h(x)\sin(m\pi x/L)dx.$$

Let us see what this gives for the tent function.

Exercise 82. *Show that for the tent function (above figure), the value of C_m is*

$$C_m = \frac{8A}{(n\pi)^2}\sin(n\pi/2).$$

Therefore

$$\begin{aligned} h(x) &= -\frac{8A}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{(2n-1)^2}\sin((2n-1)\pi x/L) \\ &= \frac{8}{\pi^2}\left(\sin(\pi x/L) - \frac{1}{9}\sin(3\pi x/L) + \frac{1}{25}\sin(5\pi x/L) - \cdots\right). \end{aligned}$$

We can now write the solution $u(x,t)$ as

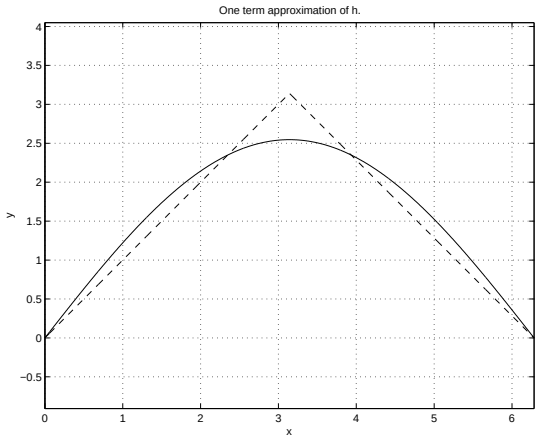
$$u(x,t) = \frac{8A}{\pi^2}\sum_{n=1}^{\infty}\sin(n\pi/2)e^{-\mathcal{D}\pi^2 n^2 t/L^2}\sin(n\pi x/L).$$

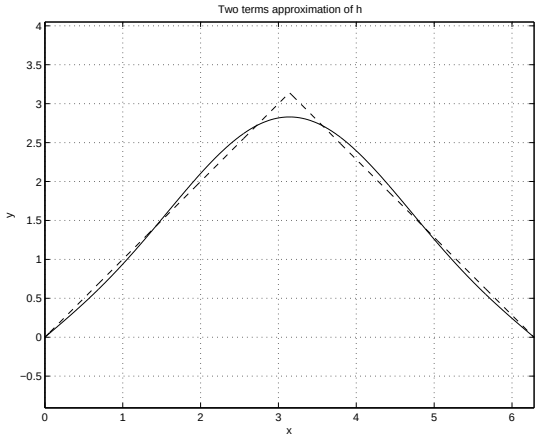
Notice that $u(x,t) \to 0$ as $t \to \infty$.

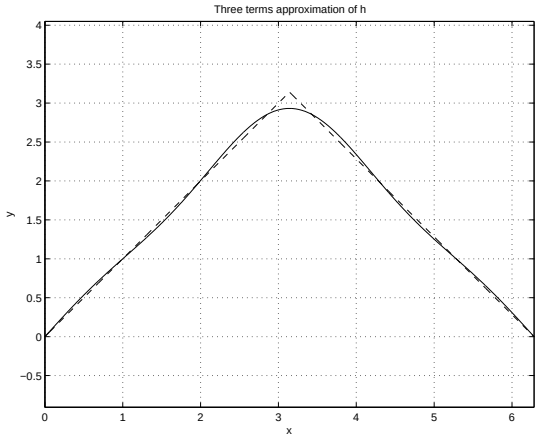
Without going into the issue of convergence of the series, we can see that it makes sense numerically. The following figures show the graphs of the series for $h(x)$ truncated at finite values of n. We take $A = \pi = L/2$.

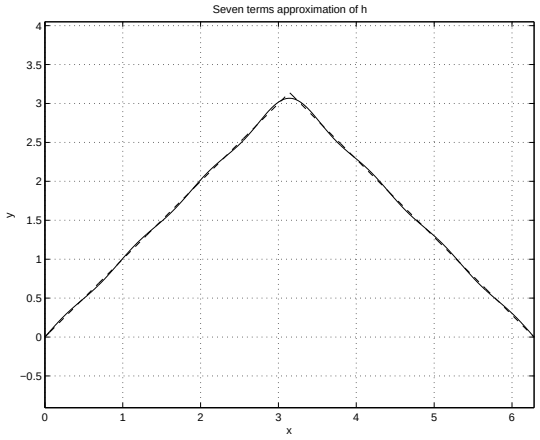
The following graph was obtained by Matlab using the commands (obvious modifications will give the other three):

```
>>x=0:2*pi/1000:2*pi;
>> h1=(8/pi)*sin(x/2);
>> plot(x,h1)
>> grid
>> hinf=pi-abs(x-pi);
>> hold on
>> plot(x,hinf,'--')
>> axis equal
>> xlabel('x')
>> ylabel('y')
>> title('One term approximation of h.')
```



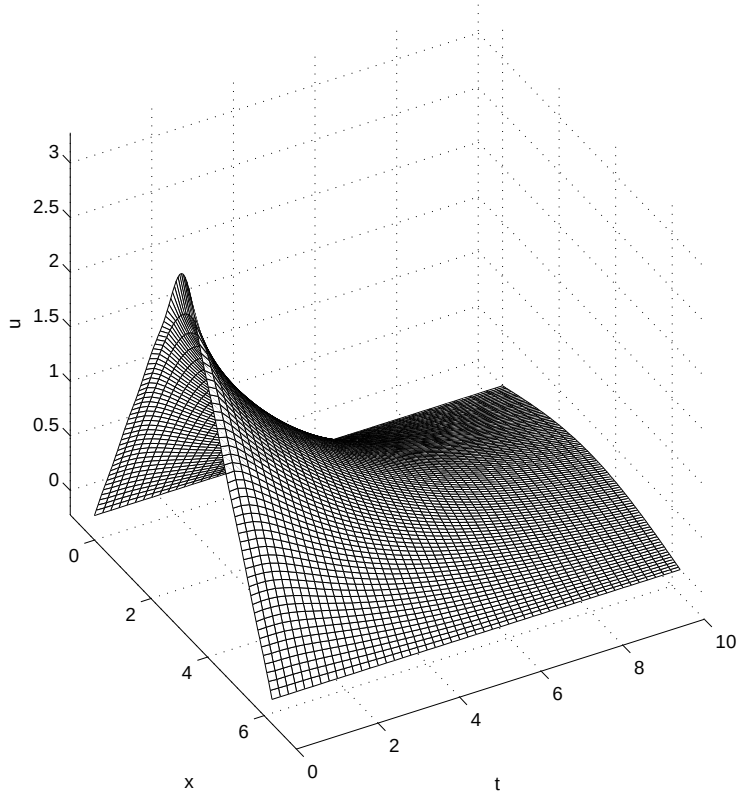

Two terms approximation of h
y
x


Three terms approximation of h
y
x


Seven terms approximation of h
y
x

The graph of $u(x,t)$ is shown next. We choose $\mathcal{D}=1$ and approximate $u(x,t)$ to seven terms:

$$u(x,t) \approx \frac{8}{\pi}\left[e^{-t/4}\sin(x/2) - \frac{1}{3^2}e^{-3^2t/4}\sin(3x/2) + \cdots + \frac{1}{13^2}e^{-13^2t/4}\sin(13x/2)\right].$$



The previous graph was obtained using the Matlab commands:

```
>> x=0:2*pi/100:2*pi;
>> t=0:0.2:10;
>> [x,t]=meshgrid(x,t);
>> u=(8/pi)*(exp(-t/4).*sin(x/2)-exp(-9*t/4).*sin(3*x/2)/9+
exp(-25*t/4).*sin(5*x/2)/25-
exp(-49*t/4).*sin(7*x/2)/49+exp(-81*t/4).*sin(9*x/2)/81-
exp(-121*t/4).*sin(11*x/2)/121+
exp(-169*t/4).*sin(13*x/2)/169);
>> mesh(x,t,u)
>> xlabel('x')
>> ylabel('t')
>> zlabel('u')
```

Observe that as t grows, each term in the series decays exponentially to 0. The rate of decay is faster for terms of higher order. Therefore, for large enough values of t the approximation

$$u(x,t) \approx \frac{8}{\pi} e^{-t/4} \sin(x/2)$$

already can be quite reasonable.

16.8. **Eigenfunctions of ∇^2 in Dimension 2.** What we say in the previous section points to the importance of determining the eigenfunctions of the Laplacian operator ∇^2 (with boundary condition 0). Recall that finding these eigenfunctions amounts to solving the boundary value problem:

$$\nabla^2 \phi = \lambda \phi$$
$$\phi(\mathbf{x}) = 0 \text{ for all } \mathbf{x} \in \partial \mathcal{R}.$$

We use in the previous section that the functions $\phi_m(x) = \sin(n\pi x/L)$ are eigenfunctions of $\frac{\partial^2}{\partial x^2}$ that vanish on the endpoints of the interval $\mathcal{R} = [0,L]$ and claimed without proof that this family is *complete* in the sense that other (say, continuous) functions on $\mathcal{R}$ that are zero at 0 and L can be expressed as an infinite sum of the form $\sum_m c_m \phi_m$.

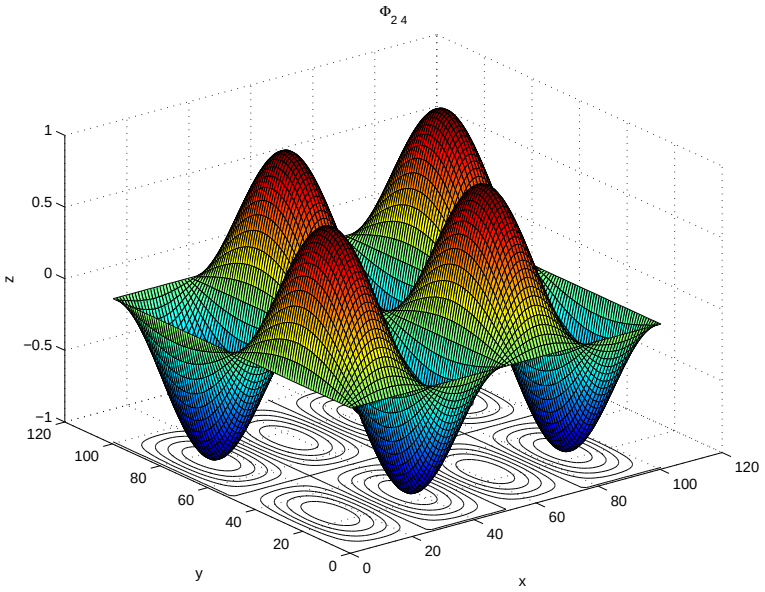
Without elaborating on the issue of completeness of the ϕ_m, we can use it to convince ourselves of the following fact, which is left for you to check as an exercise.

Exercise 83. *Show that the functions*

$$\Phi_{mn}(x,y) = \sin(n\pi x/L)\sin(m\pi y/M)$$

for m,n positive integers, constitute a complete set of eigenfunctions of the Laplacian in dimension 2, vanishing on the boundary of the rectangle $\mathbf{R} = [0,L] \times [0,M]$.

The next figure shows a contour plot of Φ_{24}, $L = M = 100$.



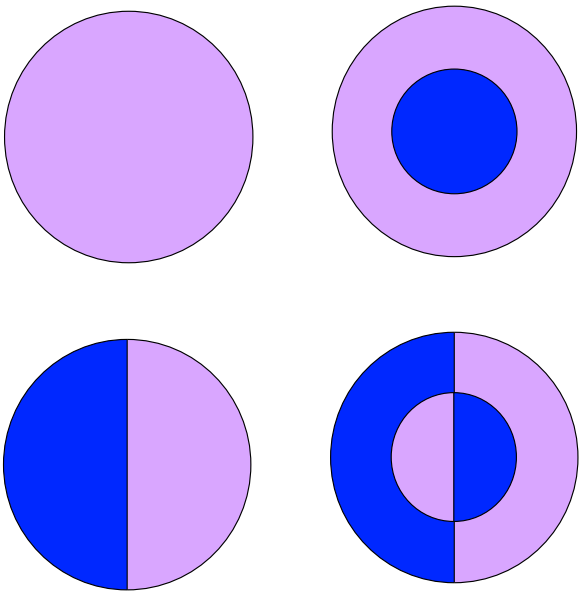
Exercise 84. *Find the Fourier series (in terms of the $\Phi_{mn}(x,y)$) of the function $f(x,y)$ on $\mathcal{R} = [-2,2]\times[-2,2]$ which takes the value 1 on $[-1,1]\times[-1,1]$ and 0 on the complement. (Although this is not a continuous function, the series will still converge to the function at all points except on the points of discontinuity, on which the series converges to $1/2$.)*

The eigenvalues and eigenfunctions of the Laplacian reflect the shape of the region. Let us look at the somewhat more complicated case of a disc. (What makes the disc, or any other shape, "more complicated" than the rectangle is that the latter is a Cartesian product of intervals, allowing us to recycle the results for intervals we've already obtained.)

Since the problem is rotationally symmetric (both ∇^2 and the boundary condition are invariant under rotations about the origin), it should be helpful to work in polar coordinates. We first need to express the Laplacian in polar coordinates: $x = r\cos\theta, y = r\sin\theta$.

Exercise 85. *Show that the Laplacian can be written in polar coordinates as follows: If $F(r,\theta) = f(x,y)$, then*

$$\nabla^2 f = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial F}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 F}{\partial \theta^2}.$$



To develop: More on the Laplacian and coordinate change, spectrum on a disc and other geometries and relationship with special functions. Bessel functions, etc.

17. Numerical Study of a Reaction-Diffusion System

Two interacting substances diffuse on $\mathbb{R} = [0,1]$, having concentrations $u(x,t)$ and $v(x,t)$, for $x, \in \mathcal{R}$ and $t \in [0,\infty)$, that satisfy the system:

$$\frac{\partial u}{\partial t} = a - u + u^2 v + \alpha \frac{\partial^2 u}{\partial x^2}$$
$$\frac{\partial u}{\partial t} = b - u^2 v + \beta \frac{\partial^2 v}{\partial x^2}.$$

We choose the following boundary conditions:

$$\frac{\partial u}{\partial x}(0,t) = -B_0$$
$$\frac{\partial u}{\partial x}(1,t) = B_1$$
$$\frac{\partial v}{\partial x}(0,t) = -C_0$$
$$\frac{\partial v}{\partial x}(1,t) = C_1.$$

The constants B_0, B_1, C_0, C_1 specify the fluxes in (if they are positive) or out (if negative) of each of the two substances through both ends of the interval. (The negative sign in front of B_0 and C_0 are necessary if we wish to interpret flows out of $[0,1]$ to correspond to positive B_0, B_1, C_0, C_1, for either end.) It the system is isolated, all constants are 0.

It is also necessary to specify initial conditions: $u(x,0) = h(x)$ and $v(x,0) = g(x)$. We thus have altogether 8 parameters and two functions that can be (mathematically) varied at will: $a, b, \alpha, \beta, B_0, B_1, C_0, C_1$; and two functions $h(x), g(x)$.

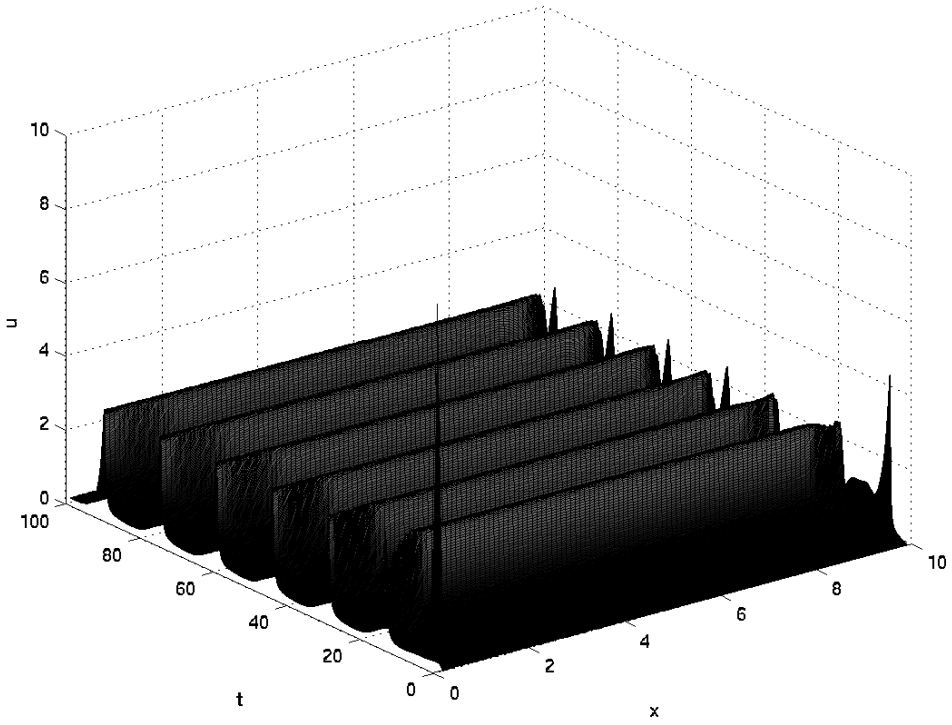
In this section we simply play with this system numerically, varying the the parameters of the equations, the boundary constants, and the initial conditions, to see what can happen. To this end we need to approximate it by a system of algebraic equations. We do this next.

The Matlab script is shown next. This saved as an "m file" (the name I used is "reacdiff.m") and is run by calling `>> reacdiff` at the Matlab's command window. You should experiment with different values for the diffusion and reaction constants.

```
% Simulation of a system of reaction-diffusion equations.
% The two dependent variable are the
% concentrations, U(x,t) and V(x,t), of two substances.
% Here x lies on the interval [0,L] and t on the interval
% [0,T].
% The equations are:
% Ut = a - U + U^ 2*V + alpha*Uxx
% Vt = b - U^ 2*V + beta*Vxx
% where Ut, Vt indicate partial derivatives in t, and
% Uxx, Vxx indicate second order partial derivatives
% in x.
% Initial and boundary conditions are defined below.
% The result of the computation is a pair of matrices
% U and V of order N-1 by M. The columns are the discretized
```

```
% functions x->U(x,t) and x->V(x,t) for the different values of
% the discretized time.
% parameters to specify:
clear
% time duration:
T=100;
% length of interval:
L=10;
% number of sub-intervals:
N=100;
% length of sub-intervals:
dx=L/N;
% number of time sub-intervals:
M=8000;
% time step:
dt=T/M;
% diffusivity of substance with concentration "u":
alpha=0.05*dx^ 2/(2*dt);
% diffusivity of substance with concentration "v":
beta=1*dx^ 2/(2*dt);
% boundary constants (specify flows at the ends of interval [0,L].
% -B0 (resp., -C0) is the outflow of u (resp., v) at 0.
% B1 (resp., C1) is the outflow of u (resp., v) at L.
B0=0;
B1=0;
C0=0;
C1=0;
% initial values for u and v (which are column vectors
% dimension N-1):
u0=zeros(N-1,1);
v0=zeros(N-1,1);
u0(1)=1/dx;
v0(N-1)=1/dx;
% in this example, u0 is a pulse of mass one at the beginning of
% the interval.  v0 is a pulse of mass one at the end.
% these are the two reaction constants:
a=1/(6*sqrt(3));
b=2/(3*sqrt(3));
% the above are all the variables for which values must be provided.
% the following are auxiliary parameters related to diffusivity:
D0=dx^ 2/(2*dt);
Du=alpha/D0;
Dv=beta/D0;
% the matrix "A" is the averaging part of the Laplacian,
% adapted for the Neumann boundary condition.
% It has 1s on the two off-diagonals (imediately above and
```

```
% below the main diagonal), 1s at the North-West and South-East
% corners, and 0s at all the other places.
A=(1/2)*(diag(ones(N-2,1),1)+diag(ones(N-2,1),1)');
A(1,1)=1/2;
A(N-1,N-1)=1/2;
% the following matrice is, essentially, the
% one-dimensional Laplacian (adapted for the
% type of boundary condition we are using):
L=A-eye(N-1);
% it will be convenient to express the boundary values
% as column vectors "B, C" of dimension N-1, as follows
% (notice that only the first and last coordinates are
% non-zero):
B=zeros(N-1,1);
C=zeros(N-1,1);
B(1)=(dx/2)*B0;
B(N-1)=(dx/2)*B1;
C(1)=(dx/2)*C0;
C(N-1)=(dx/2)*C1;
% the first column of U and V are given by
% the choice of initial conditions:
U(:,1)=u0;
V(:,1)=v0;
% Main calculation:
for j=1:M-1
U(:,j+1)=U(:,j)+dt*(a*ones(N-1,1)-U(:,j)+(U(:,j).^ 2).*V(:,j)+Du*L*U(:,j)+B);
V(:,j+1)=V(:,j)+dt*(b*ones(N-1,1) -(U(:,j).^ 2).*V(:,j)+Dv*L*V(:,j)+C);
end
% plotting U(x,t):
i=1:N-1;
j=1:M;
r=dx*i;
s=dt*j;
[x,t]=meshgrid(r,s);
surf(x,t,U')
```



(Find different ranges of parameters with spacial waves or some other interesting phenomenon. Improve numerics, particularly in time direction, so that we can take a longer time period; replace Runge-Kutta for the Euler method, used above.)

17.1. **Exact Results.** We take here a more theoretical approach to study the reaction-diffusion system described numerically above. (Perturbation methods, wave fronts, chaotic behavior, etc.)

18. Pattern Formation

Turing instability and bifurcations; activator-inhibitor systems; mathematical model of cover patterns in animals. (See hand-outs. Notes from Murray, and Britton.)

19. Probability Theory - A Little Less Pedestrian

To develop: Brownian motion, ergodicity, Maxwell-Boltzmann, Gibbs ensembles and Gibbs probabilities, collision theory and reaction rates, Arrhenius constant, energy of activation, energy of reaction; random flights: Poisson time, Gaussian impulse, Rayleigh flight, Levy flight, etc.

19.1. **An Informal Introduction to Brownian Motion.** A "deterministic" system of ordinary differential equations is specified by a vector field. Say that X denotes a differentiable vector field on $\mathbb{R}^n$, possibly time dependent. Then a solution to the initial value problem

$$\dot{x} = X(t,x), x(0) = x_0 \tag{2}$$

is a function $x(t)$ that describes a differentiable curve passing through the point $x_0 \in \mathbb{R}^n$ whose derivative vector (velocity) is equal to $X(t,x(t))$ at each point $x(t)$ along the curve.

For example, consider the initial value problem

$$\ddot{z} + a\dot{z} + bz = f(t)$$
$$z(0) = z_1$$
$$\dot{z}(0) = z_2.$$

The solution might represent, for example, the motion of a mass attached to a spring, with position and velocity at time 0 specified by the two initial conditions. If one defines

$$x = \begin{pmatrix} z \\ \dot{z} \end{pmatrix},\ A = \begin{pmatrix} 0 & 1 \\ -b & -a \end{pmatrix},\ F(t) = \begin{pmatrix} 0 \\ f(t) \end{pmatrix},\ X(t,x) = Ax + F(t),\ x_0 = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

then the second order equation can be expressed by the first order system and initial condition 2. The solution can be written as

$$x(t) = e^{tA}x_0 + e^{At}\int_0^t e^{-As}F(s)ds. \tag{3}$$

In a more realistic description of the motion, $x(0)$ might be regarded as a random quantity, taking a range of values according to some given probability distribution. Furthermore, the forces acting on the mass may also have a random component. For example, if we imagine that the mass is very small and that the motion can be affected to some small degree by random motion of air molecules, then it would be appropriate to describe the total force acting on the mass as also being a random - or "noisy" - quantity. We could then try to represent the problem by an equation of the type

$$\dot{x} = X(t,x) + N(t) \tag{4}$$

where $N(t)$ corresponds to the noise, ("random component") of the driving vector field. If N were an ordinary continuous function, the solution to 4 would be

$$x(t) = x_1(t) + e^{At}\int_0^t e^{-As}N(s)ds, \tag{5}$$

where $x_1(t)$ is the solution to 2.

We are faced with the issue of defining $N(t)$. Although a random process $N(t)$ having the properties one would like to assign to "noise" cannot be easily defined, it turns out that $B(t) = \int_0^t N(s)ds$ can be given a relatively simple mathematical interpretation and serve as the foundation for a calculus with noise.

Our first order of business will be to obtain a mathematical characterization of $B(t)$ (first on $\mathbb{R}^n$ and, later, on general Riemannian manifolds).

Some of the desirable properties that $B(t)$ should have are:

(1) Unpredictability: the increments $\Delta B(t) = B(t+\Delta t) - B(t)$ are completely unpredictable, that is, $\Delta B(t)$ is independent of its past $\{B(s) : s \leq t\}$;
(2) Stationarity: $B(t)$ is stationary, in the sense that the probability distribution of $\Delta B(t)$ does not depend on t;
(3) Continuity: if we imagine that the erratic, unpredictable behaviour of $B(t)$ is the result of a large number of relatively weak independent factors, then it is reasonable to expect that $B(t)$ does not have finite jumps, that is, that $B(t)$ is continuous in some sense.

It will be seen that the previous properties characterize the so called *Brownian motion*, or *Wiener*, process.

It is natural to ask whether one can find a probability density $\rho(t,x)$ that gives the probability of finding $x(t)$ in a region $K \subset \mathbb{R}^2$ as an integral

$$P(x(t) \in K) = \int_K \rho(t,x)dx.$$

It will be seen later that ρ is a solution of a parabolic (deterministic) partial differential equation - a diffusion equation.

19.2. **The Microscopic View.** We would like to take here a closer look at noise from a microscopic viewpoint, illustrating the main points with a simple mechanical example.

Suppose that a body of mass M, whose motion we would like to describe, has the shape of a parallelepiped and moves freely, without friction, on a horizontal rail. The body will be imagined to have sufficiently small mass for its motion to be affected by the thermal motion of gas molecules around it.

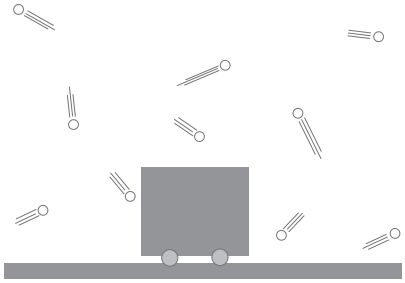
To study the motion of the body it will be convenient to keep in mind a number of different time scales: First, there is the "macroscopic scale," in which some observable displacement can be detected (say, a few minutes.) Next, we may consider the "calculus scale". Quantities that we will write as Δt (such as when we approximate the derivative of a differentiable function by the quotient $(f(t+\Delta t) - f(t))/\Delta t$) will belong to this scale. Our Δt might be of the order of, say, hundredths of a second. Finally, we will distinguish the "microscopic scale" measuring the typical time interval between collisions of gas molecules with the body. The mean time between collisions will be denoted by τ, and will depend on the gas density and mean velocity.

In order to be a little more precise it will be convenient to define

$$\gamma := \frac{1-\frac{m}{M}}{1+\frac{m}{M}},$$

where m is the mass of individual gas molecules. We assume that we can choose times scales so that the following hold:

(1) Δt should be sufficiently bigger than τ that $\Delta t/\tau$ will come "very close" to being the number of collisions during an interval $[t, t+\Delta t]$. (The law of large numbers should lie behind this property.
(2) The ratio m/M is so small that γ^i is "very close" to 1 for all $i \leq [\Delta t/\tau]$.



The component of the molecules' velocity along the direction of the rail will be written v. It will be regarded as a random variable with mean 0 and mean square

$$\mathbb{E}(v^2) = \frac{kT}{m},$$

where k is a physical quantity known as the *Boltzmann constant* and T is the *temperature* of the gas. (The temperature T may be defined, up to constant, as the mean kinetic energy of gas molecules. This will be further discussed later on.)

The effect that a collision with a gas molecule will have on the body can be calculated as follows. Denoting by V, V', and v, v' the velocities (along the rail) of the body (capitalized) and molecule (lower case) before and after the collision, then we have

$$MV + mv = MV' + mv'$$
$$\frac{1}{2}MV^2 + \frac{1}{2}mv^2 = \frac{1}{2}MV'^2 + \frac{1}{2}mv'^2,$$

where the first equation describes conservation of momentum and the second describes concervation of energy. (we are assuming that the collisions are perfectly elastic.)

These equations are easly solved, giving:

$$v' = (1+\gamma)V - \gamma v$$
$$V' = \gamma V + (1-\gamma)v.$$

Now suppose that the initial velocity of the body is $V_0 = V(0)$, and that the collision times after $t = 0$ are $T_1 < T_2 < T_2 < \dots$. The horizontal components of the velocities of the colliding molecules at the respective times are $v_1, v_2, v_2, \dots$. We assume that v_i are independent identically distributed random variables (with mean 0 and mean square kT/m) and that $\tau_i = T_{i+1} - T_i$ are independent identically distributed random variables with mean τ.

Then, for $t \in [T_i, T_{i+1})$, and writing $V_i := V(T_i)$,

$$V_i = V(t) = \gamma V_{i-1} + (1-\gamma)v_i.$$

Iterating the previous equation we obtain, after n collisions,

$$V_n = \gamma^n V_0 + (1-\gamma)\sum_{i=0}^{n-1} \gamma^i v_{n-i}.$$

By the law of large numbers, the number of collisions up to time t is, approximately, $n = [t/\tau]$. Therefore, we write

$$V(t) = e^{-\alpha t} V_0 + (1-\gamma) \sum_{i=0}^{[\frac{t}{\tau}]-1} \gamma^i v_{[t/\tau]-i},$$

where $\alpha := (\ln \gamma^{-1})/\tau$.

The exponentially decaying term $e^{-\alpha t} V_0$ shows that after a little while any memory of the initial velocity is lost. From now on, we ignore that term by assuming that $V_0 = 0$.

We make now a key observation. Write $\Delta V_t := V(t+\Delta t) - V(t)$. Then

$$\begin{aligned}
\Delta V_t &= (1-\gamma) \left(\sum_{i=0}^{[\frac{t+\Delta t}{\tau}]-1} \gamma^i v_{[\frac{t+\Delta t}{\tau}]-i} - \sum_{i=0}^{[\frac{t}{\tau}]-1} \gamma^i v_{[\frac{t}{\tau}]-i} \right) \\
&= (1-\gamma) \left(\sum_{i=1}^{[\frac{t+\Delta t}{\tau}]} \gamma^{[\frac{t+\Delta t}{\tau}]-i} v_i - \sum_{i=1}^{[\frac{t}{\tau}]} \gamma^{[\frac{t}{\tau}]-i} v_i \right) \\
&= (1-\gamma) \left(\sum_{i=[\frac{t}{\tau}]+1}^{[\frac{t+\Delta t}{\tau}]} \gamma^{[\frac{t+\Delta t}{\tau}]-i} v_i - \sum_{i=1}^{[\frac{t}{\tau}]} \left(\gamma^{[\frac{t+\Delta t}{\tau}]-i} - \gamma^{[\frac{t}{\tau}]-i} \right) v_i \right).
\end{aligned}$$

We are assuming that the v_i are independent and identically distributed, with mean 0 and mean square $\nu^2 := kT/m$. In particular, $\mathbb{E}[v_i v_j] = 0$ whenever $i \neq j$. With that in mind, we write:

$$\begin{aligned}
\mathbb{E}[(\Delta V_t)^2] &= \nu^2 (1-\gamma)^2 \left(\sum_{i=[\frac{t}{\tau}]+1}^{[\frac{t+\Delta t}{\tau}]} \gamma^{2([\frac{t+\Delta t}{\tau}]-i)} - \sum_{i=1}^{[\frac{t}{\tau}]} \left(\gamma^{2([\frac{t+\Delta t}{\tau}]-i)} - \gamma^{2([\frac{t}{\tau}]-i)} \right) \right) \\
&= \nu^2 (1-\gamma)^2 \left(\sum_{i=0}^{[\frac{t+\Delta t}{\tau}]-[\frac{t}{\tau}]-1} \gamma^{2i} - \left(\gamma^{2([\frac{t+\Delta t}{\tau}]-[\frac{t}{\tau}])} - 1 \right) \sum_{i=0}^{[\frac{t}{\tau}]-1} \gamma^{2i} \right)
\end{aligned}$$

We now use the scale assumptions. A more precise statement of the assumptions that is actually used is the following. The quantities γ and $\Delta t/\tau$ are such that $(1-\gamma^i)^2$ is much smaller than $1-\gamma^2$ for all $i = 1, \ldots, [\Delta t/\tau] + 1$. (This is clearly possible since the limit of $(1-\gamma^i)^2/(1-\gamma^2)$ as $\gamma \to 1$ is 0.)

Under these approximations, the second sum is close to 0, γ^{2i} is close to 1 for all $i = 1, \ldots, [\Delta t/\tau] + 1$, and we have (disregarding small errors)

$$\mathbb{E}[(\Delta V_t)^2] = \nu^2 (1-\gamma)^2 \left([\frac{t+\Delta t}{\tau}] - [\frac{t}{\tau}] \right) = \text{Constant} \Delta t.$$

The above calculation actually shows that ΔV_t is close to

$$(1-\gamma)(\sum_{i=[\frac{t}{\tau}]+1}^{[\frac{t+\Delta t}{\tau}]} v_i),$$

a sum of independent identically distributed random variables with mean 0 and finite variance. Since, by our scale assumption, $\Delta t/\tau$ is big, the central limit theorem implies that ΔV_t is a centered Gaussian random variable, with variance $C\Delta t$.

Therefore, the random process $V(t)$ has the following properties: ΔV_t is a Gaussian random variable with mean 0, variance $C\Delta t$ and it is independent of $\{V(s) : s \leq t\}$. Later on, we will use precisely these properties to define Brownian motion.

Notice that in this physical situation it is the velocity process that is a Brownian motion, whereas the position process is obtained by integrating $V(t)$.

19.3. **Temperature and the Maxwell-Boltzmann distribution.** Consider a collection of d noninteracting point masses of equal mass m, moving freely inside a rectangular box B with solid walls. As each particle reaches a wall of the box, it bounces off according to the usual equal angles law. The total energy of the gas as a function of the velocities is

$$E(v_1, \ldots, v_d) := \sum_{i=1}^{d} \frac{1}{2} m |v_i|^2.$$

It is assumed constant and equal to E. The mean energy per molecule will be written as

$$\frac{E}{d} =: \frac{3}{2} kT$$

where k is a constant independent of E and d (the Boltzmann constant) and T is the *temperature* of the gas. The phase space of the system (whose points represent the positions and velocities of the particles) is

$$(B \times \mathbb{R}^3)^d = B^d \times \mathbb{R}^{3d}.$$

Assuming that the total energy is constant and equal to E (the box is thermally insulated so that no energy exchange takes place with the outside), the part of phase space the gas may occupy is $\Omega := B^d \times S$, where S is the sphere in $\mathbb{R}^{3d}$ with center 0 and radius $R = \sqrt{3kT/m}\, d^{1/2} = cd^{1/2}$. We now make the assumption that the particles are distributed in Ω according to the uniform distribution. This means that P will be taken to be the normalized volume measure on Ω.

We would like to determine the probability that a velocity component of a given particle will fall in the interval $[a,b]$. Clearly, the position of the particle is immaterial, so the problem has the following geometric formulation. Let S denote the sphere in $\mathbb{R}^{3d}$ with center 0 and radius $R = cd^{1/2}$, given the probability measure P corresponding to normalized area measure. Let $(x_1, \ldots, x_n)$, $n = 3d$, denote the coordinates of $\mathbb{R}^{3d}$. Then the problem is to find the normalized area of the subset of S determined by $a \leq x_n \leq b$. In other words, we look for the measure of the interval $[a,b]$ with respect to the measure $\mu_d := x_{n*}P$. (By spherical symmetry, μ_d does not depend on the coordinate chosen.)

Assume that $-R < a < b < R$. (Notice that μ_d is supported on $[-R,R]$, that is, the complement of this interval has measure 0.) A simple calculation shows that the probability is given by

$$\mu_d([a,b]) = \frac{\int_a^b [1 - u^2/R^2]^{(3d-3)/2}\, du}{\int_{-R}^{R} [1 - u^2/R^2]^{(3d-3)/2}\, du}.$$

Using $\lim_{n\to\infty}(1 - \alpha/n)^n = e^{-\alpha}$ and $\int_{-\infty}^{\infty} e^{-\frac{mu^2}{2kT}} du = (\frac{m}{2\pi kT})^{1/2}$ we obtain

$$\lim_{d\to\infty} \mu_d([a,b]) = (\frac{m}{2\pi kT})^{1/2} \int_a^b e^{-\frac{mu^2}{2kT}} du.$$

Therefore, the components of the velocity of the gas molecules are normally distributed, with density

$$\rho(x) = (\frac{m}{2\pi kT})^{1/2} e^{-\frac{mx^2}{2kT}}.$$

It follows that the probability density for the velocity of individual molecules is given by the *Maxwell-Boltzmann* distribution

$$\rho(v) = (\frac{m}{2\pi kT})^{3/2} e^{-\frac{m|v|^2}{2kT}}.$$

This shows that v is a normal random variable with mean 0 and variance kT/m.

In order to understand the significance of the temperature parameter T, consider the following remark. Suppose that the insulated box contains two types of (point) particles that can be distinguished by their masses m_1 and m_2. The total number of molecules is d, a fraction f_1 comprised of molecules of mass m_1 and and a fraction $f_2 = 1 - f_1$ of molecules of mass m_2. The box is thermally insulated so that the total energy E is kept constant. Let E_1 and E_2 be the total (kinetic) energies of the gas components of type 1 and 2, respectively, so that $E_1 + E_2 = E$. We write, for $i = 1, 2$:

$$E_i = \frac{3}{2} kT_i f_i d.$$

Also introduce the parameter $T = f_1 T_1 + f_2 T_2$. This is chosen so that $E = (3/2)kTd$.

The (velocity part of the) phase space for the gas can be described by the ellipsoid

$$\{(v, w) \in \mathbb{R}^{3d_1} \times \mathbb{R}^{3d_2} \mid \frac{1}{2} m_1 |v|^2 + \frac{1}{2} m_2 |w|^2 = E\}.$$

As before, we assume that the distribution of velocities is uniform over the ellipsoid, that is, the probability P is proportional to the area measure and we ask for the probability distribution of T_1. A simple calculation (that uses the observation that the subset of the phase space corresponding to a given value of T_1 is the cartesian product of two spheres of radii $(2E_i/m_i)^{1/2}$, for $i = 1, 2$) shows that for large values of d (the total number of particles) the probability of $x := f_1 T_1/T$, over the interval $[0, 1]$, has density

$$c[x^{f_1}(1-x)^{f_2}]^d.$$

Again for large values of d, this density has a sharp maximum at the point $x = f_1/(f_1 + f_2)$, which corresponds to the value

$$T_1 = T = T_2.$$

The conclusion is that, under the uniformity assumption made about P, the part of phase space most likely to be occupied by the gas corresponds to that for which the temperature parameters in both components of the gas mixture coincide.

This argument suggests that the probability density for the velocity of a particle of mass m immersed in a monoatomic gas at temperature T will approach in time the equilibrium value given by the Maxwell-Boltzmann distribution with mass m and temperature T.

We would like now to take a closer look at how the microscopic interactions between a particle of mass m suspended on a gas at temperature T of monoatomic molecules of much smaller mass, can lead to the probability density given in the previous paragraph. We call the larger particle a *Brownian particle*.

The Brownian particle is assumed to be under two kinds of forces. One is a frictional force due to viscosity, given by $-m\beta v$, where β is a constant and v is the particle's velocity. The second force is due to the combined effect of individual collisions with the surrounding molecules and has a highly fluctuating and chaotic behavior. We denote it by $mf(t)$. By Newton's second law of motion, $v(t)$ is described by the differential equation

$$\frac{dv}{dt} = -\beta v + f(t).$$

We regard $f(t)$ as a random variable, and expect that a solution of this equation (if one exists in some appropriate sense) will also be a random variable. If $\rho_{v_0}(v,t)$ denotes the density of the probability distribution of $v(t)$, (conditioned by $v(0) = v_0$) we expect, given the earlier discussion, that as t grows,

$$\rho_{v_0}(v,t) \to (\frac{m}{2\pi kT})^{3/2} e^{-\frac{m|v|^2}{2kT}}$$

and as t approaches 0, $\rho_{v_0}(v,t)$ approaches the Dirac delta function concentrated at v_0. The above equation is called the *Langevin equation*. Using the equilibrium value of the velocity distribution, we would like to determine the statistical properties of f.

The formal solution of the Langevin equation is

$$v(t) = v_0 e^{-\beta t} + e^{-\beta t} \int_0^t e^{\beta s} f(s)\, ds.$$

As the first term on the right-hand side of the equation goes to 0, the probability density of $\int_0^t e^{-\beta(t-s)} f(s)\, ds$ should approach, for large t, the Maxwell-Boltzmann equilibrium density of v.

Consider the Riemann sum approximation:

$$\int_0^t e^{-\beta(t-s)} f(s)\, ds \approx e^{-\beta t} \sum_i e^{\beta i \Delta t} f(i\Delta t)\, \Delta t.$$

The random variables $\Delta b_i := f(i\Delta t)\, \Delta t$ express the accelerations that the Brownian particle gains during the interval $[i\Delta t, (i+1)\Delta t]$. For large t, we have:

$$v \approx \sum_i e^{\beta i \Delta t - t} \Delta b_i.$$

Δb_i is assumed to be the sum of the accelerations due to a large number of independent collisions with the surrounding molecules, taking place during the interval $[i\Delta t, (i+1)\Delta t]$. It is thus natural to suppose that the Δb_i are independent random variables with the same kind of probability distribution as for the molecular velocities, that is, the Δb_i are assumed to be independent equally distributed normal random variables of 0 mean. To determine the variance $V(\Delta t)$ of Δb_i, we use that the limit (Maxwell-Boltzmann) distribution has variance $\frac{kT}{m}$, so that

$$E\left[|v|^2\right] \to \frac{kT}{m}.$$

On the other hand (using that the Δb_i are independent),

$$\begin{aligned} E\left[|v|^2\right] &\approx E\left[\sum_{i,j} e^{\beta(i\Delta t - t) + \beta(j\Delta t - t)} \langle \Delta b_i, \Delta b_j \rangle\right] \\ &= \sum_{i,j} e^{\beta(i\Delta t - t) + \beta(j\Delta t - t)} E[\langle \Delta b_i, \Delta b_j \rangle] \\ &= \sum_i e^{2\beta(i\Delta t - t)} E\left[|\Delta b_i|^2\right] \\ &= \sum_i e^{2\beta(i\Delta t - t)} V(\Delta t). \end{aligned}$$

In order for the Riemann sums to converge as $\Delta t \to 0$ it is now apparent that we must require $V(\Delta t)/\Delta t$ to have a nonzero finite limit. We call the limit σ^2. Therefore,

$$E\left[|v|^2\right] \approx \sum_i e^{2\beta(i\Delta t - t)} \sigma^2 \Delta t \approx \sigma^2 \int_0^t e^{-2\beta(t-s)}\, dt = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})$$

and we obtain $\sigma^2 = 2\beta kT/m$.

The conclusion is that the "force" f that accounts for the Maxwell-Boltzmann distribution of the Brownian particle is expected to be $f(t)\Delta t = \Delta b$, where Δb is a normal random variable of mean 0 and variance $2\beta kT/m$.

19.4. **Random Walk on the Real Line.** Let $(\Omega, \mathcal{F}, P)$ denote the coin-tossing probability space. (In particular, $\Omega = \{0,1\}^{\mathbb{N}}$.) Rather than work with the coordinates x_i as before, it will be more convenient here to use $\pi_i : \Omega \to \{-1,1\}$, such that $\pi_i(\omega) = 1$ if $x_i(\omega) = 0$ and $\pi_i(\omega) = -1$ if $x_i(\omega) = 1$.

Fix $n \in \mathbb{N}$ and define a motion on the real line with velocity given by the following random function of t:

$$v^{(n)}(t) := \sqrt{n}\pi_{[nt]+1} : \Omega \to \{-\sqrt{n}, \sqrt{n}\}.$$

In other words, $v^{(n)}(t) = \sqrt{n}\pi_k$ when t is in the interval $[(k-1)/n, k/n)$. The motion on $\mathbb{R}$ is obtained by integrating the velocity process. It is described by the random process $x^{(n)}(t)$, $t \geq 0$, such that $x^{(0)}$ will be chosen to be 0 and

$$x^{(n)}(t) = \frac{1}{\sqrt{n}} \left\{ \pi_1 + \pi_2 + \cdots + \pi_{[nt]} + (nt - [nt])\pi_{[nt]+1} \right\}.$$

Note that

$$x^{(n)}(\frac{m}{n}) = \frac{1}{\sqrt{n}} \left\{ \pi_1 + \pi_2 + \cdots + \pi_m \right\} = \sqrt{\frac{m}{n}} \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \pi_i.$$

We now fix t and consider what happens to $x^{(n)}(t)$ as $n \to \infty$ and $\frac{m}{n} \to t$. The *central limit theorem* immediately implies that the probability distribution of $x^{(n)}(t)$ converges to a Gaussian probability distribution. Thus we have:

Proposition 86. *As $n \to \infty$,*

$$P(a \leq x^{(n)}(t) \leq b) \to \int_{a/\sqrt{t}}^{b/\sqrt{t}} \frac{e^{\frac{x^2}{2}}}{\sqrt{2\pi}} dx = \int_a^b \frac{e^{\frac{x^2}{2t}}}{\sqrt{2\pi t}} dx.$$

Proof. This is a consequence of the central limit theorem, which states that

$$P(a \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \pi_i \leq b) \to \int_a^b \frac{e^{\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

as $n \to \infty$. □

19.5. **Khinchine's Proof of the Central Limit Theorem.** We sketch in this section a simple proof of the central limit theorem due to Khinchine, from 1933.

Consider the function

$$u(x,t) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} e^{-\frac{(x-y)^2}{2t}} f(y) dy.$$

It is a simple exercise to verify that u is a solution of the heat equation:

$$u_t = \frac{1}{2} u_{xx},$$

for $t > 0$ and $x \in \mathbb{R}$, and satisfying the initial condition $u(x,0) = f(x)$.

Define

$$x^{(n)}(\frac{m}{n}) = \frac{1}{\sqrt{n}} \left\{ \pi_1 + \pi_2 + \cdots + \pi_m \right\}.$$

Also define, for $t = m/n$, the functions

$$u^{(n)}\left(\frac{k}{\sqrt{n}},t\right) = E[f(x^{(n)}(t))|x^{(n)}(0) = \frac{k}{\sqrt{n}}].$$

Then $u^{(n)}$ satisfies the equation

$$u^{(n)}\left(\frac{k}{\sqrt{n}},\frac{m}{n}\right) = \sum_{j=-\infty}^{+\infty} p(k,j;m)\, f(\frac{j}{\sqrt{n}}),$$

where $p(k,j;m)$ is the probability that, having started at $k/\sqrt{n}$ at time 0, the particle describing a random walk along the line will find itself at $j/\sqrt{n}$ at time m/n. Thus

$$p(k,j;m) = P(\pi_1 + \cdots + \pi_m = j-k).$$

Exercise 87. *Show that*

$$p(k,j;m+1) = \frac{1}{2}[p(k+1,j;m) + p(k-1,j;m)].$$

A different way to express the equation in the previous exercise is to write

$$p(k,j;m+1) - p(k,j;m) = \frac{1}{2}[p(k+1,j;m) - 2p(k,j;m) + p(k-1,j;m)].$$

Notice how this looks like a discretized form of the heat equation. In fact, the proof will proceed by comparing solutions of the continuous and discrete equations. We first set some notation to facilitate this comparison. Write $\delta = 1/\sqrt{n}$, $\tau = 1/n$, and define $u^{(n)}_{xx} = \Delta^{(n)}u^{(n)}$, where

$$\Delta^{(n)}h\,(k\delta, m\tau) = \frac{h((k+1)\delta, m\tau) - 2h(k\delta, m\tau) + h((k-1)\delta, m\tau)}{\delta^2},$$

which is a discretized second derivative in x. The first derivative in t has the following discrete form:

$$u^{(n)}_t(k\delta, m\tau) = \frac{u^{(n)}(k\delta,(m+1)\tau) - u^{(n)}(k\delta, m\tau)}{\tau}.$$

We now have:

$$\begin{aligned}
\tau u^{(n)}_t(k\delta, m\tau) &= \sum_{j=-\infty}^{+\infty} [p(k,j;m+1) - p(k,j;m)]\, f(j\delta) \\
&= \sum_{j=-\infty}^{+\infty} \frac{1}{2}[p(k+1,j;m) - 2p(k,j;m) + p(k-1,j;m)]\, f(j\delta) \\
&= \frac{1}{2}\delta^2 u^{(n)}_{xx}(k\delta, m\tau).
\end{aligned}$$

Since $\delta^2 = \tau$, we have

$$u^{(n)}_t(k\delta, m\tau) = \frac{1}{2}u^{(n)}_{xx}(k\delta, m\tau).$$

Our goal is to show that

$$u^{(n)}(\frac{k}{\sqrt{n}}, \frac{m}{n}) \to u(x,t)$$

as $n \to \infty$, $m/n \to t$, and $k/\sqrt{n} \to x$.

Since u satisfies the heat equation, then

$$\frac{u(k\delta, \frac{m}{n} + \tau) - u(k\delta, \frac{m}{n})}{\tau} = \Delta^{(n)}u(k\delta, \frac{m}{n}) + o(1).$$

Here, $o(1) \to 0$ uniformly in k and m as $n \to \infty$.

After some algebraic simplification, we write the discretized equations in a recursive form:

$$u(k\delta, \frac{m}{n} + \tau) = \frac{1}{2}\left[u((k+1)\delta, \frac{m}{n}) + u((k-1)\delta, \frac{m}{n})\right] + \tau o(1)$$
$$u^{(n)}(k\delta, \frac{m}{n} + \tau) = \frac{1}{2}\left[u^{(n)}((k+1)\delta, \frac{m}{n}) + u^{(n)}((k-1)\delta, \frac{m}{n})\right].$$

Using the initial condition $u^{(n)}(k\delta, 0) = f(k\delta) = u(k\delta, 0)$ and the above recursive formulas we finally get

$$u^{(n)}(k\delta, \frac{m}{n}) - u(k\delta, \frac{m}{n}) = \frac{m}{n} o(1).$$

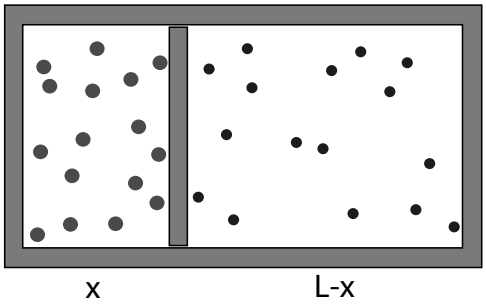
We can now pass to the limit to conclude the proof.

19.6. **A Geometric Interpretation of the Central Limit Theorem.** Let $\mathcal{C}$ be the subset of $\mathbb{R}^n$ that consists of all the vertex points of the n-dimensional unit cube $[0,1]^n$. Note that $\mathcal{C}$ has 2^n elements. Let V be the vertex that is farthest away from the origin (which is also a vertex), and C the center of the cube. The center of the cube corresponds to the vector $c = \frac{1}{2}\overrightarrow{OV}$. Define the unit vector $u = \frac{\overrightarrow{OV}}{|\overrightarrow{OV}|}$. Also define the map $\pi : \mathbb{R}^n \to \mathbb{R}$ such that $\pi(x) = (x - c) \cdot u$ (the ordinary dot product of the two vectors). This is the orthogonal projection of the translate $x - c$ along the direction from O to V.

Let μ_n be the measure on $\mathcal{C}$ that assigns mass $1/2^n$ to each vertex and let $P_n := \pi_* \mu_n$. Then, as $n \to \infty$, P_n converges to the standard Gaussian measure on the real line.

20. More on Ideal Gases

We have used before the following fact: one mole of any gas at standard temperature and pressure occupies the same volume, 22.4 liters. To understand why this is so, consider the following thought experiment.



A closed chamber has two compartments separated by a moving wall and each compartment contains gases of possibly different masses, M and m. We suppose that the number of molecules of each kind is the same and that the moving wall can transmit heat. After equilibrium the temperature and pressure in the two compartments are the same. We wish to show that the moving wall will be at the center of the chamber, that is, $x = L - x$.

Let $\frac{1}{2}M\langle|\mathbf{V}|^2\rangle$ and $\frac{1}{2}m\langle|\mathbf{V}|^2\rangle$ be the mean kinetic energies of molecules on each compartment. Since the temperatures are the same, we have equality

$$\frac{1}{2}M\langle|\mathbf{V}|^2\rangle = \frac{1}{2}m\langle|\mathbf{V}|^2\rangle.$$

The mean value of $|V_x|^2$ is $\frac{1}{3}|\mathbf{V}|^2$, so the mean time that it takes a particle of mass M to traverse the length of its compartment, x, is $\sqrt{3}x/\langle|\mathbf{V}|^2\rangle^{\frac{1}{2}}$. During an interval of time Δt, the average number of times a particle of mass M will hit the middle wall is $\Delta t/(2\sqrt{3}x/\langle|\mathbf{V}|^2\rangle^{\frac{1}{2}})$. So the mean momentum tranfered to the wall by the gas of mass M during that interval of time is

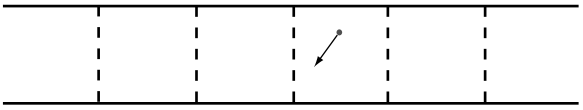
$$\frac{n\Delta t\langle|\mathbf{V}|^2\rangle^{\frac{1}{2}}}{2\sqrt{3}x}M\langle|\mathbf{V}|^2\rangle^{\frac{1}{2}}.$$

That momentum must be the same on both sides of the wall since the pressures are equal. Therefore

$$\frac{n\Delta t}{\sqrt{3}}\frac{M\langle|\mathbf{V}|^2\rangle}{2}\frac{1}{x}=\frac{n\Delta t}{\sqrt{3}}\frac{m\langle|\mathbf{v}|^2\rangle}{2}\frac{1}{L-x}.$$

Since the mean kinetic energies are the same, this equation simplifies to $x=L-x$ as claimed.

21. Billiard Models of Gas Diffusion



22. Reaction and Diffusion