Chapter 11

Systems of Differential Equations

Contents

Linear systems. A **linear system** is a system of differential equations of the form

$$
(1)\qquad \begin{array}{rcccccccc}
x_1' & = & a_{11}x_1 & + & \cdots & + & a_{1n}x_n & + & f_1, \\
x_2' & = & a_{21}x_1 & + & \cdots & + & a_{2n}x_n & + & f_2, \\
\vdots & & & \vdots & \cdots & \vdots & & \vdots & \\
x_m' & = & a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & + & f_m,
\end{array}
$$

where $' = d/dt$. Given are the functions $a_{ij}(t)$ and $f_j(t)$ on some interval $a < t < b$. The unknowns are the functions $x_1(t), \ldots, x_n(t)$.

The system is called **homogeneous** if all $f_j = 0$, otherwise it is called **non-homogeneous**.

Matrix Notation for Systems. A non-homogeneous system of linear equations (1) is written as the equivalent vector-matrix system

$$\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t),$$

where

$$\vec{\mathbf{x}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \vec{\mathbf{f}} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \dots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

11.1 Examples of Systems

Brine Tank Cascade

Let brine tanks A, B, C be given of volumes 20, 40, 60, respectively, as in Figure 1.

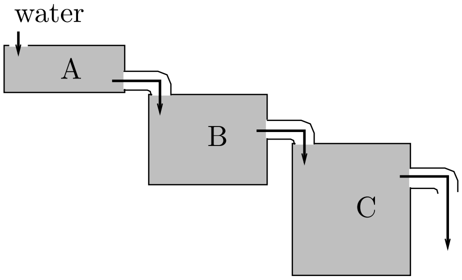

Figure 1. Three brine tanks in cascade.

It is supposed that fluid enters tank A at rate r, drains from A to B at rate r, drains from B to C at rate r, then drains from tank C at rate r. Hence the volumes of the tanks remain constant. Let $r = 10$, to illustrate the ideas.

Uniform stirring of each tank is assumed, which implies **uniform salt concentration** throughout each tank.

Let $x_1(t)$, $x_2(t)$, $x_3(t)$ denote the amount of salt at time t in each tank. We suppose **water containing no salt** is added to tank A . Therefore, the salt in all the tanks is eventually lost from the drains. The cascade is modeled by the **chemical balance law**

$$\text{rate of change} \quad = \quad \text{input rate} \quad - \quad \text{output rate}.$$

Application of the balance law, justified below in *compartment analysis*, results in the triangular differential system

$$\begin{aligned} x_1' &= -\frac{1}{2}x_1, \\ x_2' &= \frac{1}{2}x_1 - \frac{1}{4}x_2, \\ x_3' &= \frac{1}{4}x_2 - \frac{1}{6}x_3. \end{aligned}$$

The solution, to be justified later in this chapter, is given by the equations

$$\begin{aligned} x_1(t) &= x_1(0)e^{-t/2}, \\ x_2(t) &= -2x_1(0)e^{-t/2} + (x_2(0) + 2x_1(0))e^{-t/4}, \\ x_3(t) &= \frac{3}{2}\,x_1(0)e^{-t/2} - 3(x_2(0) + 2x_1(0))e^{-t/4} \\ &\quad + (x_3(0) - \frac{3}{2}x_1(0) + 3(x_2(0) + 2x_1(0)))e^{-t/6}. \end{aligned}$$

Cascades and Compartment Analysis

A **linear cascade** is a diagram of **compartments** in which input and output rates have been assigned from one or more different compartments. The diagram is a succinct way to summarize and document the various rates.

The method of **compartment analysis** translates the diagram into a system of linear differential equations. The method has been used to derive applied models in diverse topics like ecology, chemistry, heating and cooling, kinetics, mechanics and electricity.

The method. Refer to Figure 2. A compartment diagram consists of the following components.

Variable Names	Each **compartment** is labelled with a variable X.

Arrows Each arrow is labelled with a **flow rate** R.

Input Rate An arrowhead pointing at compartment X documents **input rate** R.

Output Rate An arrowhead pointing away from compartment X documents **output rate** R.

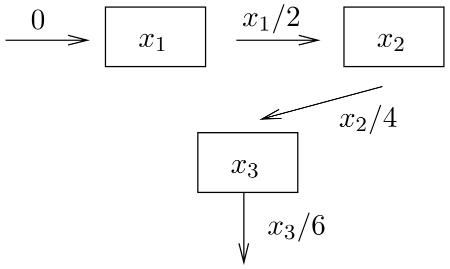

Figure 2. Compartment analysis diagram.
The diagram represents the classical brine tank problem of Figure 1.

Assembly of the single linear differential equation for a diagram compartment X is done by writing dX/dt for the left side of the differential equation and then algebraically adding the input and output rates to obtain the right side of the differential equation, according to the **balance law**

$$\frac{dX}{dt} = \text{sum of input rates} - \text{sum of output rates}$$

By convention, a compartment with no arriving arrowhead has input zero, and a compartment with no exiting arrowhead has output zero. Applying the balance law to Figure 2 gives one differential equation for each of the three compartments $\boxed{x_1}$, $\boxed{x_2}$, $\boxed{x_3}$.

$$\begin{aligned} x_1' &= 0 - \frac{1}{2}x_1, \\ x_2' &= \frac{1}{2}x_1 - \frac{1}{4}x_2, \\ x_3' &= \frac{1}{4}x_2 - \frac{1}{6}x_3. \end{aligned}$$

Recycled Brine Tank Cascade

Let brine tanks A, B, C be given of volumes 60, 30, 60, respectively, as in Figure 3.

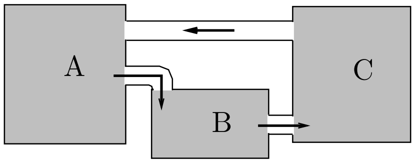

Figure 3. Three brine tanks in cascade with recycling.

Suppose that fluid drains from tank A to B at rate r, drains from tank B to C at rate r, then drains from tank C to A at rate r. The tank volumes remain constant due to constant recycling of fluid. For purposes of illustration, let $r = 10$.

Uniform stirring of each tank is assumed, which implies **uniform salt concentration** throughout each tank.

Let $x_1(t)$, $x_2(t)$, $x_3(t)$ denote the amount of salt at time t in each tank. No salt is lost from the system, due to recycling. Using compartment analysis, the recycled cascade is modeled by the non-triangular system

$$\begin{aligned}
x_1' &= -\frac{1}{6}x_1 && + \frac{1}{6}x_3, \\
x_2' &= \frac{1}{6}x_1 - \frac{1}{3}x_2, && \\
x_3' &= \frac{1}{3}x_2 && - \frac{1}{6}x_3.
\end{aligned}$$

The solution is given by the equations

$$\begin{aligned}
x_1(t) &= c_1 + (c_2 - 2c_3)e^{-t/3}\cos(t/6) + (2c_2 + c_3)e^{-t/3}\sin(t/6), \\
x_2(t) &= \frac{1}{2}c_1 + (-2c_2 - c_3)e^{-t/3}\cos(t/6) + (c_2 - 2c_3)e^{-t/3}\sin(t/6), \\
x_3(t) &= c_1 + (c_2 + 3c_3)e^{-t/3}\cos(t/6) + (-3c_2 + c_3)e^{-t/3}\sin(t/6).
\end{aligned}$$

At infinity, $x_1 = x_3 = c_1$, $x_2 = c_1/2$. The meaning is that the total amount of salt is uniformly distributed in the tanks, in the ratio 2 : 1 : 2.

Pond Pollution

Consider three ponds connected by streams, as in Figure 4. The first pond has a pollution source, which spreads via the connecting streams to the other ponds. The plan is to determine the amount of pollutant in each pond.

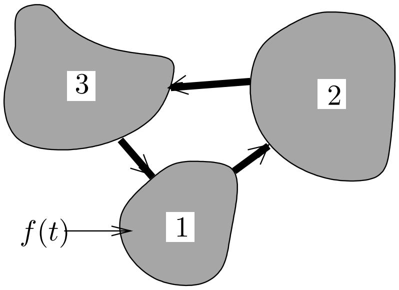

Figure 4. Three ponds 1, 2, 3 of volumes V_1, V_2, V_3 connected by streams. The pollution source $f(t)$ is in pond 1.

Assume the following.

- Symbol $f(t)$ is the pollutant flow rate into pond 1 (lb/min).
- Symbols f_1, f_2, f_3 denote the pollutant flow rates out of ponds 1, 2, 3, respectively (gal/min). It is assumed that the pollutant is well-mixed in each pond.

- The three ponds have volumes V_1, V_2, V_3 (gal), which remain constant.

- Symbols $x_1(t)$, $x_2(t)$, $x_3(t)$ denote the amount (lbs) of pollutant in ponds 1, 2, 3, respectively.

The pollutant flux is the flow rate times the pollutant concentration, e.g., pond 1 is emptied with flux f_1 times $x_1(t)/V_1$. A compartment analysis is summarized in the following diagram.

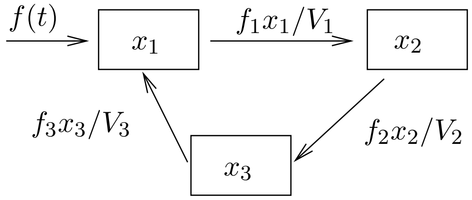

Figure 5. Pond diagram. The compartment diagram represents the three-pond pollution problem of Figure 4.

The diagram plus compartment analysis gives the following differential equations.

$$\begin{aligned}
x_1'(t) &= \frac{f_3}{V_3}\,x_3(t) - \frac{f_1}{V_1}\,x_1(t) + f(t),\\
x_2'(t) &= \frac{f_1}{V_1}\,x_1(t) - \frac{f_2}{V_2}\,x_2(t),\\
x_3'(t) &= \frac{f_2}{V_2}\,x_2(t) - \frac{f_3}{V_3}\,x_3(t).
\end{aligned}$$

For a specific numerical example, take $f_i/V_i = 0.001$, $1 \le i \le 3$, and let $f(t) = 0.125$ lb/min for the first 48 hours (2880 minutes), thereafter $f(t) = 0$. We expect due to uniform mixing that after a long time there will be $(0.125)(2880) = 360$ pounds of pollutant uniformly deposited, which is 120 pounds per pond.

Initially, $x_1(0) = x_2(0) = x_3(0) = 0$, if the ponds were pristine. The specialized problem for the first 48 hours is

$$\begin{aligned}
x_1'(t) &= 0.001\,x_3(t) - 0.001\,x_1(t) + 0.125,\\
x_2'(t) &= 0.001\,x_1(t) - 0.001\,x_2(t),\\
x_3'(t) &= 0.001\,x_2(t) - 0.001\,x_3(t),\\
&x_1(0) = x_2(0) = x_3(0) = 0.
\end{aligned}$$

The solution to this system is

$$\begin{aligned}
x_1(t) &= e^{-\frac{3t}{2000}}\left(\frac{125\sqrt{3}}{9}\sin\left(\frac{\sqrt{3}t}{2000}\right) - \frac{125}{3}\cos\left(\frac{\sqrt{3}t}{2000}\right)\right) + \frac{125}{3} + \frac{t}{24},\\
x_2(t) &= -\frac{250\sqrt{3}}{9}e^{-\frac{3t}{2000}}\sin\left(\frac{\sqrt{3}t}{2000}\right) + \frac{t}{24},\\
x_3(t) &= e^{-\frac{3t}{2000}}\left(\frac{125}{3}\cos\left(\frac{\sqrt{3}t}{2000}\right) + \frac{125\sqrt{3}}{9}\sin\left(\frac{\sqrt{3}t}{2000}\right)\right) + \frac{t}{24} - \frac{125}{3}.
\end{aligned}$$

After 48 hours elapse, the approximate pollutant amounts in pounds are

$$x_1(2880) = 162.30, \quad x_2(2880) = 119.61, \quad x_3(2880) = 78.08.$$

It should be remarked that the system above is altered by replacing 0.125 by zero, in order to predict the state of the ponds after 48 hours. The corresponding homogeneous system has an equilibrium solution $x_1(t) = x_2(t) = x_3(t) = 120$. This constant solution is the limit at infinity of the solution to the homogeneous system, using the initial values $x_1(0) \approx 162.30$, $x_2(0) \approx 119.61$, $x_3(0) \approx 78.08$.

Home Heating

Consider a typical home with attic, basement and insulated main floor.

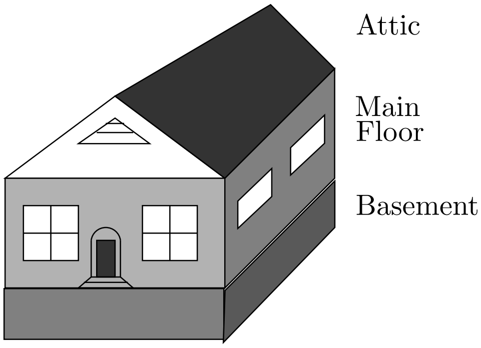

Figure 6. Typical home with attic and basement. The below-grade basement and the attic are un-insulated. Only the main living area is insulated.

It is usual to surround the main living area with insulation, but the attic area has walls and ceiling without insulation. The walls and floor in the basement are insulated by earth. The basement ceiling is insulated by air space in the joists, a layer of flooring on the main floor and a layer of drywall in the basement. We will analyze the changing temperatures in the three levels using Newton's cooling law and the variables

$$\begin{aligned} z(t) &= \text{Temperature in the attic,} \\ y(t) &= \text{Temperature in the main living area,} \\ x(t) &= \text{Temperature in the basement,} \\ t &= \text{Time in hours.} \end{aligned}$$

Initial data. Assume it is winter time and the outside temperature in constantly 35°F during the day. Also assumed is a basement earth temperature of 45°F. Initially, the heat is off for several days. The initial values at noon ($t = 0$) are then $x(0) = 45$, $y(0) = z(0) = 35$.

Portable heater. A small electric heater is turned on at noon, with thermostat set for 100°F. When the heater is running, it provides a 20°F rise per hour, therefore it takes some time to reach 100°F (probably never!). Newton's cooling law

$$\text{Temperature rate} = \text{k(Temperature difference)}$$

will be applied to five boundary surfaces: (0) the basement walls and floor, (1) the basement ceiling, (2) the main floor walls, (3) the main floor ceiling, and (4) the attic walls and ceiling. Newton's cooling law gives positive cooling constants k_0, k_1, k_2, k_3, k_4 and the equations

$$\begin{aligned} x' &= k_0(45 - x) + k_1(y - x), \\ y' &= k_1(x - y) + k_2(35 - y) + k_3(z - y) + 20, \\ z' &= k_3(y - z) + k_4(35 - z). \end{aligned}$$

The insulation constants will be defined as $k_0 = 1/2$, $k_1 = 1/2$, $k_2 = 1/4$, $k_3 = 1/4$, $k_4 = 3/4$ to reflect insulation quality. The reciprocal $1/k$ is approximately the amount of time in hours required for 63% of the temperature difference to be exchanged. For instance, 4 hours elapse for the main floor. The model:

$$\begin{aligned} x' &= \frac{1}{2}(45 - x) + \frac{1}{2}(y - x), \\ y' &= \frac{1}{2}(x - y) + \frac{1}{4}(35 - y) + \frac{1}{4}(z - y) + 20, \\ z' &= \frac{1}{4}(y - z) + \frac{3}{4}(35 - z). \end{aligned}$$

The homogeneous solution in vector form is given in terms of constants $a = 1 + \sqrt{5}/4$, $b = 1 - \sqrt{5}/4$, and arbitrary constants c_1, c_2, c_3 by the formula

$$\begin{pmatrix} x_h(t) \\ y_h(t) \\ z_h(t) \end{pmatrix} = c_1 e^{-t} \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix} + c_2 e^{-at} \begin{pmatrix} 2 \\ \sqrt{5} \\ 1 \end{pmatrix} + c_3 e^{-bt} \begin{pmatrix} 2 \\ -\sqrt{5} \\ 1 \end{pmatrix}.$$

A particular solution is an equilibrium solution

$$\begin{pmatrix} x_p(t) \\ y_p(t) \\ z_p(t) \end{pmatrix} = \begin{pmatrix} \frac{620}{11} \\ \frac{745}{11} \\ \frac{475}{11} \end{pmatrix}.$$

The homogeneous solution has limit zero at infinity, hence the temperatures of the three spaces hover around $x = 56.4$, $y = 67.7$, $z = 43.2$ degrees Fahrenheit. Specific information can be gathered by solving for c_1, c_2, c_3 according to the initial data $x(0) = 45$, $y(0) = z(0) = 35$. The answers are

$$c_1 = 5, \quad c_2 = \frac{25}{2} + \frac{7}{2}\sqrt{5}, \quad c_3 = \frac{25}{2} - \frac{7}{2}\sqrt{5}.$$

Underpowered heater. To the main floor each hour is added 20°F, but the heat escapes at a substantial rate, so that after one hour $y \approx 68$°F.

After five hours, $y \approx 68$°F. The heater in this example is so inadequate that even after many hours, the main living area is still under 69°F.

Forced air furnace. Replacing the space heater by a normal furnace adds the difficulty of **switches** in the input, namely, the thermostat turns off the furnace when the main floor temperature reaches 70°F, and it turns it on again after a 4°F temperature drop. We will suppose that the furnace has four times the BTU rating of the space heater, which translates to an 80°F temperature rise per hour. The study of the forced air furnace requires two differential equations, one with 20 replaced by 80 (DE 1, furnace on) and the other with 20 replaced by 0 (DE 2, furnace off). The plan is to use the first differential equation on time interval $0 \le t \le t_1$, then switch to the second differential equation for time interval $t_1 \le t \le t_2$. The time intervals are selected so that $y(t_1) = 70$ (the thermostat setting) and $y(t_2) = 66$ (thermostat setting less 4 degrees). Numerical work gives the following results.

Time in minutes	Main floor temperature	Model	Furnace
31.6	70	DE 1	on
40.9	66	DE 2	off
45.3	70	DE 1	on
54.6	66	DE 2	off

The reason for the non-uniform times between furnace cycles can be seen from the model. Each time the furnace cycles, heat enters the main floor, then escapes through the other two levels. Consequently, the initial conditions on each floor applied to models 1 and 2 are changing, resulting in different solutions to the models on each switch.

Chemostats and Microorganism Culturing

A vessel into which nutrients are pumped, to feed a microorganism, is called a **chemostat**[1]. Uniform distributions of microorganisms and nutrients are assumed, for example, due to stirring effects. The pumping is matched by draining to keep the volume constant.

[1] The October 14, 2004 issue of the journal *Nature* featured a study of the co-evolution of a common type of bacteria, Escherichia coli, and a virus that infects it, called bacteriophage T7. Postdoctoral researcher Samantha Forde set up "microbial communities of bacteria and viruses with different nutrient levels in a series of chemostats – glass culture tubes that provide nutrients and oxygen and siphon off wastes."

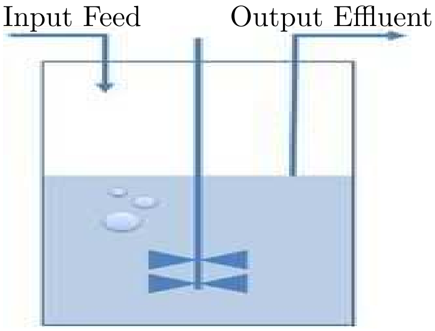

Figure 7. A Basic Chemostat. A stirred bio-reactor operated as a chemostat, with continuous inflow and outflow. The flow rates are controlled to maintain a constant culture volume.

In a typical chemostat, one nutrient is kept in short supply while all others are abundant. We consider here the question of **survival** of the organism subject to the limited resource. The problem is quantified as follows:

$x(t) =$ the concentration of the limited nutrient in the vessel,

$y(t) =$ the concentration of organisms in the vessel.

A special case of the derivation in J.M. Cushing's text for the organism *E. Coli*[2] is the set of **nonlinear** differential equations[3]

$$\begin{aligned} x' &= -0.075x + (0.075)(0.005) - \frac{1}{63}g(x)y, \\ y' &= -0.075y + g(x)y, \end{aligned} \tag{2}$$

where $g(x) = 0.68x(0.0016 + x)^{-1}$. Of special interest to the study of this equation are two linearized equations at equilibria, given by

$$\begin{aligned} u_1' &= -0.075\,u_1 - 0.008177008175\,u_2, \\ u_2' &= 0.4401515152\,u_2, \end{aligned} \tag{3}$$

$$\begin{aligned} v_1' &= -1.690372243\,v_1 - 0.001190476190\,v_2, \\ v_2' &= 101.7684513\,v_1. \end{aligned} \tag{4}$$

Although we cannot solve the nonlinear system explicitly, nevertheless there are explicit formulas for u_1, u_2, v_1, v_2 that complete the picture of how solutions $x(t)$, $y(t)$ behave at $t = \infty$. The result of the analysis is that *E. Coli* survives indefinitely in this vessel at concentration $y \approx 0.3$.

[2] In a biology Master's thesis, two strains of Escherichia coli were grown in a glucose-limited chemostat coupled to a modified Robbins device containing plugs of silicone rubber urinary catheter material. Reference: Jennifer L. Adams and Robert J. C. McLean, Applied and Environmental Microbiology, September 1999, p. 4285-4287, Vol. 65, No. 9.

[3] More details can be found in *The Theory of the Chemostat Dynamics of Microbial Competition*, ISBN-13: 9780521067348, by Hal Smith and Paul Waltman, June 2008.

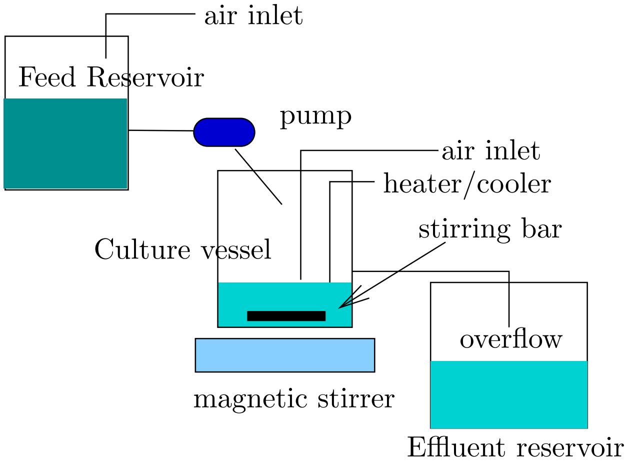

Figure 8. Laboratory Chemostat.
The components are the **Feed reservoir**, which contains the nutrients, a stirred chemical reactor labeled the **Culture vessel**, and the **Effluent reservoir**, which holds the effluent overflow from the reactor.

Irregular Heartbeats and Lidocaine

The human malady of **ventricular arrhythmia** or irregular heartbeat is treated clinically using the drug **lidocaine**.

Figure 9. Xylocaine label, a brand name for the drug lidocaine.

To be effective, the drug has to be maintained at a bloodstream concentration of 1.5 milligrams per liter, but concentrations above 6 milligrams per liter are considered lethal in some patients. The actual dosage depends upon body weight. The adult dosage maximum for ventricular tachycardia is reported at 3 mg/kg.[4] The drug is supplied in 0.5%, 1% and 2% solutions, which are stored at room temperature.

A differential equation model for the dynamics of the drug therapy uses

$$x(t) = \text{amount of } \textit{lidocaine} \text{ in the bloodstream,}$$
$$y(t) = \text{amount of } \textit{lidocaine} \text{ in body tissue.}$$

A typical set of equations, valid for a special body weight only, appears below; for more detail see J.M. Cushing's text [**?**].

$$\begin{aligned} x'(t) &= -0.09x(t) + 0.038y(t), \\ y'(t) &= 0.066x(t) - 0.038y(t). \end{aligned} \tag{5}$$

[4] Source: **Family Practice Notebook**, `http://www.fpnotebook.com/`. The author is Scott Moses, MD, who practises in Lino Lakes, Minnesota.

The physically significant initial data is zero drug in the bloodstream $x(0)=0$ and injection dosage $y(0)=y_0$. The answers:

$$\begin{aligned} x(t) &= -0.3367y_0e^{-0.1204t}+0.3367y_0e^{-0.0076t},\\ y(t) &= 0.2696y_0e^{-0.1204t}+0.7304y_0e^{-0.0076t}. \end{aligned}$$

The answers can be used to estimate the maximum possible safe dosage y_0 and the duration of time that the drug *lidocaine* is effective.

Nutrient Flow in an Aquarium

Consider a vessel of water containing a radioactive isotope, to be used as a tracer for the food chain, which consists of aquatic plankton varieties A and B.

Plankton are aquatic organisms that drift with the currents, typically in an environment like Chesapeake Bay. Plankton can be divided into two groups, phytoplankton and zooplankton. The phytoplankton are *plant-like* drifters: diatoms and other alga. Zooplankton are *animal-like* drifters: copepods, larvae, and small crustaceans.

Figure 10. Left: Bacillaria paxillifera, phytoplankton. Right: Anomura Galathea zoea, zooplankton.

Let

$$\begin{aligned} x(t) &= \text{isotope concentration in the water},\\ y(t) &= \text{isotope concentration in } A,\\ z(t) &= \text{isotope concentration in } B. \end{aligned}$$

Typical differential equations are

$$\begin{aligned} x'(t) &= -3x(t)+6y(t)+5z(t),\\ y'(t) &= 2x(t)-12y(t),\\ z'(t) &= x(t)+6y(t)-5z(t). \end{aligned}$$

The answers are

$$\begin{aligned} x(t) &= 6c_1+(1+\sqrt{6})c_2e^{(-10+\sqrt{6})t}+(1-\sqrt{6})c_3e^{(-10-\sqrt{6})t},\\ y(t) &= c_1+c_2e^{(-10+\sqrt{6})t}+c_3e^{(-10-\sqrt{6})t},\\ z(t) &= \frac{12}{5}c_1-\left(2+\sqrt{1.5}\right)c_2e^{(-10+\sqrt{6})t}+\left(-2+\sqrt{1.5}\right)c_3e^{(-10-\sqrt{6})t}. \end{aligned}$$

The constants c_1, c_2, c_3 are related to the initial radioactive isotope concentrations $x(0) = x_0$, $y(0) = 0$, $z(0) = 0$, by the 3×3 system of linear algebraic equations

$$\begin{array}{rcrcrcr} 6c_1 & + & (1+\sqrt{6})c_2 & + & (1-\sqrt{6})c_3 & = & x_0, \\ c_1 & + & c_2 & + & c_3 & = & 0, \\ \dfrac{12}{5}c_1 & - & \left(2+\sqrt{1.5}\right)c_2 & + & \left(-2+\sqrt{1.5}\right)c_3 & = & 0. \end{array}$$

Biomass Transfer

Consider a European forest having one or two varieties of trees. We select some of the oldest trees, those expected to die off in the next few years, then follow the cycle of living trees into dead trees. The dead trees eventually decay and fall from seasonal and biological events. Finally, the fallen trees become humus.

Figure 11. Forest Biomass. Total biomass is a parameter used to assess atmospheric carbon that is harvested by trees. Forest management uses biomass subclasses to classify fire risk.

Let variables x, y, z, t be defined by

$x(t) =$ biomass decayed into humus,

$y(t) =$ biomass of dead trees,

$z(t) =$ biomass of living trees,

$t =$ time in decades (*decade* $= 10$ years).

A typical biological model is

$$\begin{aligned} x'(t) &= -x(t) + 3y(t), \\ y'(t) &= -3y(t) + 5z(t), \\ z'(t) &= -5z(t). \end{aligned}$$

Suppose there are no dead trees and no humus at $t = 0$, with initially z_0 units of living tree biomass. These assumptions imply initial conditions $x(0) = y(0) = 0$, $z(0) = z_0$. The solution is

$$\begin{aligned} x(t) &= \frac{15}{8} z_0 \left(e^{-5t} - 2e^{-3t} + e^{-t} \right), \\ y(t) &= \frac{5}{2} z_0 \left(-e^{-5t} + e^{-3t} \right), \\ z(t) &= z_0 e^{-5t}. \end{aligned}$$

The live tree biomass $z(t) = z_0 e^{-5t}$ decreases according to a Malthusian decay law from its initial size z_0. It decays to 60% of its original biomass in one year. Interesting calculations that can be made from the other formulas include the future dates when the dead tree biomass and the humus biomass are maximum. The predicted dates are approximately 2.5 and 8 years hence, respectively.

The predictions made by this model are trends extrapolated from rate observations in the forest. Like weather prediction, it is a calculated guess that disappoints on a given day and from the outset has no predictable answer.

Total biomass is considered an important parameter to assess atmospheric carbon that is harvested by trees. Biomass estimates for forests since 1980 have been made by satellite remote sensing data with instances of 90% accuracy (*Science* 87(5), September 2004).

Pesticides in Soil and Trees

A Washington cherry orchard was sprayed with pesticides.

Figure 12. Cherries in June.

Assume that a negligible amount of pesticide was sprayed on the soil. Pesticide applied to the trees has a certain outflow rate to the soil, and conversely, pesticide in the soil has a certain uptake rate into the trees. Repeated applications of the pesticide are required to control the insects, which implies the pesticide levels in the trees varies with time. Quantize the pesticide spraying as follows.

$$
\begin{aligned}
x(t) &= \text{amount of pesticide in the trees},\\
y(t) &= \text{amount of pesticide in the soil},\\
r(t) &= \text{amount of pesticide applied to the trees},\\
t &= \text{time in years}.
\end{aligned}
$$

A typical model is obtained from input-output analysis, similar to the brine tank models:

$$
\begin{aligned}
x'(t) &= 2x(t) - y(t) + r(t),\\
y'(t) &= 2x(t) - 3y(t).
\end{aligned}
$$

In a pristine orchard, the initial data is $x(0) = 0$, $y(0) = 0$, because the trees and the soil initially harbor no pesticide. The solution of the model obviously depends on $r(t)$. The nonhomogeneous dependence is treated by the method of variation of parameters *infra*. Approximate formulas are

$$
\begin{aligned}
x(t) &\approx \int_0^t \left(1.10e^{1.6(t-u)} - 0.12e^{-2.6(t-u)}\right) r(u)du,\\
y(t) &\approx \int_0^t \left(0.49e^{1.6(t-u)} - 0.49e^{-2.6(t-u)}\right) r(u)du.
\end{aligned}
$$

The exponential rates 1.6 and -2.6 represent respectively the accumulation of the pesticide into the soil and the decay of the pesticide from the trees. The application rate $r(t)$ is typically a step function equal to a positive constant over a small interval of time and zero elsewhere, or a sum of such functions, representing periodic applications of pesticide.

Forecasting Prices

A cosmetics manufacturer has a marketing policy based upon the price $x(t)$ of its salon shampoo.

Figure 13. Salon shampoo sample. The marketing strategy for the shampoo is to set the price $x(t)$ dynamically to reflect demand for the product.

The **production** $P(t)$ and the **sales** $S(t)$ are given in terms of the **price** $x(t)$ and the **change in price** $x'(t)$ by the equations

$$\begin{aligned} P(t) &= 4 - \frac{3}{4}x(t) - 8x'(t) \quad \text{(Production)},\\ S(t) &= 15 - 4x(t) - 2x'(t) \quad \text{(Sales)}.\end{aligned}$$

The differential equations for the price $x(t)$ and inventory level $I(t)$ are

$$\begin{aligned} x'(t) &= k(I(t) - I_0),\\ I'(t) &= P(t) - S(t).\end{aligned}$$

The inventory level $I_0 = 50$ represents the desired level. The equations can be written in terms of $x(t)$, $I(t)$ as follows.

$$\begin{array}{rclclcl} x'(t) & = & & & kI(t) & - & kI_0,\\ I'(t) & = & \frac{13}{4}x(t) & - & 6kI(t) & + & 6kI_0 - 11.\end{array}$$

If $k = 1$, $x(0) = 10$ and $I(0) = 7$, then the solution is given by

$$\begin{aligned} x(t) &= \frac{44}{13} + \frac{86}{13}e^{-13t/2},\\ I(t) &= 50 - 43e^{-13t/2}.\end{aligned}$$

The **forecast** of price $x(t) \approx 3.39$ dollars at inventory level $I(t) \approx 50$ is based upon the two limits

$$\lim_{t\to\infty} x(t) = \frac{44}{13}, \quad \lim_{t\to\infty} I(t) = 50.$$

Coupled Spring-Mass Systems

Three masses are attached to each other by four springs as in Figure 14.

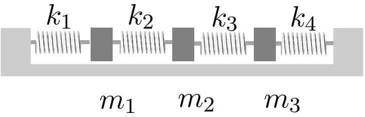

Figure 14. Three masses connected by springs. The masses slide along a frictionless horizontal surface.

The analysis uses the following constants, variables and assumptions.

Mass Constants The masses m_1, m_2, m_3 are assumed to be point masses concentrated at their center of gravity.

Spring Constants The mass of each spring is negligible. The springs operate according to Hooke's law: Force = k(elongation). Constants k_1, k_2, k_3, k_4 denote the Hooke's constants. The springs restore after compression and extension.

Position Variables The symbols $x_1(t)$, $x_2(t)$, $x_3(t)$ denote the mass positions along the horizontal surface, measured from their equilibrium positions, plus right and minus left.

Fixed Ends The first and last spring are attached to fixed walls.

The **competition method** is used to derive the equations of motion. In this case, the law is

$$\text{Newton's Second Law Force} = \text{Sum of the Hooke's Forces.}$$

The model equations are

$$\begin{array}{lrcl} & m_1 x_1''(t) &=& -k_1 x_1(t) + k_2[x_2(t) - x_1(t)], \\ (6) & m_2 x_2''(t) &=& -k_2[x_2(t) - x_1(t)] + k_3[x_3(t) - x_2(t)], \\ & m_3 x_3''(t) &=& -k_3[x_3(t) - x_2(t)] - k_4 x_3(t). \end{array}$$

The equations are justified in the case of all positive variables by observing that the first three springs are elongated by x_1, $x_2 - x_1$, $x_3 - x_2$, respectively. The last spring is compressed by x_3, which accounts for the minus sign.

Another way to justify the equations is through mirror-image symmetry: interchange $k_1 \longleftrightarrow k_4$, $k_2 \longleftrightarrow k_3$, $x_1 \longleftrightarrow x_3$, then equation 2 should be unchanged and equation 3 should become equation 1.

Matrix Formulation. System (6) can be written as a second order vector-matrix system

$$\begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix} \begin{pmatrix} x_1'' \\ x_2'' \\ x_3'' \end{pmatrix} = \begin{pmatrix} -k_1 - k_2 & k_2 & 0 \\ k_2 & -k_2 - k_3 & k_3 \\ 0 & k_3 & -k_3 - k_4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

More succinctly, the system is written as

$$M\vec{\mathbf{x}}''(t) = K\vec{\mathbf{x}}(t)$$

where the **displacement** $\vec{\mathbf{x}}$, **mass matrix** M and **stiffness matrix** K are defined by the formulas

$$\vec{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \; M = \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix}, \; K = \begin{pmatrix} -k_1 - k_2 & k_2 & 0 \\ k_2 & -k_2 - k_3 & k_3 \\ 0 & k_3 & -k_3 - k_4 \end{pmatrix}.$$

Numerical example. Let $m_1 = 1$, $m_2 = 1$, $m_3 = 1$, $k_1 = 2$, $k_2 = 1$, $k_3 = 1$, $k_4 = 2$. Then the system is given by

$$\begin{pmatrix} x_1'' \\ x_2'' \\ x_3'' \end{pmatrix} = \begin{pmatrix} -3 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

The vector solution is given by the formula

$$\begin{aligned}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= (a_1 \cos t + b_1 \sin t) \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \\ &+ \left(a_2 \cos \sqrt{3}t + b_2 \sin \sqrt{3}t\right) \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \\ &+ (a_3 \cos 2t + b_3 \sin 2t) \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix},\end{aligned}$$

where a_1, a_2, a_3, b_1, b_2, b_3 are arbitrary constants.

Boxcars

A special case of the coupled spring-mass system is three boxcars on a level track connected by springs, as in Figure 15.

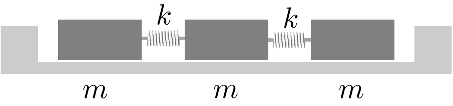

Figure 15. Three identical boxcars connected by identical springs.

Except for the springs on fixed ends, this problem is the same as the one of the preceding illustration, therefore taking $k_1 = k_4 = 0$, $k_2 = k_3 = k$, $m_1 = m_2 = m_3 = m$ gives the system

$$\begin{pmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m \end{pmatrix} \begin{pmatrix} x_1'' \\ x_2'' \\ x_3'' \end{pmatrix} = \begin{pmatrix} -k & k & 0 \\ k & -2k & k \\ 0 & k & -k \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Take $k/m = 1$ to obtain the illustration

$$\vec{\mathbf{x}}'' = \begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{pmatrix} \vec{\mathbf{x}},$$

which has vector solution

$$\begin{aligned}\vec{\mathbf{x}} &= (a_1 + b_1 t) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (a_2 \cos t + b_2 \sin t) \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \\ &+ \left(a_3 \cos \sqrt{3}t + b_3 \sin \sqrt{3}t\right) \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix},\end{aligned}$$

where a_1, a_2, a_3, b_1, b_2, b_3 are arbitrary constants.

The solution expression can be used to discover what happens to the boxcars when the springs act normally upon compression but disengage upon expansion. An interesting physical situation is when one car moves along the track, contacts two stationary cars, then transfers its momentum to the other cars, followed by disengagement.

Monatomic Crystals

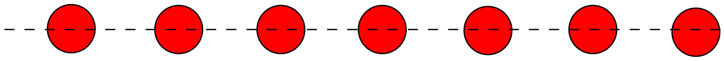

Figure 16. A Crystal Model.
The n crystals are identical masses m assumed connected by equal springs of Hooke's constant k. The last mass is connected to the first mass.

The scalar differential equations for Figure 16 are written for mass positions $x_1, \ldots, x_n$, with $x_0 = x_n$, $x_{n+1} = x_1$ to account for the ring of identical masses (periodic boundary condition). Then for $k = 1, \ldots, n$

$$m\frac{d^2x_k}{dt^2} = k(x_{k+1} - x_k) + k(x_{k-1} - x_k) = k(x_{k-1} - 2x_k + x_{k+1}).$$

These equations represent a system $x'' = Ax$, where the symmetric matrix of coefficients $A = M^{-1}K$ is given for $n = 5$ and $k/m = 1$ by

$$A = \begin{pmatrix} -2 & 1 & 0 & 0 & 1 \\ 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \\ 1 & 0 & 0 & 1 & -2 \end{pmatrix}.$$

If $n = 3$ and $k/m = 1$, then $A = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix}$ and the solutions x_1, x_2, x_3 are linear combinations of the functions 1, t, $\cos\sqrt{3}t$, $\sin\sqrt{3}t$.

Electrical Network I

Consider the LR-network of Figure 17.

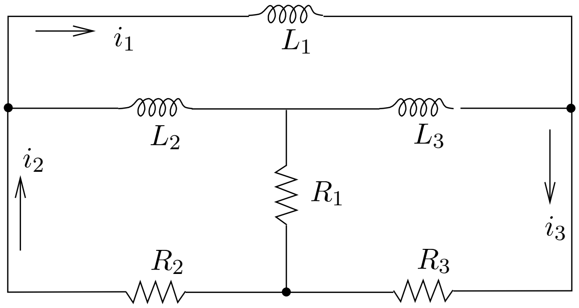

Figure 17. An electrical network. There are three resistors R_1, R_2, R_3 and three inductors L_1, L_2, L_3. The currents i_1, i_2, i_3 are defined between nodes (black dots).

The derivation of the differential equations for the loop currents i_1, i_2, i_3 uses Kirchhoff's laws and the voltage drop formulas for resistors and inductors. The black dots in the diagram are the **nodes** that determine the beginning and end of each of the currents i_1, i_2, i_3. Currents are defined only on the outer boundary of the network. Kirchhoff's node law determines the currents across L_2, L_3 (arrowhead right) as $i_2 - i_1$ and $i_3 - i_1$, respectively. Similarly, $i_2 - i_3$ is the current across R_1 (arrowhead down). Using Ohm's law $V_R = RI$ and Faraday's law $V_L = LI'$ plus Kirchhoff's loop law *algebraic sum of the voltage drops is zero* around a closed loop (see the `maple` code below), we arrive at the model

$$\begin{array}{rcrlcrl}
i_1' &=& - & \left(\dfrac{R_2}{L_1}\right) i_2 &-& & \left(\dfrac{R_3}{L_1}\right) i_3, \\
i_2' &=& - & \left(\dfrac{R_2}{L_2} + \dfrac{R_2}{L_1}\right) i_2 &+& & \left(\dfrac{R_1}{L_2} - \dfrac{R_3}{L_1}\right) i_3, \\
i_3' &=& & \left(\dfrac{R_1}{L_3} - \dfrac{R_2}{L_1}\right) i_2 &-& & \left(\dfrac{R_1}{L_3} + \dfrac{R_3}{L_1} + \dfrac{R_3}{L_3}\right) i_3
\end{array}$$

A computer algebra system is helpful to obtain the differential equations from the closed loop formulas. Part of the theory is that the number of equations equals the number of *holes* in the network, called the **connectivity**. Here's some `maple` code for determining the equations in scalar and also in vector-matrix form.

```
loop1:=L1*D(i1)+R3*i3+R2*i2=0;
loop2:=L2*D(i2)-L2*D(i1)+R1*(i2-i3)+R2*i2=0;
loop3:=L3*D(i3)-L3*D(i1)+R3*i3+R1*(i3-i2)=0;
f1:=solve(loop1,D(i1));
f2:=solve(subs(D(i1)=f1,loop2),D(i2));
f3:=solve(subs(D(i1)=f1,loop3),D(i3));
with(linalg):
jacobian([f1,f2,f3],[i1,i2,i3]);
```

Electrical Network II

Consider the *LR*-network of Figure 18. This network produces only two differential equations, even though there are three *holes* (connectivity 3). The derivation of the differential equations parallels the previous network, so nothing will be repeated here.

A computer algebra system is used to obtain the differential equations from the closed loop formulas. Below is `maple` code to generate the equations $i_1' = f_1$, $i_2' = f_2$, $i_3 = f_3$.

```
loop1:=L1*D(i1)+R2*(i1-i2)+R1*(i1-i3)=0;
loop2:=L2*D(i2)+R3*(i2-i3)+R2*(i2-i1)=0;
```

```
loop3:=R3*(i3-i2)+R1*(i3-i1)=E;
f3:=solve(loop3,i3);
f1:=solve(subs(i3=f3,loop1),D(i1));
f2:=solve(subs(i3=f3,loop2),D(i2));
```

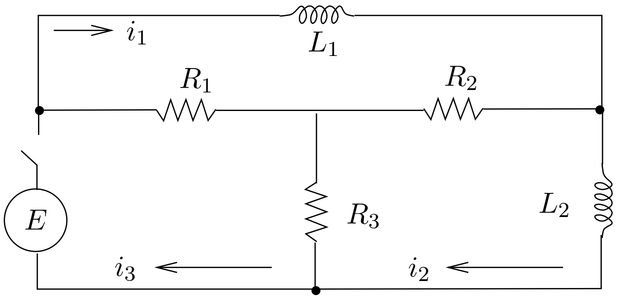

Figure 18. An electrical network.
There are three resistors R_1, R_2, R_3, two inductors L_1, L_2 and a battery E. The currents i_1, i_2, i_3 are defined between nodes (black dots).

The model, in the special case $L_1 = L_2 = 1$ and $R_1 = R_2 = R_3 = R$:

$$\begin{array}{rcrcrcr}
i_1' & = & -\dfrac{3R}{2}i_1 & + & \dfrac{3R}{2}i_2 & + & \dfrac{E}{2}, \\
i_2' & = & \dfrac{3R}{2}i_1 & - & \dfrac{3R}{2}i_2 & + & \dfrac{E}{2}, \\
i_3 & = & \dfrac{1}{2}i_1 & + & \dfrac{1}{2}i_2 & + & \dfrac{E}{2R}.
\end{array}$$

It is easily justified that the solution of the differential equations for initial conditions $i_1(0) = i_2(0) = 0$ is given by

$$i_1(t) = \frac{E}{2}t, \quad i_2(t) = \frac{E}{2}t.$$

Logging Timber by Helicopter

Certain sections of National Forest in the USA do not have logging access roads. In order to log the timber in these areas, helicopters are employed to move the felled trees to a nearby loading area, where they are transported by truck to the mill. The felled trees are slung beneath the helicopter on cables.

Figure 19. Helicopter logging.
Left: An Erickson helicopter lifts felled trees.
Right: Two trees are attached to the cable to lower transportation costs.

The payload for two trees approximates a double pendulum, which oscillates during flight. The angles of oscillation θ_1, θ_2 of the two connecting cables, measured from the gravity vector direction, satisfy the following differential equations, in which g is the gravitation constant, m_1, m_2 denote the masses of the two trees and L_1, L_2 are the cable lengths.

$$\begin{array}{rcrcr} (m_1+m_2)L_1^2\theta_1'' & + & m_2L_1L_2\theta_2'' & + & (m1+m_2)L_1g\theta_1 = 0, \\ m_2L_1L_2\theta_1'' & + & m_2L_2^2\theta_2'' & + & m_2L_2g\theta_2 = 0. \end{array}$$

This model is derived assuming small displacements θ_1, θ_2, that is, $\sin\theta \approx \theta$ for both angles, using the following diagram.

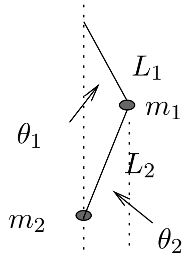

Figure 20. Logging Timber by Helicopter. The cables have lengths L_1, L_2. The angles θ_1, θ_2 are measured from vertical.

The lengths L_1, L_2 are adjusted on each trip for the length of the trees, so that the trees do not collide in flight with each other nor with the helicopter. Sometimes, three or more smaller trees are bundled together in a package, which is treated here as identical to a single, very thick tree hanging on the cable.

Vector-matrix model. The angles θ_1, θ_2 satisfy the second-order vector-matrix equation

$$\begin{pmatrix} (m_1+m_2)L_1 & m_2L_2 \\ L_1 & L_2 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}'' = -\begin{pmatrix} m_1g+m_2g & 0 \\ 0 & g \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}.$$

This system is equivalent to the second-order system

$$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}'' = \begin{pmatrix} -\dfrac{m_1g+m_2g}{L_1m_1} & \dfrac{m_2g}{L_1m_1} \\ \dfrac{m_1g+m_2\,g}{L_2m_1} & -\dfrac{(m_1+m_2)\,g}{L_2m_1} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}.$$

Earthquake Effects on Buildings

A horizontal earthquake oscillation $F(t) = F_0\cos\omega t$ affects each floor of a 5-floor building; see Figure 21. The effect of the earthquake depends upon the natural frequencies of oscillation of the floors.

In the case of a single-floor building, the center-of-mass position $x(t)$ of the building satisfies $mx'' + kx = E$ and the natural frequency of oscillation is $\sqrt{k/m}$. The earthquake force E is given by Newton's second law: $E(t) = -mF''(t)$. If $\omega \approx \sqrt{k/m}$, then the amplitude of $x(t)$ is large compared to the amplitude of the force E. The amplitude increase in $x(t)$ means that a small-amplitude earthquake wave can resonant with the building and possibly demolish the structure.

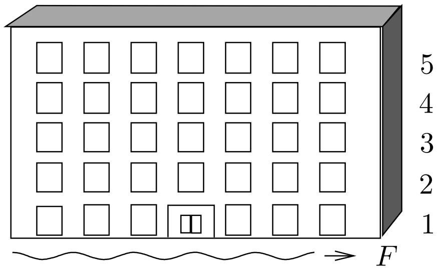

Figure 21. A 5-Floor Building.
A horizontal earthquake wave F affects every floor. The actual wave has wavelength many times larger than the illustration.

The following assumptions and symbols are used to quantize the oscillation of the 5-floor building.

- Each floor is considered a point mass located at its center-of-mass. The floors have masses $m_1, \ldots, m_5$.
- Each floor is restored to its equilibrium position by a linear restoring force or Hooke's force $-k$(elongation). The Hooke's constants are $k_1, \ldots, k_5$.
- The locations of masses representing the 5 floors are $x_1, \ldots, x_5$. The equilibrium position is $x_1 = \cdots = x_5 = 0$.
- Damping effects of the floors are ignored. This is a *frictionless* system.

The differential equations for the model are obtained by **competition**: the Newton's second law force is set equal to the sum of the Hooke's forces and the external force due to the earthquake wave. This results in the following system, where $k_6 = 0$, $E_j = m_j F''$ for $j = 1, 2, 3, 4, 5$ and $F = F_0 \cos \omega t$.

$$\begin{array}{rcl} m_1 x_1'' & = & -(k_1 + k_2)x_1 + k_2 x_2 + E_1, \\ m_2 x_2'' & = & k_2 x_1 - (k_2 + k_3)x_2 + k_3 x_3 + E_2, \\ m_3 x_3'' & = & k_3 x_2 - (k_3 + k_4)x_3 + k_4 x_4 + E_3, \\ m_4 x_4'' & = & k_4 x_3 - (k_4 + k_5)x_4 + k_5 x_5 + E_4, \\ m_5 x_5'' & = & k_5 x_4 - (k_5 + k_6)x_5 + E_5. \end{array}$$

In particular, the equations for a floor depend only upon the neighboring floors. The bottom floor and the top floor are exceptions: they have just one neighboring floor.

Vector-matrix second order system. Define

$$M = \begin{pmatrix} m_1 & 0 & 0 & 0 & 0 \\ 0 & m_2 & 0 & 0 & 0 \\ 0 & 0 & m_3 & 0 & 0 \\ 0 & 0 & 0 & m_4 & 0 \\ 0 & 0 & 0 & 0 & m_5 \end{pmatrix}, \quad \vec{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}, \quad \vec{\mathbf{H}} = \begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \\ E_5 \end{pmatrix},$$

$$K = \begin{pmatrix} -k_1-k_2 & k_2 & 0 & 0 & 0 \\ k_2 & -k_2-k_3 & k_3 & 0 & 0 \\ 0 & k_3 & -k_3-k_4 & k_4 & 0 \\ 0 & 0 & k_4 & -k_4-k_5 & k_5 \\ 0 & 0 & 0 & k_5 & -k_5-k_6 \end{pmatrix}.$$

In the last row, $k_6 = 0$, to reflect the absence of a floor above the fifth. The second order system is

$$M\vec{\mathbf{x}}''(t) = K\vec{\mathbf{x}}(t) + \vec{\mathbf{H}}(t).$$

The matrix M is called the **mass matrix** and the matrix K is called the **Hooke's matrix**. The **external force** $\vec{\mathbf{H}}(t)$ can be written as a scalar function $E(t) = -F''(t)$ times a constant vector:

$$\vec{\mathbf{H}}(t) = -\omega^2 F_0 \cos\omega t \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ m_5 \end{pmatrix}.$$

Identical floors. Let us assume that all floors have the same mass m and the same Hooke's constant k. Then $M = mI$ and the equation becomes

$$\vec{\mathbf{x}}'' = m^{-1}\begin{pmatrix} -2k & k & 0 & 0 & 0 \\ k & -2k & k & 0 & 0 \\ 0 & k & -2k & k & 0 \\ 0 & 0 & k & -2k & k \\ 0 & 0 & 0 & k & -k \end{pmatrix}\vec{\mathbf{x}} - F_0\omega^2\cos(\omega t)\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

The Hooke's matrix K is symmetric ($K^T = K$) with negative entries only on the diagonal. The last diagonal entry is $-k$ (a common error is to write $-2k$).

Particular solution. The method of undetermined coefficients predicts a trial solution $\vec{\mathbf{x}}_p(t) = \vec{\mathbf{c}}\cos\omega t$, because each differential equation has nonhomogeneous term $-F_0\omega^2\cos\omega t$. The constant vector $\vec{\mathbf{c}}$ is found by trial solution substitution. Cancel the common factor $\cos\omega t$ in the substituted equation to obtain the equation $\left(m^{-1}K + \omega^2 I\right)\vec{\mathbf{c}} = F_0\omega^2\vec{\mathbf{b}}$, where $\vec{\mathbf{b}}$ is column vector of ones in the preceding display. Let $B(\omega) = m^{-1}K + \omega^2 I$. Then the formula $B^{-1} = \dfrac{\mathbf{adj}(B)}{\det(B)}$ gives

$$\vec{\mathbf{c}} = F_0\omega^2\frac{\mathbf{adj}(B(\omega))}{\det(B(\omega))}\vec{\mathbf{b}}.$$

The constant vector $\vec{\mathbf{c}}$ can have a large magnitude when $\det(B(\omega)) \approx 0$. This occurs when $-\omega^2$ is nearly an eigenvalue of $m^{-1}K$.

Homogeneous solution. The theory of this chapter gives the homogeneous solution $\vec{\mathbf{x}}_h(t)$ as the sum

$$\vec{\mathbf{x}}_h(t) = \sum_{j=1}^{5}(a_j \cos\omega_j t + b_j \sin\omega_j t)\vec{\mathbf{v}}_j$$

where $r = \omega_j$ and $\vec{\mathbf{v}} = \vec{\mathbf{v}}_j \neq \vec{\mathbf{0}}$ satisfy

$$\left(\frac{1}{m}K + r^2\, I\right)\vec{\mathbf{v}} = \vec{\mathbf{0}}.$$

Special case $k/m = 10$. Then

$$\frac{1}{m}K = \begin{pmatrix} -20 & 10 & 0 & 0 & 0 \\ 10 & -20 & 10 & 0 & 0 \\ 0 & 10 & -20 & 10 & 0 \\ 0 & 0 & 10 & -20 & 10 \\ 0 & 0 & 0 & 10 & -10 \end{pmatrix}$$

and the values $\omega_1, \ldots, \omega_5$ are found by solving the determinant equation $\det((1/m)K + \omega^2 I) = 0$, to obtain the values in Table 1.

Table 1. The natural frequencies for the special case $k/m = 10$**.**

Frequency	Value
ω_1	0.900078068
ω_2	2.627315231
ω_3	4.141702938
ω_4	5.320554507
ω_5	6.068366391

General solution. Superposition implies $\vec{\mathbf{x}}(t) = \vec{\mathbf{x}}_h(t) + \vec{\mathbf{x}}_p(t)$. Both terms of the general solution represent bounded oscillations.

Resonance effects. The special solution $\vec{\mathbf{x}}_p(t)$ can be used to obtain some insight into practical resonance effects between the incoming earthquake wave and the building floors. When ω is close to one of the frequencies $\omega_1, \ldots, \omega_5$, then the amplitude of a component of $\vec{\mathbf{x}}_p$ can be very large, causing the floor to take an excursion that is too large to maintain the structural integrity of the floor.

The **physical interpretation** is that an earthquake wave of the proper frequency, having time duration sufficiently long, can demolish a floor and hence demolish the entire building. The amplitude of the earthquake wave does not have to be large: a fraction of a centimeter might be enough to start the oscillation of the floors.

Earthquakes and Tsunamis

Seismic wave shape was studied for first order equations in Chapter 2, page 151. Recorded here are some historical notes about seismic waves and earthquake events.

The original **Richter scale**, with deprecated use in seismology, was invented by seismologist C. Richter to rank earthquake power.

The moment magnitude scale (M_W) has largely replaced the original Richter scale and its modified versions. The highest reported magnitude is 9.5 M_W by the United States Geological Survey for the Concepción, Chile earthquake of May 22, 1960. News reports and the general public still refer to earthquake magnitude using the term *Richter Scale*.

The Sumatra earthquake of December 26, 2004 occurred close to a deep-sea trench, a subduction zone where one tectonic plate slips beneath another. Most of the earthquake energy is released in these areas as the two plates grind towards each other. Estimates of magnitude 8.8 M_W to 9.3 M_W followed the event. The US Geological Survey estimated 9.2 M_W.

The Chile earthquake and tsunami of 1960 has been documented well. Here is an account by Dr. Gerard Fryer of the Hawaii Institute of Geophysics and Planetology, in Honolulu.

> The tsunami was generated by the Chile earthquake of May 22, 1960, the largest earthquake ever recorded: it was magnitude 9.6. What happened in the earthquake was that a piece of the Pacific seafloor (or strictly speaking, the Nazca Plate) about the size of California slid fifty feet beneath the continent of South America. Like a spring, the lower slopes of the South American continent offshore snapped upwards as much as twenty feet while land along the Chile coast dropped about ten feet. This change in the shape of the ocean bottom changed the shape of the sea surface. Since the sea surface likes to be flat, the pile of excess water at the surface collapsed to create a series of waves — the tsunami.
>
> The tsunami, together with the coastal subsidence and flooding, caused tremendous damage along the Chile coast, where about 2,000 people died. The waves spread outwards across the Pacific. About 15 hours later the waves flooded Hilo, on the island of Hawaii, where they built up to 30 feet and caused 61 deaths along the waterfront. Seven hours after that, 22 hours after the earthquake, the waves flooded the coastline of Japan where 10-foot waves caused 200 deaths. The waves also caused damage in the Marquesas, in Samoa, and in New Zealand. Tidal gauges throughout the Pacific measured anomalous oscillations for about three days as the waves bounced from one side of the ocean to the other.

11.2 Basic First-order System Methods

Solving 2×2 Systems

It is shown here that *any* constant linear system

$$\vec{\mathbf{u}}' = A\vec{\mathbf{u}}, \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

can be solved by one of the following elementary methods.

(a) The integrating factor method for $y' = p(x)y + q(x)$.

(b) The second order constant coefficient formulas in Theorem 1, Chapter 5.

Triangular A. Let's assume $b = 0$, so that A is lower triangular. The upper triangular case is handled similarly. Then $\vec{\mathbf{u}}' = A\vec{\mathbf{u}}$ has the scalar form

$$\begin{array}{rcl} u_1' & = & au_1, \\ u_2' & = & cu_1 + du_2. \end{array}$$

The first differential equation is solved by the growth/decay formula:

$$u_1(t) = u_0 e^{at}.$$

Then substitute the answer just found into the second differential equation to give

$$u_2' = du_2 + cu_0 e^{at}.$$

This is a linear first order equation of the form $y' = p(x)y + q(x)$, to be solved by the integrating factor method. Therefore, a triangular system can always be solved by the first order integrating factor method.

An illustration. Let us solve $\vec{\mathbf{u}}' = A\vec{\mathbf{u}}$ for the triangular matrix

$$A = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \quad \text{representing} \quad \left\{ \begin{array}{rcl} u_1' & = & u_1, \\ u_2' & = & 2u_1 + u_2. \end{array} \right.$$

The first equation $u_1' = u_1$ has solution $u_1 = c_1 e^t$. The second equation $u_2' = 2u_1 + u_2$ becomes upon substitution of $u_1 = c_1 e^t$ the new equation

$$u_2' = 2c_1 e^t + u_2,$$

which is a first order linear differential equation with linear integrating factor method solution $u_2 = (2c_1 t + c_2)e^t$. The general solution of $\vec{\mathbf{u}}' = A\vec{\mathbf{u}}$ in scalar form is

$$u_1 = c_1 e^t, \quad u_2 = 2c_1 t e^t + c_2 e^t.$$

The **vector form** of the general solution is

$$\vec{\mathbf{u}}(t) = c_1 \begin{pmatrix} e^t \\ 2te^t \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ e^t \end{pmatrix}.$$

The **vector basis** is the set

$$\mathcal{B} = \left\{ \begin{pmatrix} e^t \\ 2te^t \end{pmatrix}, \begin{pmatrix} 0 \\ e^t \end{pmatrix} \right\}.$$

Non-Triangular A. In order that A be non-triangular, both $b \neq 0$ and $c \neq 0$ must be satisfied. The scalar form of the system $\vec{\mathbf{u}}' = A\vec{\mathbf{u}}$ is

$$\begin{aligned} u_1' &= au_1 + bu_2, \\ u_2' &= cu_1 + du_2. \end{aligned}$$

Theorem 1 (Solving Non-Triangular $\vec{\mathbf{u}}' = A\vec{\mathbf{u}}$)
Solutions u_1, u_2 of $\vec{\mathbf{u}}' = A\vec{\mathbf{u}}$ are linear combinations of the list of Euler solution atoms obtained from the roots r of the quadratic equation

$$\det(A - rI) = 0.$$

Proof: The method: differentiate the first equation, then use the equations to eliminate u_2, u_2'. The result is a second order differential equation for u_1. The same differential equation is satisfied also for u_2. The details:

$$\begin{aligned} u_1'' &= au_1' + bu_2' && \text{Differentiate the first equation.} \\ &= au_1' + bcu_1 + bdu_2 && \text{Use equation } u_2' = cu_1 + du_2. \\ &= au_1' + bcu_1 + d(u_1' - au_1) && \text{Use equation } u_1' = au_1 + bu_2. \\ &= (a+d)u_1' + (bc - ad)u_1 && \text{Second order equation for } u_1 \text{ found} \end{aligned}$$

The characteristic equation of $u_1'' - (a+d)u_1' + (ad - bc)u_1 = 0$ is

$$r^2 - (a+d)r + (bc - ad) = 0.$$

Finally, we show the expansion of $\det(A - rI)$ is the same characteristic polynomial:

$$\begin{aligned} \det(A - rI) &= \begin{vmatrix} a - r & b \\ c & d - r \end{vmatrix} \\ &= (a - r)(d - r) - bc \\ &= r^2 - (a + d)r + ad - bc. \end{aligned}$$

The proof is complete.

The reader can verify that the differential equation for u_1 or u_2 is exactly

$$u'' - \mathbf{trace}(A)u' + \det(A)u = 0.$$

Assume below that A is non-triangular, meaning $b \neq 0$ and $c \neq 0$.

Finding u_1. Apply the second order formulas, Theorem 1 in Chapter 5, to solve for u_1. This involves writing a list of Euler solution atoms corresponding to the two roots of the characteristic equation $r^2 - (a + d)r + ad - bc = 0$, followed by expressing u_1 as a linear combination of the two Euler atoms.

Finding u_2. Isolate u_2 in the first differential equation by division:

$$u_2 = \frac{1}{b}(u_1' - au_1).$$

The two formulas for u_1, u_2 represent the general solution of the system $\vec{\mathbf{u}}' = A\vec{\mathbf{u}}$, when A is 2×2.

An illustration. Let's solve $\vec{\mathbf{u}}' = A\vec{\mathbf{u}}$ when

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad \text{representing} \quad \begin{cases} u_1' &= u_1 + 2u_2, \\ u_2' &= 2u_1 + u_2. \end{cases}$$

The equation $\det(A - rI) = 0$ is $(1-r)^2 - 4 = 0$ with roots $r = -1$ and $r = 3$. The Euler solution atom list is $L = \{e^{-t}, e^{3t}\}$. Then the linear combination of Euler atoms is $u_1 = c_1e^{-t} + c_2e^{3t}$. The first equation $u_1' = u_1 + 2u_2$ implies $u_2 = \frac{1}{2}(u_1' - u_1)$. The scalar general solution of $\vec{\mathbf{u}}' = A\vec{\mathbf{u}}$ is then

$$u_1 = c_1e^{-t} + c_2e^{3t}, \quad u_2 = -c_1e^{-t} + c_2e^{3t}.$$

In vector form, the general solution is

$$\vec{\mathbf{u}} = c_1 \begin{pmatrix} e^{-t} \\ -e^{-t} \end{pmatrix} + c_2 \begin{pmatrix} e^{3t} \\ e^{3t} \end{pmatrix}.$$

Triangular Methods

Diagonal $n \times n$ **matrix** $A = \mathbf{diag}(a_1, \ldots, a_n)$. Then the system $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ is a set of uncoupled scalar growth/decay equations:

$$\begin{aligned} x_1'(t) &= a_1x_1(t), \\ x_2'(t) &= a_2x_2(t), \\ &\vdots \\ x_n'(t) &= a_nx_n(t). \end{aligned}$$

The solution to the system is given by the formulas

$$\begin{aligned} x_1(t) &= c_1e^{a_1t}, \\ x_2(t) &= c_2e^{a_2t}, \\ &\vdots \\ x_n(t) &= c_ne^{a_nt}. \end{aligned}$$

The numbers $c_1, \ldots, c_n$ are arbitrary constants.

Triangular $n \times n$ matrix A. If a linear system $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ has a square triangular matrix A, then the system can be solved by first order scalar methods. To illustrate the ideas, consider the 3×3 linear system

$$\vec{\mathbf{x}}' = \begin{pmatrix} 2 & 0 & 0 \\ 3 & 3 & 0 \\ 4 & 4 & 4 \end{pmatrix} \vec{\mathbf{x}}.$$

The coefficient matrix A is *lower triangular.* In scalar form, the system is given by the equations

$$\begin{aligned} x_1'(t) &= 2x_1(t), \\ x_2'(t) &= 3x_1(t) + 3x_2(t), \\ x_3'(t) &= 4x_1(t) + 4x_2(t) + 4x_3(t). \end{aligned}$$

A recursive method. The system is solved recursively by first order scalar methods only, starting with the first equation $x_1'(t) = 2x_1(t)$. This growth equation has general solution $x_1(t) = c_1e^{2t}$. The second equation then becomes the first order linear equation

$$\begin{aligned} x_2'(t) &= 3x_1(t) + 3x_2(t) \\ &= 3x_2(t) + 3c_1e^{2t}. \end{aligned}$$

The integrating factor method applies to find the general solution $x_2(t) = -3c_1e^{2t}+c_2e^{3t}$. The third and last equation becomes the first order linear equation

$$\begin{aligned} x_3'(t) &= 4x_1(t) + 4x_2(t) + 4x_3(t) \\ &= 4x_3(t) + 4c_1e^{2t} + 4(-3c_1e^{2t} + c_2e^{3t}). \end{aligned}$$

The integrating factor method is repeated to find the general solution $x_3(t) = 4c_1e^{2t} - 4c_2e^{3t} + c_3e^{4t}$.

In summary, the scalar general solution to the system is given by the formulas

$$\begin{aligned} x_1(t) &= c_1e^{2t}, \\ x_2(t) &= -3c_1e^{2t} + c_2e^{3t}, \\ x_3(t) &= 4c_1e^{2t} - 4c_2e^{3t} + c_3e^{4t}. \end{aligned}$$

Structure of solutions. A system $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ for $n \times n$ triangular A has component solutions $x_1(t)$, ..., $x_n(t)$ given as polynomials times exponentials. The exponential factors $e^{a_{11}t}$, ..., $e^{a_{nn}t}$ are expressed in terms of the diagonal elements a_{11}, ..., a_{nn} of the matrix A. Fewer than n distinct exponential factors may appear, due to duplicate diagonal elements. These duplications cause the polynomial factors to appear. The reader is invited to work out the solution to the system below, which has duplicate diagonal entries $a_{11} = a_{22} = a_{33} = 2$.

$$\begin{aligned} x_1'(t) &= 2x_1(t), \\ x_2'(t) &= 3x_1(t) + 2x_2(t), \\ x_3'(t) &= 4x_1(t) + 4x_2(t) + 2x_3(t). \end{aligned}$$

The solution, given below, has polynomial factors t and t^2, appearing because of the duplicate diagonal entries $2, 2, 2$, and only one exponential factor e^{2t}.

$$\begin{array}{rcl} x_1(t) & = & c_1 e^{2t}, \\ x_2(t) & = & 3c_1 t e^{2t} + c_2 e^{2t}, \\ x_3(t) & = & 4c_1 t e^{2t} + 6c_1 t^2 e^{2t} + 4c_2 t e^{2t} + c_3 e^{2t}. \end{array}$$

Conversion to Systems

Routinely converted to a system of equations of first order are scalar second order linear differential equations, systems of scalar second order linear differential equations and scalar linear differential equations of higher order.

Scalar second order linear equations. Consider an equation $au'' + bu' + cu = f$ where $a \neq 0$, b, c, f are allowed to depend on t, $' = d/dt$. Define the **position-velocity substitution**

$$x(t) = u(t), \quad y(t) = u'(t).$$

Then $x' = u' = y$ and $y' = u'' = (-bu' - cu + f)/a = -(b/a)y - (c/a)x + f/a$. The resulting system is equivalent to the second order equation, in the sense that the position-velocity substitution equates solutions of one system to the other:

$$\begin{array}{rcl} x'(t) & = & y(t), \\ y'(t) & = & -\dfrac{c(t)}{a(t)}x(t) - \dfrac{b(t)}{a(t)}y(t) + \dfrac{f(t)}{a(t)}. \end{array}$$

The case of constant coefficients and f a function of t arises often enough to isolate the result for further reference.

Theorem 2 (System Equivalent to Second Order Linear)
Let $a \neq 0$, b, c be constants and $f(t)$ continuous. Then $au''+bu'+cu = f(t)$ is equivalent to the first order system

$$a\vec{\mathbf{w}}\,'(t) = \begin{pmatrix} 0 & a \\ -c & -b \end{pmatrix} \vec{\mathbf{w}}\,(t) + \begin{pmatrix} 0 \\ f(t) \end{pmatrix}, \quad \vec{\mathbf{w}}\,(t) = \begin{pmatrix} u(t) \\ u'(t) \end{pmatrix}.$$

Converting second order systems to first order systems. A similar position-velocity substitution can be carried out on a system of two second order linear differential equations. Assume

$$\begin{array}{rcl} a_1 u_1'' + b_1 u_1' + c_1 u_1 & = & f_1, \\ a_2 u_2'' + b_2 u_2' + c_2 u_2 & = & f_2. \end{array}$$

Then the preceding methods for the scalar case give the equivalence

$$\begin{pmatrix} a_1 & 0 & 0 & 0 \\ 0 & a_1 & 0 & 0 \\ 0 & 0 & a_2 & 0 \\ 0 & 0 & 0 & a_2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_1' \\ u_2 \\ u_2' \end{pmatrix}' = \begin{pmatrix} 0 & a_1 & 0 & 0 \\ -c_1 & -b_1 & 0 & 0 \\ 0 & 0 & 0 & a_2 \\ 0 & 0 & -c_2 & -b_2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_1' \\ u_2 \\ u_2' \end{pmatrix} + \begin{pmatrix} 0 \\ f_1 \\ 0 \\ f_2 \end{pmatrix}.$$

Coupled spring-mass systems. Springs connecting undamped coupled masses were considered at the beginning of this chapter, page 786. Typical equations are

$$(1) \qquad \begin{aligned} m_1 x_1''(t) &= -k_1 x_1(t) + k_2[x_2(t) - x_1(t)], \\ m_2 x_2''(t) &= -k_2[x_2(t) - x_1(t)] + k_3[x_3(t) - x_2(t)], \\ m_3 x_3''(t) &= -k_3[x_3(t) - x_2(t)] - k_4 x_3(t). \end{aligned}$$

The equations can be represented by a second order linear system of dimension 3 of the form $M\vec{\mathbf{x}}'' = K\vec{\mathbf{x}}$, where the **position** $\vec{\mathbf{x}}$, the **mass matrix** M and the **Hooke's matrix** K are given by the equalities

$$\vec{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad M = \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix},$$

$$K = \begin{pmatrix} -(k_1+k_2) & k_2 & 0 \\ k_2 & -(k_2+k_3) & k_3 \\ 0 & -k_3 & -(k_3+k_4) \end{pmatrix}.$$

Systems of second order linear equations. A second order system $M\vec{\mathbf{x}}'' = K\vec{\mathbf{x}} + \vec{\mathbf{F}}(t)$ is called a **forced system** and $\vec{\mathbf{F}}$ is called the **external vector force**. Such a system can always be converted to a second order system where the mass matrix is the identity, by multiplying by M^{-1}:

$$\vec{\mathbf{x}}'' = M^{-1}K\vec{\mathbf{x}} + M^{-1}\vec{\mathbf{F}}(t).$$

The benign form $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}} + \vec{\mathbf{G}}(t)$, where $A = M^{-1}K$ and $\vec{\mathbf{G}} = M^{-1}\vec{\mathbf{F}}$, admits a block matrix conversion into a first order system:

$$\frac{d}{dt}\begin{pmatrix} \vec{\mathbf{x}}(t) \\ \vec{\mathbf{x}}'(t) \end{pmatrix} = \left(\begin{array}{c|c} 0 & I \\ \hline A & 0 \end{array}\right) \begin{pmatrix} \vec{\mathbf{x}}(t) \\ \vec{\mathbf{x}}'(t) \end{pmatrix} + \begin{pmatrix} \vec{\mathbf{0}} \\ \vec{\mathbf{G}}(t) \end{pmatrix}.$$

Damped second order systems. The addition of a dashpot to each of the masses gives a **damped second order system** with forcing

$$M\vec{\mathbf{x}}'' = B\vec{\mathbf{x}}' + K\vec{\mathbf{X}} + \vec{\mathbf{F}}(t).$$

In the case of one scalar equation, the matrices M, B, K are constants m, $-c$, $-k$ and the external force is a scalar function $f(t)$, hence the system becomes the classical damped spring-mass equation

$$mx'' + cx' + kx = f(t).$$

A useful way to write the first order system is to introduce variable $\vec{\mathbf{u}} = M\vec{\mathbf{x}}$, in order to obtain the two equations

$$\vec{\mathbf{u}}' = M\vec{\mathbf{x}}', \quad \vec{\mathbf{u}}'' = B\vec{\mathbf{x}}' + K\vec{\mathbf{x}} + \vec{\mathbf{F}}(t).$$

Then a first order system in block matrix form is given by

$$\left(\begin{array}{c|c} M & 0 \\ \hline 0 & M \end{array}\right) \frac{d}{dt} \begin{pmatrix} \vec{\mathbf{x}}(t) \\ \vec{\mathbf{x}}'(t) \end{pmatrix} = \left(\begin{array}{c|c} 0 & M \\ \hline K & B \end{array}\right) \begin{pmatrix} \vec{\mathbf{x}}(t) \\ \vec{\mathbf{x}}'(t) \end{pmatrix} + \begin{pmatrix} \vec{\mathbf{0}} \\ \vec{\mathbf{F}}(t) \end{pmatrix}.$$

The benign form $\vec{\mathbf{x}}'' = M^{-1}B\vec{\mathbf{x}}' + M^{-1}K\vec{\mathbf{x}} + M^{-1}\vec{\mathbf{F}}(t)$, obtained by left-multiplication by M^{-1}, can be similarly written as a first order system in block matrix form.

$$\frac{d}{dt} \begin{pmatrix} \vec{\mathbf{x}}(t) \\ \vec{\mathbf{x}}'(t) \end{pmatrix} = \left(\begin{array}{c|c} 0 & I \\ \hline M^{-1}K & M^{-1}B \end{array}\right) \begin{pmatrix} \vec{\mathbf{x}}(t) \\ \vec{\mathbf{x}}'(t) \end{pmatrix} + \begin{pmatrix} \vec{\mathbf{0}} \\ M^{-1}\vec{\mathbf{F}}(t) \end{pmatrix}.$$

Higher order linear equations. Every homogeneous nth order constant-coefficient linear differential equation

$$y^{(n)} = p_0 y + \cdots + p_{n-1} y^{(n-1)}$$

can be converted to a linear homogeneous vector-matrix system

$$\frac{d}{dx} \begin{pmatrix} y \\ y' \\ y'' \\ \vdots \\ y^{(n-1)} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & & & \vdots & \\ 0 & 0 & 0 & \cdots & 1 \\ p_0 & p_1 & p_2 & \cdots & p_{n-1} \end{pmatrix} \begin{pmatrix} y \\ y' \\ y'' \\ \vdots \\ y^{(n-1)} \end{pmatrix}.$$

This is a linear system $\vec{\mathbf{u}}' = A\vec{\mathbf{u}}$ where $\vec{\mathbf{u}}$ is the $n \times 1$ column vector consisting of y and its successive derivatives, while the $n \times n$ matrix A is the classical **companion matrix** of the characteristic polynomial

$$r^n = p_0 + p_1 r + p_2 r^2 + \cdots + p_{n-1} r^{n-1}.$$

To illustrate, the companion matrix for $r^4 = a + br + cr^2 + dr^3$ is

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ a & b & c & d \end{pmatrix}.$$

The preceding companion matrix has the following block matrix form, which is representative of all companion matrices.

$$A = \left(\begin{array}{c|ccc} \vec{\mathbf{0}} & & I & \\ \hline a & b & c & d \end{array}\right).$$

Continuous coefficients. It is routinely observed that the methods above for conversion to a first order system apply equally as well to higher order linear differential equations with continuous coefficients. To illustrate, the fourth order linear equation $y^{iv} = a(x)y + b(x)y' + c(x)y'' + d(x)y'''$ has first order system form $\vec{\mathbf{u}}\,' = A\vec{\mathbf{u}}$ where A is the companion matrix for the polynomial $r^4 = a(x) + b(x)r + c(x)r^2 + d(x)r^3$, x held fixed.

Forced higher order linear equations. All that has been said above applies equally to a forced linear equation like

$$y^{iv} = 2y + \sin(x)y' + \cos(x)y'' + x^2y''' + f(x).$$

It has a conversion to a first order nonhomogeneous linear system

$$\vec{\mathbf{u}}\,' = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 2 & \sin x & \cos x & x^2 \end{pmatrix} \vec{\mathbf{u}} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ f(x) \end{pmatrix}, \quad \vec{\mathbf{u}} = \begin{pmatrix} y \\ y' \\ y'' \\ y''' \end{pmatrix}.$$

11.3 Structure of Linear Systems

Linear systems. A **linear system** is a system of differential equations of the form

$$(1)\qquad \begin{array}{rcrcccrcr} x_1' & = & a_{11}x_1 & + & \cdots & + & a_{1n}x_n & + & f_1, \\ x_2' & = & a_{21}x_1 & + & \cdots & + & a_{2n}x_n & + & f_2, \\ \vdots & & & & \vdots \;\cdots\; \vdots & & & & \vdots \\ x_m' & = & a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & + & f_m, \end{array}$$

where $' = d/dt$. Given are the functions $a_{ij}(t)$ and $f_j(t)$ on some interval $a < t < b$. The unknowns are the functions $x_1(t), \ldots, x_n(t)$.

The system is called **homogeneous** if all $f_j = 0$, otherwise it is called **non-homogeneous**.

Matrix Notation for Systems. A non-homogeneous system of linear equations (1) is written as the equivalent vector-matrix system

$$\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t),$$

where

$$\vec{\mathbf{x}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \vec{\mathbf{f}} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ldots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

Existence-uniqueness. The fundamental theorem of Picard and Lindelöf applied to the matrix system $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t)$ says that a unique solution $\vec{\mathbf{x}}(t)$ exists for each initial value problem and the solution exists on the common interval of continuity of the entries in $A(t)$ and $\vec{\mathbf{f}}(t)$.

Three special results are isolated here, to illustrate how the Picard theory is applied to linear systems.

Theorem 3 (Unique Zero Solution)
Let $A(t)$ be an $m \times n$ matrix with entries continuous on $a < t < b$. Then the initial value problem

$$\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}}, \quad \vec{\mathbf{x}}(0) = \vec{\mathbf{0}}$$

has unique solution $\vec{\mathbf{x}}(t) = \vec{\mathbf{0}}$ on $a < t < b$.

Theorem 4 (Existence-Uniqueness for Constant Linear Systems)
Let $A(t) = A$ be an $m \times n$ matrix with constant entries and let $\vec{\mathbf{x}}_0$ be any m-vector. Then the initial value problem

$$\vec{\mathbf{x}}' = A\vec{\mathbf{x}}, \quad \vec{\mathbf{x}}(0) = \vec{\mathbf{x}}_0$$

has a unique solution $\vec{\mathbf{x}}(t)$ defined for all values of t.

Theorem 5 (Uniqueness and Solution Crossings)
Let $A(t)$ be an $m \times n$ matrix with entries continuous on $a < t < b$ and assume $\vec{\mathbf{f}}(t)$ is also continuous on $a < t < b$. If $\vec{\mathbf{x}}(t)$ and $\vec{\mathbf{y}}(t)$ are solutions of $\vec{\mathbf{u}}' = A(t)\vec{\mathbf{u}} + \vec{\mathbf{f}}(t)$ on $a < t < b$ and $\vec{\mathbf{x}}(t_0) = \vec{\mathbf{y}}(t_0)$ for some t_0, $a < t_0 < b$, then $\vec{\mathbf{x}}(t) = \vec{\mathbf{y}}(t)$ for $a < t < b$.

Superposition. Linear homogeneous systems have **linear structure** and the solutions to nonhomogeneous systems obey a **principle of superposition**.

Theorem 6 (Linear Structure)
Let $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}}$ have two solutions $\vec{\mathbf{x}}_1(t)$, $\vec{\mathbf{x}}_2(t)$. If k_1, k_2 are constants, then $\vec{\mathbf{x}}(t) = k_1\,\vec{\mathbf{x}}_1(t) + k_2\,\vec{\mathbf{x}}_2(t)$ is also a solution of $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}}$.

The standard basis $\{\vec{\mathbf{w}}_k\}_{k=1}^n$. The Picard-Lindelöf theorem applied to initial conditions $\vec{\mathbf{x}}(t_0) = \vec{\mathbf{x}}_0$, with $\vec{\mathbf{x}}_0$ successively set equal to the columns of the $n \times n$ identity matrix, produces n solutions $\vec{\mathbf{w}}_1, \ldots, \vec{\mathbf{w}}_n$ to the equation $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}}$, all of which exist on the same interval $a < t < b$.

The linear structure theorem implies that for any choice of the constants $c_1, \ldots, c_n$, the vector linear combination

$$\vec{\mathbf{x}}(t) = c_1\vec{\mathbf{w}}_1(t) + c_2\vec{\mathbf{w}}_2(t) + \cdots + c_n\vec{\mathbf{w}}_n(t) \tag{2}$$

is a solution of $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}}$.

Conversely, if $c_1, \ldots, c_n$ are taken to be the components of a given vector $\vec{\mathbf{x}}_0$, then the above linear combination must be the unique solution of the initial value problem with $\vec{\mathbf{x}}(t_0) = \vec{\mathbf{x}}_0$. Therefore, all solutions of the equation $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}}$ are given by the expression above, where $c_1, \ldots, c_n$ are taken to be **arbitrary constants**. In summary:

Theorem 7 (Basis)
The solution set of $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}}$ is an n-dimensional subspace of the vector space of all vector-valued functions $\vec{\mathbf{x}}(t)$. Every solution has a unique basis expansion (2).

Theorem 8 (Superposition Principle)
Let $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t)$ have a particular solution $\vec{\mathbf{x}}_p(t)$. If $\vec{\mathbf{x}}(t)$ is any solution of $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t)$, then $\vec{\mathbf{x}}(t)$ can be decomposed as **homogeneous plus particular**:

$$\vec{\mathbf{x}}(t) = \vec{\mathbf{x}}_h(t) + \vec{\mathbf{x}}_p(t).$$

The term $\vec{\mathbf{x}}_h(t)$ is a certain solution of the homogeneous differential equation $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}}$, which means arbitrary constants c_1, c_2, ... have been assigned certain values. The particular solution $\vec{\mathbf{x}}_p(t)$ can be selected to be free of any unresolved or arbitrary constants.

Theorem 9 (Difference of Solutions)
Let $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t)$ have two solutions $\vec{\mathbf{x}} = \vec{\mathbf{u}}(t)$ and $\vec{\mathbf{x}} = \vec{\mathbf{v}}(t)$. Define $\vec{\mathbf{y}}(t) = \vec{\mathbf{u}}(t) - \vec{\mathbf{v}}(t)$. Then $\vec{\mathbf{y}}(t)$ satisfies the homogeneous equation

$$\vec{\mathbf{y}}' = A(t)\vec{\mathbf{y}}.$$

General Solution. We explain **general solution** by example. If a formula $x = c_1e^t + c_2e^{2t}$ is called a general solution, then it means that all possible solutions of the differential equation are expressed by this formula. In particular, it means that a given solution can be represented by the formula, by specializing values for the constants c_1, c_2. We expect the number of arbitrary constants to be the least possible number.

The general solution of $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t)$ is an expression involving arbitrary constants c_1, c_2, $\ldots$ and certain functions. The expression is often given in vector notation, although scalar expressions are commonplace and perfectly acceptable. Required is that the expression represents all solutions of the differential equation, in the following sense:

(a) Every **assignment of constants** produces a solution of the differential equation.

(b) Every possible solution is uniquely obtained from the expression by **specializing the constants**.

Due to the superposition principle, the constants in the general solution are identified as multipliers against solutions of the homogeneous differential equation. The general solution has some recognizable structure.

Theorem 10 (General Solution)
Let $A(t)$ be $n \times n$ and $\vec{\mathbf{f}}(t)$ $n \times 1$, both continuous on an interval $a < t < b$. The linear nonhomogeneous system $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t)$ has general solution $\vec{\mathbf{x}}$ given by the expression

$$\vec{\mathbf{x}} = \vec{\mathbf{x}}_h(t) + \vec{\mathbf{x}}_p(t).$$

The term $\vec{\mathbf{y}} = \vec{\mathbf{x}}_h(t)$ is a general solution of the homogeneous equation $\vec{\mathbf{y}}' = A(t)\vec{\mathbf{y}}$, in which are to be found n arbitrary constants c_1, $\ldots$, c_n. The term $\vec{\mathbf{x}} = \vec{\mathbf{x}}_p(t)$ is a particular solution of $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t)$, in which there are present no unresolved nor arbitrary constants.

Recognition of homogeneous solution terms. An expression $\vec{\mathbf{x}}$ for the general solution of a nonhomogeneous equation $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t)$ involves arbitrary constants $c_1, \ldots, c_n$. It is possible to isolate both terms $\vec{\mathbf{x}}_h$ and $\vec{\mathbf{x}}_p$ by a simple procedure.

To find $\vec{\mathbf{x}}_p$, set to zero all arbitrary constants c_1, c_2, $\ldots$; the resulting expression is free of unresolved and arbitrary constants.

To find $\vec{\mathbf{x}}_h$, we find first the vector solutions $\vec{\mathbf{y}} = \vec{\mathbf{u}}_k(t)$ of $\vec{\mathbf{y}}' = A(t)\vec{\mathbf{y}}$, which are multiplied by constants c_k. Then the general solution $\vec{\mathbf{x}}_h$ of the homogeneous equation $\vec{\mathbf{y}}' = A(t)\vec{\mathbf{y}}$ is given by

$$\vec{\mathbf{x}}_h(t) = c_1\vec{\mathbf{u}}_1(t) + c_2\vec{\mathbf{u}}_2(t) + \cdots + c_n\vec{\mathbf{u}}_n(t).$$

Use partial derivatives on expression $\vec{\mathbf{x}}$ to find the column vectors

$$\vec{\mathbf{u}}_k(t) = \frac{\partial}{\partial c_k}\vec{\mathbf{x}}.$$

This technique isolates the vector components of the homogeneous solution from any form of the general solution, including scalar formulas for the components of $\vec{\mathbf{x}}$. In any case, the general solution $\vec{\mathbf{x}}$ of the linear system $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t)$ is represented by the expression

$$\vec{\mathbf{x}} = c_1\vec{\mathbf{u}}_1(t) + c_2\vec{\mathbf{u}}_2(t) + \cdots + c_n\vec{\mathbf{u}}_n(t) + \vec{\mathbf{x}}_p(t).$$

In this expression, each *assignment* of the constants $c_1, \ldots, c_n$ produces a solution of the nonhomogeneous system, and conversely, each possible solution of the nonhomogeneous system is obtained by a unique *specialization* of the constants $c_1, \ldots, c_n$.

To illustrate the ideas, consider a 3×3 linear system $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t)$ with general solution

$$\vec{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

given in scalar form by the expressions

$$\begin{aligned} x_1 &= c_1e^t + c_2e^{-t} + t, \\ x_2 &= (c_1 + c_2)e^t + c_3e^{2t}, \\ x_3 &= (2c_2 - c_1)e^{-t} + (4c_1 - 2c_3)e^{2t} + 2t. \end{aligned}$$

To find the vector form of the general solution, we take partial derivatives $\vec{\mathbf{u}}_k = \dfrac{\partial\vec{\mathbf{x}}}{\partial c_k}$ with respect to the variable names c_1, c_2, c_3:

$$\vec{\mathbf{u}}_1 = \begin{pmatrix} e^t \\ e^t \\ -e^{-t} + 4e^{2t} \end{pmatrix}, \quad \vec{\mathbf{u}}_2 = \begin{pmatrix} e^{-t} \\ e^t \\ 2e^{-t} \end{pmatrix}, \quad \vec{\mathbf{u}}_3 = \begin{pmatrix} 0 \\ e^{2t} \\ -2e^{2t} \end{pmatrix}.$$

To find $\vec{\mathbf{x}}_p(t)$, set $c_1 = c_2 = c_3 = 0$:

$$\vec{\mathbf{x}}_p(t) = \begin{pmatrix} t \\ 0 \\ 2t \end{pmatrix}.$$

Finally,

$$\begin{aligned}
\vec{\mathbf{x}} &= c_1\vec{\mathbf{u}}_1(t) + c_2\vec{\mathbf{u}}_2(t) + c_3\vec{\mathbf{u}}_3(t) + \vec{\mathbf{x}}_p(t) \\
&= c_1\begin{pmatrix} e^t \\ e^t \\ -e^{-t}+4e^{2t} \end{pmatrix} + c_2\begin{pmatrix} e^{-t} \\ e^t \\ 2e^{-t} \end{pmatrix} + c_3\begin{pmatrix} 0 \\ e^{2t} \\ -2e^{2t} \end{pmatrix} + \begin{pmatrix} t \\ 0 \\ 2t \end{pmatrix}.
\end{aligned}$$

The expression $\vec{\mathbf{x}} = c_1\vec{\mathbf{u}}_1(t) + c_2\vec{\mathbf{u}}_2(t) + c_3\vec{\mathbf{u}}_3(t) + \vec{\mathbf{x}}_p(t)$ satisfies required elements (a) and (b) in the definition of general solution. We will develop now a way to routinely test the uniqueness requirement in (b).

Independence. Constants $c_1, \ldots, c_n$ in the general solution $\vec{\mathbf{x}} = \vec{\mathbf{x}}_h + \vec{\mathbf{x}}_p$ appear exactly in the expression $\vec{\mathbf{x}}_h$, which has the form

$$\vec{\mathbf{x}}_h = c_1\vec{\mathbf{u}}_1 + c_2\vec{\mathbf{u}}_2 + \cdots + c_n\vec{\mathbf{u}}_n.$$

A solution $\vec{\mathbf{x}}$ uniquely determines the constants. In particular, the zero solution of the homogeneous equation is uniquely represented, which can be stated this way:

$$c_1\vec{\mathbf{u}}_1 + c_2\vec{\mathbf{u}}_2 + \cdots + c_n\vec{\mathbf{u}}_n = \vec{\mathbf{0}} \quad \text{implies} \quad c_1 = c_2 = \cdots = c_n = 0.$$

This statement equivalently says that the list of n vector-valued functions $\vec{\mathbf{u}}_1(t), \ldots, \vec{\mathbf{u}}_n(t)$ is **linearly independent**.

It is possible to write down a candidate general solution to some 3×3 linear system $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ via equations like

$$\begin{aligned}
x_1 &= c_1e^t + c_2e^t + c_3e^{2t}, \\
x_2 &= c_1e^t + c_2e^t + 2c_3e^{2t}, \\
x_3 &= c_1e^t + c_2e^t + 4c_3e^{2t}.
\end{aligned}$$

This example was constructed to contain a classic mistake, for purposes of illustration.

How can we detect a mistake, given only that this expression is supposed to represent the general solution? First of all, we can test that $\vec{\mathbf{u}}_1 = \partial\vec{\mathbf{x}}/\partial c_1$, $\vec{\mathbf{u}}_2 = \partial\vec{\mathbf{x}}/\partial c_2$, $\vec{\mathbf{u}}_3 = \partial\vec{\mathbf{x}}/\partial c_3$ are indeed solutions. But to insure the unique representation requirement, the vector functions $\vec{\mathbf{u}}_1$, $\vec{\mathbf{u}}_2$, $\vec{\mathbf{u}}_3$ must be linearly independent. We compute

$$\vec{\mathbf{u}}_1 = \begin{pmatrix} e^t \\ e^t \\ e^t \end{pmatrix}, \quad \vec{\mathbf{u}}_2 = \begin{pmatrix} e^t \\ e^t \\ e^t \end{pmatrix}, \quad \vec{\mathbf{u}}_3 = \begin{pmatrix} e^{2t} \\ 2e^{2t} \\ 4e^{2t} \end{pmatrix}.$$

Therefore, $\vec{\mathbf{u}}_1 = \vec{\mathbf{u}}_2$, which implies that the functions $\vec{\mathbf{u}}_1$, $\vec{\mathbf{u}}_2$, $\vec{\mathbf{u}}_3$ fail to be independent. While is is possible to test independence by a rudimentary test based upon the definition, we prefer the following test due to Norwegian mathematician N. H. Abel (1802-1829).

Theorem 11 (Abel's Formula and the Wronskian)
Let $\vec{\mathbf{x}}_h(t) = c_1\vec{\mathbf{u}}_1(t) + \cdots + c_n\vec{\mathbf{u}}_n(t)$ be a candidate general solution to the equation $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}}$. In particular, the vector functions $\vec{\mathbf{u}}_1(t)$, ..., $\vec{\mathbf{u}}_n(t)$ are solutions of $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}}$. Define the **Wronskian** by

$$w(t) = \det(\langle \vec{\mathbf{u}}_1(t)|\cdots|\vec{\mathbf{u}}_n(t))\rangle.$$

Then **Abel's formula** holds:

$$w(t) = e^{\int_{t_0}^{t} \mathbf{trace}(A(s))ds} w(t_0).^5$$

In particular, $w(t)$ is either everywhere nonzero or everywhere zero, accordingly as $w(t_0) \neq 0$ or $w(t_0) = 0$.

Theorem 12 (Abel's Wronskian Test for Independence)
The vector solutions $\vec{\mathbf{u}}_1$, ..., $\vec{\mathbf{u}}_n$ of $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}}$ are independent if and only if the Wronskian $w(t)$ is nonzero for some $t = t_0$.

Clever use of the point t_0 in Abel's Wronskian test can lead to succinct independence tests. For instance, let

$$\vec{\mathbf{u}}_1 = \begin{pmatrix} e^t \\ e^t \\ e^t \end{pmatrix}, \quad \vec{\mathbf{u}}_2 = \begin{pmatrix} e^t \\ e^t \\ e^t \end{pmatrix}, \quad \vec{\mathbf{u}}_3 = \begin{pmatrix} e^{2t} \\ 2e^{2t} \\ 4e^{2t} \end{pmatrix}.$$

Then $w(t)$ might appear to be complicated, but $w(0)$ is obviously zero because it has two duplicate columns. Therefore, Abel's Wronskian test detects **dependence** of $\vec{\mathbf{u}}_1$, $\vec{\mathbf{u}}_2$, $\vec{\mathbf{u}}_3$.

To illustrate Abel's Wronskian test when it detects independence, consider the column vectors

$$\vec{\mathbf{u}}_1 = \begin{pmatrix} e^t \\ e^t \\ -e^{-t} + 4e^{2t} \end{pmatrix}, \quad \vec{\mathbf{u}}_2 = \begin{pmatrix} e^{-t} \\ e^t \\ 2e^{-t} \end{pmatrix}, \quad \vec{\mathbf{u}}_3 = \begin{pmatrix} 0 \\ e^{2t} \\ -2e^{2t} \end{pmatrix}.$$

At $t = t_0 = 0$, they become the column vectors

$$\vec{\mathbf{u}}_1 = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}, \quad \vec{\mathbf{u}}_2 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \quad \vec{\mathbf{u}}_3 = \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix}.$$

Then $w(0) = \det(\langle \vec{\mathbf{u}}_1(0)|\vec{\mathbf{u}}_2(0)|\vec{\mathbf{u}}_3(0))\rangle = 1$ is nonzero, testing **independence** of $\vec{\mathbf{u}}_1$, $\vec{\mathbf{u}}_2$, $\vec{\mathbf{u}}_3$.

[5] The **trace** of a square matrix is the sum of its diagonal elements. In literature, the formula is called the **Abel-Liouville** formula.

Initial value problems and the rref method. An **initial value problem** is the problem of solving for $\vec{\mathbf{x}}$, given

$$\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}} + \vec{\mathbf{f}}(t), \quad \vec{\mathbf{x}}(t_0) = \vec{\mathbf{x}}_0.$$

If a general solution is known,

$$\vec{\mathbf{x}} = c_1\vec{\mathbf{u}}_1(t) + \cdots + c_n\vec{\mathbf{u}}_n(t) + \vec{\mathbf{x}}_p(t),$$

then the problem of finding $\vec{\mathbf{x}}$ reduces to finding $c_1, \ldots, c_n$ in the relation

$$c_1\vec{\mathbf{u}}_1(t_0) + \cdots + c_n\vec{\mathbf{u}}_n(t_0) + \vec{\mathbf{x}}_p(t_0) = \vec{\mathbf{x}}_0.$$

This is a matrix equation for the unknown constants $c_1, \ldots, c_n$ of the form $B\vec{\mathbf{c}} = \vec{\mathbf{d}}$, where

$$B = \langle \vec{\mathbf{u}}_1(t_0)|\cdots|\vec{\mathbf{u}}_n(t_0)\rangle, \quad \vec{\mathbf{c}} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \quad \vec{\mathbf{d}} = \vec{\mathbf{x}}_0 - \vec{\mathbf{x}}_p(t_0).$$

The **rref**-method applies to find $\vec{\mathbf{c}}$. The method is to perform swap, combination and multiply operations to $C = \langle B|\vec{\mathbf{d}}\rangle$ until $\mathbf{rref}(C) = \langle I|\vec{\mathbf{c}}\rangle$.

To illustrate the method, consider the general solution

$$\begin{array}{rcl} x_1 &=& c_1e^t + c_2e^{-t} + t, \\ x_2 &=& (c_1 + c_2)e^t + c_3e^{2t}, \\ x_3 &=& (2c_2 - c_1)e^{-t} + (4c_1 - 2c_3)e^{2t} + 2t. \end{array}$$

We shall solve for c_1, c_2, c_3, given the initial condition $x_1(0) = 1$, $x_2(0) = 0$, $x_3(0) = -1$. The above relations evaluated at $t = 0$ give the system

$$\begin{array}{rcl} 1 &=& c_1e^0 + c_2e^0 + 0, \\ 0 &=& (c_1 + c_2)e^0 + c_3e^0, \\ -1 &=& (2c_2 - c_1)e^0 + (4c_1 - 2c_3)e^0 + 0. \end{array}$$

In standard scalar form, this is the 3×3 linear system

$$\begin{array}{rcrcrcr} c_1 &+& c_2 & & &=& 1, \\ c_1 &+& c_2 &+& c_3 &=& 0, \\ 3c_1 &+& 2c_2 &-& 2c_3 &=& -1. \end{array}$$

The augmented matrix C, to be reduced to **rref** form, is given by

$$C = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 3 & 2 & -2 & -1 \end{pmatrix}.$$

After the **rref** process is completed, we obtain

$$\mathbf{rref}(C) = \begin{pmatrix} 1 & 0 & 0 & -5 \\ 0 & 1 & 0 & 6 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

From this display, we read off the answer $c_1 = -5$, $c_2 = 6$, $c_3 = -1$ and report the final answer

$$\begin{aligned} x_1 &= -5e^t + 6e^{-t} + t, \\ x_2 &= e^t - e^{2t}, \\ x_3 &= 17e^{-t} - 18e^{2t} + 2t. \end{aligned}$$

Equilibria. An equilibrium point $\vec{\mathbf{x}}_0$ of a linear system $\vec{\mathbf{x}}' = A(t)\vec{\mathbf{x}}$ is a constant solution, $\vec{\mathbf{x}}(t) = \vec{\mathbf{x}}_0$ for all t. Mostly, this makes sense when $A(t)$ is constant, although the definition applies to continuous systems. For a solution $\vec{\mathbf{x}}$ to be constant means $\vec{\mathbf{x}}' = \vec{\mathbf{0}}$, hence all equilibria are determined from the equation

$$A(t)\vec{\mathbf{x}}_0 = \vec{\mathbf{0}} \quad \text{for all } t.$$

This is a homogeneous system of linear algebraic equations to be solved for $\vec{\mathbf{x}}_0$. It is not allowed for the answer $\vec{\mathbf{x}}_0$ to depend on t (if it does, then it is **not** an equilibrium). The theory for a constant matrix $A(t) \equiv A$ says that either $\vec{\mathbf{x}}_0 = \vec{\mathbf{0}}$ is the unique solution or else there are infinitely many answers for $\vec{\mathbf{x}}_0$ (the nullity of A is positive).

11.4 Matrix Exponential

The problem

$$\frac{d}{dt}\vec{\mathbf{x}}(t) = A\vec{\mathbf{x}}(t), \quad \vec{\mathbf{x}}(0) = \vec{\mathbf{x}}_0$$

has a unique solution, according to the Picard-Lindelöf theorem. Solve the problem n times, when $\vec{\mathbf{x}}_0$ equals a column of the identity matrix, and write $\vec{\mathbf{w}}_1(t)$, …, $\vec{\mathbf{w}}_n(t)$ for the n solutions so obtained. Define the **matrix exponential** e^{At} by packaging these n solutions into a matrix:

$$e^{At} \equiv \langle \vec{\mathbf{w}}_1(t) | \dots | \vec{\mathbf{w}}_n(t) \rangle.$$

By construction, any possible solution of $\frac{d}{dt}\vec{\mathbf{x}} = A\vec{\mathbf{x}}$ can be uniquely expressed in terms of the matrix exponential e^{At} by the formula

$$\vec{\mathbf{x}}(t) = e^{At}\vec{\mathbf{x}}(0).$$

Matrix Exponential Identities

Announced here and proved below are various formulas and identities for the matrix exponential e^{At}:

$\frac{d}{dt}\left(e^{At}\right) = Ae^{At}$	Columns satisfy $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$.
$e^{\vec{\mathbf{0}}} = I$	Where $\vec{\mathbf{0}}$ is the zero matrix.
$Be^{At} = e^{At}B$	If $AB = BA$.
$e^{At}e^{Bt} = e^{(A+B)t}$	If $AB = BA$.
$e^{At}e^{As} = e^{A(t+s)}$	Since At and As commute.
$\left(e^{At}\right)^{-1} = e^{-At}$	Equivalently, $e^{At}e^{-At} = I$.
$e^{At} = r_1(t)P_1 + \cdots + r_n(t)P_n$	Putzer's spectral formula — see page 816.
$e^{At} = e^{\lambda_1 t}I + \frac{e^{\lambda_1 t} - e^{\lambda_2 t}}{\lambda_1 - \lambda_2}(A - \lambda_1 I)$	A is 2×2, $\lambda_1 \neq \lambda_2$ real.
$e^{At} = e^{\lambda_1 t}I + te^{\lambda_1 t}(A - \lambda_1 I)$	A is 2×2, $\lambda_1 = \lambda_2$ real.
$e^{At} = e^{at}\cos bt\, I + \frac{e^{at}\sin bt}{b}(A - aI)$	A is 2×2, $\lambda_1 = \overline{\lambda}_2 = a + ib$, $b > 0$.
$e^{At} = \sum_{n=0}^{\infty} A^n \frac{t^n}{n!}$	Picard series. See page 818.
$e^{At} = P^{-1}e^{Jt}P$	Jordan form $J = PAP^{-1}$.

Putzer's Spectral Formula

The spectral formula of Putzer applies to a system $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ to find its general solution. The method uses matrices $P_1, \ldots, P_n$ constructed from A and the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A, matrix multiplication, and the solution $\vec{\mathbf{r}}(t)$ of the first order $n \times n$ initial value problem

$$\vec{\mathbf{r}}'(t) = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & \lambda_2 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \lambda_3 & \cdots & 0 & 0 \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & 1 & \lambda_n \end{pmatrix} \vec{\mathbf{r}}(t), \quad \vec{\mathbf{r}}(0) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The system is solved by first order scalar methods and back-substitution. We will derive the formula separately for the 2×2 case (the one used most often) and the $n \times n$ case.

Spectral Formula 2×2

The general solution of the 2×2 system $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ is given by the formula

$$\vec{\mathbf{x}}(t) = (r_1(t)P_1 + r_2(t)P_2)\,\vec{\mathbf{x}}(0),$$

where r_1, r_2, P_1, P_2 are defined as follows.

The eigenvalues $r = \lambda_1, \lambda_2$ are the two roots of the quadratic equation

$$\det(A - rI) = 0.$$

Define 2×2 matrices P_1, P_2 by the formulas

$$P_1 = I, \quad P_2 = A - \lambda_1 I.$$

The functions $r_1(t)$, $r_2(t)$ are defined by the differential system

$$\left\{ \begin{array}{lcll} r_1' & = & \lambda_1 r_1, & r_1(0) = 1, \\ r_2' & = & \lambda_2 r_2 + r_1, & r_2(0) = 0. \end{array} \right.$$

Proof: The Cayley-Hamilton formula $(A - \lambda_1 I)(A - \lambda_2 I) = \vec{\mathbf{0}}$ is valid for any 2×2 matrix A and the two roots $r = \lambda_1, \lambda_2$ of the determinant equality $\det(A - rI) = 0$. The Cayley-Hamilton formula is the same as $(A - \lambda_2)P_2 = \vec{\mathbf{0}}$, which implies the identity $AP_2 = \lambda_2 P_2$. Compute as follows.

$$\begin{aligned} \vec{\mathbf{x}}'(t) &= (r_1'(t)P_1 + r_2'(t)P_2)\,\vec{\mathbf{x}}(0) \\ &= (\lambda_1 r_1(t)P_1 + r_1(t)P_2 + \lambda_2 r_2(t)P_2)\,\vec{\mathbf{x}}(0) \\ &= (r_1(t)A + \lambda_2 r_2(t)P_2)\,\vec{\mathbf{x}}(0) \\ &= (r_1(t)A + r_2(t)AP_2)\,\vec{\mathbf{x}}(0) \\ &= A\,(r_1(t)I + r_2(t)P_2)\,\vec{\mathbf{x}}(0) \\ &= A\vec{\mathbf{x}}(t). \end{aligned}$$

This proves that $\vec{\mathbf{x}}(t)$ is a solution. Because $\Phi(t) \equiv r_1(t)P_1 + r_2(t)P_2$ satisfies $\Phi(0) = I$, then any possible solution of $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ can be represented by the given formula. The proof is complete.

Real Distinct Eigenvalues. Suppose A is 2×2 having real distinct eigenvalues λ_1, λ_2 and $\vec{\mathbf{x}}(0)$ is real. Then

$$r_1 = e^{\lambda_1 t}, \quad r_2 = \frac{e^{\lambda_1 t} - e^{\lambda_2 T}}{\lambda_1 - \lambda_2}$$

and

$$\vec{\mathbf{x}}(t) = \left(e^{\lambda_1 t} I + \frac{e^{\lambda_1 t} - e^{\lambda_2 t}}{\lambda_1 - \lambda_2}(A - \lambda_1 I) \right) \vec{\mathbf{x}}(0).$$

The matrix exponential formula for real distinct eigenvalues:

$$e^{At} = e^{\lambda_1 t} I + \frac{e^{\lambda_1 t} - e^{\lambda_2 t}}{\lambda_1 - \lambda_2}(A - \lambda_1 I).$$

Real Equal Eigenvalues. Suppose A is 2×2 having real equal eigenvalues $\lambda_1 = \lambda_2$ and $\vec{\mathbf{x}}(0)$ is real. Then $r_1 = e^{\lambda_1 t}$, $r_2 = te^{\lambda_1 t}$ and

$$\vec{\mathbf{x}}(t) = \left(e^{\lambda_1 t} I + te^{\lambda_1 t}(A - \lambda_1 I) \right) \vec{\mathbf{x}}(0).$$

The matrix exponential formula for real equal eigenvalues:

$$e^{At} = e^{\lambda_1 t} I + te^{\lambda_1 t}(A - \lambda_1 I).$$

Complex Eigenvalues. Suppose A is 2×2 having complex eigenvalues $\lambda_1 = a + bi$ with $b > 0$ and $\lambda_2 = a - bi$. If $\vec{\mathbf{x}}(0)$ is real, then a real solution is obtained by taking the real part of the spectral formula. This formula is formally identical to the case of real distinct eigenvalues. Then

$$\begin{aligned}
\mathcal{R}\mathrm{e}(\vec{\mathbf{x}}(t)) &= (\mathcal{R}\mathrm{e}(r_1(t))I + \mathcal{R}\mathrm{e}(r_2(t)(A - \lambda_1 I)))\,\vec{\mathbf{x}}(0) \\
&= \left(\mathcal{R}\mathrm{e}(e^{(a+ib)t})I + \mathcal{R}\mathrm{e}(e^{at}\frac{\sin bt}{b}(A - (a+ib)I)) \right) \vec{\mathbf{x}}(0) \\
&= \left(e^{at}\cos bt\, I + e^{at}\frac{\sin bt}{b}(A - aI)) \right) \vec{\mathbf{x}}(0)
\end{aligned}$$

The matrix exponential formula for complex conjugate eigenvalues:

$$e^{At} = e^{at}\left(\cos bt\, I + \frac{\sin bt}{b}(A - aI)) \right).$$

How to Remember Putzer's 2×2 Formula. The expressions

$$(1)\qquad \begin{aligned} e^{At} &= r_1(t)I + r_2(t)(A - \lambda_1 I), \\ r_1(t) &= e^{\lambda_1 t}, \quad r_2(t) = \frac{e^{\lambda_1 t} - e^{\lambda_2 t}}{\lambda_1 - \lambda_2} \end{aligned}$$

are enough to generate all three formulas. Fraction r_2 is the $d/d\lambda$-Newton quotient for r_1. It has limit $te^{\lambda_1 t}$ as $\lambda_2 \to \lambda_1$, therefore the formula includes the case $\lambda_1 = \lambda_2$ by limiting. If $\lambda_1 = \overline{\lambda_2} = a + ib$ with $b > 0$, then the fraction r_2 is already real, because it has for $z = e^{\lambda_1 t}$ and $w = \lambda_1$ the form

$$r_2(t) = \frac{z - \overline{z}}{w - \overline{w}} = \frac{\sin bt}{b}.$$

Taking real parts of expression (1) gives the complex case formula.

Spectral Formula $n \times n$

The general solution of $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ is given by the formula

$$\vec{\mathbf{x}}(t) = (r_1(t)P_1 + r_2(t)P_2 + \cdots + r_n(t)P_n)\,\vec{\mathbf{x}}(0),$$

where r_1, r_2, …, r_n, P_1, P_2, …, P_n are defined as follows.

The eigenvalues $r = \lambda_1, \ldots, \lambda_n$ are the roots of the polynomial equation

$$\det(A - rI) = 0.$$

Define $n \times n$ matrices P_1, …, P_n by the formulas

$$P_1 = I, \quad P_k = P_{k-1}(A - \lambda_{k-1}I) = \Pi_{j=1}^{k-1}(A - \lambda_j I), \quad k = 2, \ldots, n.$$

The functions $r_1(t)$, …, $r_n(t)$ are defined by the differential system

$$\begin{array}{rcll} r_1' &=& \lambda_1 r_1, & r_1(0) = 1, \\ r_2' &=& \lambda_2 r_2 + r_1, & r_2(0) = 0, \\ &\vdots& & \\ r_n' &=& \lambda_n r_n + r_{n-1}, & r_n(0) = 0. \end{array}$$

Proof: The Cayley-Hamilton formula $(A - \lambda_1 I)\cdots(A - \lambda_n I) = \vec{\mathbf{0}}$ is valid for any $n \times n$ matrix A and the n roots $r = \lambda_1, \ldots, \lambda_n$ of the determinant equality $\det(A - rI) = 0$. Two facts will be used: (1) The Cayley-Hamilton formula implies $AP_n = \lambda_n P_n$; (2) The definition of P_k implies $\lambda_k P_k + P_{k+1} = AP_k$ for $1 \le k \le n-1$. Compute as follows.

$$\begin{aligned} \boxed{\mathbf{1}}\quad \vec{\mathbf{x}}'(t) &= (r_1'(t)P_1 + \cdots + r_n'(t)P_n)\,\vec{\mathbf{x}}(0) \\ \boxed{\mathbf{2}}\quad &= \left(\sum_{k=1}^{n} \lambda_k r_k(t)P_k + \sum_{k=2}^{n} r_{k-1}P_k\right)\vec{\mathbf{x}}(0) \end{aligned}$$

$$\boxed{3} \qquad = \left(\sum_{k=1}^{n-1} \lambda_k r_k(t) P_k + r_n(t)\lambda_n P_n + \sum_{k=1}^{n-1} r_k P_{k+1}\right)\vec{\mathbf{x}}(0)$$

$$\boxed{4} \qquad = \left(\sum_{k=1}^{n-1} r_k(t)(\lambda_k P_k + P_{k+1}) + r_n(t)\lambda_n P_n\right)\vec{\mathbf{x}}(0)$$

$$\boxed{5} \qquad = \left(\sum_{k=1}^{n-1} r_k(t) A P_k + r_n(t) A P_n\right)\vec{\mathbf{x}}(0)$$

$$\boxed{6} \qquad = A\left(\sum_{k=1}^{n} r_k(t) P_k\right)\vec{\mathbf{x}}(0)$$

$$\boxed{7} \qquad = A\vec{\mathbf{x}}(t).$$

Details: $\boxed{1}$ Differentiate the formula for $\vec{\mathbf{x}}(t)$. $\boxed{2}$ Use the differential equations for $r_1,\ldots,r_n$. $\boxed{3}$ Split off the last term from the first sum, then re-index the last sum. $\boxed{4}$ Combine the two sums. $\boxed{5}$ Use the recursion for P_k and the Cayley-Hamilton formula $(A - \lambda_n I)P_n = \vec{\mathbf{0}}$. $\boxed{6}$ Factor out A on the left. $\boxed{7}$ Apply the definition of $\vec{\mathbf{x}}(t)$.

This proves that $\vec{\mathbf{x}}(t)$ is a solution. Because $\Phi(t) \equiv \sum_{k=1}^{n} r_k(t)P_k$ satisfies $\Phi(0) = I$, then any possible solution of $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ can be so represented. The proof is complete.

Proofs of Matrix Exponential Properties

Verify $\left(e^{At}\right)' = Ae^{At}$. Let $\vec{\mathbf{x}}_0$ denote a column of the identity matrix. Define $\vec{\mathbf{x}}(t) = e^{At}\vec{\mathbf{x}}_0$. Then

$$\begin{aligned}\left(e^{At}\right)'\vec{\mathbf{x}}_0 &= \vec{\mathbf{x}}'(t)\\ &= A\vec{\mathbf{x}}(t)\\ &= Ae^{At}\vec{\mathbf{x}}_0.\end{aligned}$$

Because this identity holds for all columns of the identity matrix, then $(e^{At})'$ and Ae^{At} have identical columns, hence we have proved the identity $\left(e^{At}\right)' = Ae^{At}$.

Verify $AB = BA$ implies $Be^{At} = e^{At}B$. Define $\vec{\mathbf{w}}_1(t) = e^{At}B\vec{\mathbf{w}}_0$ and $\vec{\mathbf{w}}_2(t) = Be^{At}\vec{\mathbf{w}}_0$. Calculate $\vec{\mathbf{w}}_1'(t) = A\vec{\mathbf{w}}_1(t)$ and $\vec{\mathbf{w}}_2'(t) = BAe^{At}\vec{\mathbf{w}}_0 = ABe^{At}\vec{\mathbf{w}}_0 = A\vec{\mathbf{w}}_2(t)$, due to $BA = AB$. Because $\vec{\mathbf{w}}_1(0) = \vec{\mathbf{w}}_2(0) = \vec{\mathbf{w}}_0$, then the uniqueness assertion of the Picard-Lindelöf theorem implies that $\vec{\mathbf{w}}_1(t) = \vec{\mathbf{w}}_2(t)$. Because $\vec{\mathbf{w}}_0$ is any vector, then $e^{At}B = Be^{At}$. The proof is complete.

Verify $e^{At}e^{Bt} = e^{(A+B)t}$. Let $\vec{\mathbf{x}}_0$ be a column of the identity matrix. Define $\vec{\mathbf{x}}(t) = e^{At}e^{Bt}\vec{\mathbf{x}}_0$ and $\vec{\mathbf{y}}(t) = e^{(A+B)t}\vec{\mathbf{x}}_0$. We must show that $\vec{\mathbf{x}}(t) = \vec{\mathbf{y}}(t)$ for all t. Define $\vec{\mathbf{u}}(t) = e^{Bt}\vec{\mathbf{x}}_0$. We will apply the result $e^{At}B = Be^{At}$, valid for $BA = AB$. The details:

$$\begin{aligned}\vec{\mathbf{x}}'(t) &= \left(e^{At}\vec{\mathbf{u}}(t)\right)'\\ &= Ae^{At}\vec{\mathbf{u}}(t) + e^{At}\vec{\mathbf{u}}'(t)\\ &= A\vec{\mathbf{x}}(t) + e^{At}B\vec{\mathbf{u}}(t)\\ &= A\vec{\mathbf{x}}(t) + Be^{At}\vec{\mathbf{u}}(t)\\ &= (A+B)\vec{\mathbf{x}}(t).\end{aligned}$$

We also know that $\vec{\mathbf{y}}'(t) = (A+B)\vec{\mathbf{y}}(t)$ and since $\vec{\mathbf{x}}(0) = \vec{\mathbf{y}}(0) = \vec{\mathbf{x}}_0$, then the Picard-Lindelöf theorem implies that $\vec{\mathbf{x}}(t) = \vec{\mathbf{y}}(t)$ for all t. This completes the proof.

Verify $e^{At}e^{As} = e^{A(t+s)}$. Let t be a variable and consider s fixed. Define $\vec{\mathbf{x}}(t) = e^{At}e^{As}\vec{\mathbf{x}}_0$ and $\vec{\mathbf{y}}(t) = e^{A(t+s)}\vec{\mathbf{x}}_0$. Then $\vec{\mathbf{x}}(0) = \vec{\mathbf{y}}(0)$ and both satisfy the differential equation $\vec{\mathbf{u}}'(t) = A\vec{\mathbf{u}}(t)$. By the uniqueness in the Picard-Lindelöf theorem, $\vec{\mathbf{x}}(t) = \vec{\mathbf{y}}(t)$, which implies $e^{At}e^{As} = e^{A(t+s)}$. The proof is complete.

Verify $e^{At} = \sum_{n=0}^{\infty} A^n \frac{t^n}{n!}$. The idea of the proof is to apply Picard iteration. By definition, the columns of e^{At} are vector solutions $\vec{\mathbf{w}}_1(t), \ldots, \vec{\mathbf{w}}_n(t)$ whose values at $t = 0$ are the corresponding columns of the $n \times n$ identity matrix. According to the theory of Picard iterates, a particular iterate is defined by

$$\vec{\mathbf{y}}_{n+1}(t) = \vec{\mathbf{y}}_0 + \int_0^t A\vec{\mathbf{y}}_n(r)dr, \quad n \geq 0.$$

The vector $\vec{\mathbf{y}}_0$ equals some column of the identity matrix. The Picard iterates can be found explicitly, as follows.

$$\begin{aligned}
\vec{\mathbf{y}}_1(t) &= \vec{\mathbf{y}}_0 + \textstyle\int_0^t A\vec{\mathbf{y}}_0 dr \\
&= (I + At)\,\vec{\mathbf{y}}_0, \\
\vec{\mathbf{y}}_2(t) &= \vec{\mathbf{y}}_0 + \textstyle\int_0^t A\vec{\mathbf{y}}_1(r)dr \\
&= \vec{\mathbf{y}}_0 + \textstyle\int_0^t A\,(I + At)\,\vec{\mathbf{y}}_0 dr \\
&= \left(I + At + A^2t^2/2\right)\vec{\mathbf{y}}_0, \\
&\vdots \\
\vec{\mathbf{y}}_n(t) &= \left(I + At + A^2\tfrac{t^2}{2} + \cdots + A^n\tfrac{t^n}{n!}\right)\vec{\mathbf{y}}_0.
\end{aligned}$$

The Picard-Lindelöf theorem implies that for $\vec{\mathbf{y}}_0 =$ column k of the identity matrix,

$$\lim_{n\to\infty} \vec{\mathbf{y}}_n(t) = \vec{\mathbf{w}}_k(t).$$

This being valid for each index k, then the columns of the matrix sum

$$\sum_{m=0}^{N} A^m \frac{t^m}{m!}$$

converge as $N \to \infty$ to $\vec{\mathbf{w}}_1(t), \ldots, \vec{\mathbf{w}}_n(t)$. This implies the matrix identity

$$e^{At} = \sum_{n=0}^{\infty} A^n \frac{t^n}{n!}.$$

The proof is complete.

Computing e^{At}

Theorem 13 (Computing e^{Jt} for J Triangular)
If J is an upper triangular matrix, then a column $\vec{\mathbf{u}}(t)$ of e^{Jt} can be computed by solving the system $\vec{\mathbf{u}}'(t) = J\vec{\mathbf{u}}(t)$, $\vec{\mathbf{u}}(0) = \vec{\mathbf{v}}$, where $\vec{\mathbf{v}}$ is the

corresponding column of the identity matrix. This problem can always be solved by first-order scalar methods of growth-decay theory and the integrating factor method.

Theorem 14 (Exponential of a Diagonal Matrix)
For real or complex constants λ_1, ..., λ_n,

$$e^{\mathbf{diag}(\lambda_1,\ldots,\lambda_n)t} = \mathbf{diag}\left(e^{\lambda_1 t},\ldots,e^{\lambda_n t}\right).$$

Theorem 15 (Block Diagonal Matrix)
If $A = \mathbf{diag}(B_1,\ldots,B_k)$ and each of B_1, ..., B_k is a square matrix, then

$$e^{At} = \mathbf{diag}\left(e^{B_1 t},\ldots,e^{B_k t}\right).$$

Theorem 16 (Complex Exponential)
Given real a, b, then

$$e^{\begin{pmatrix} a & b \\ -b & a\end{pmatrix}t} = e^{at}\begin{pmatrix}\cos bt & \sin bt \\ -\sin bt & \cos bt\end{pmatrix}.$$

Exercises 11.4

Matrix Exponential.

1. **(Picard)** Let A be real 2×2. Write out the two initial value problems which define the columns $\vec{\mathbf{w}}_1(t)$, $\vec{\mathbf{w}}_2(t)$ of e^{At}.

2. **(Picard)** Let A be real 3×3. Write out the three initial value problems which define the columns $\vec{\mathbf{w}}_1(t)$, $\vec{\mathbf{w}}_2(t)$, $\vec{\mathbf{w}}_3(t)$ of e^{At}.

3. **(Definition)** Let A be real 2×2. Show that the solution $\vec{\mathbf{x}}(t) = e^{At}\vec{\mathbf{u}}_0$ satisfies $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ and $\vec{\mathbf{x}}(0) = \vec{\mathbf{u}}_0$.

4. **Definition** Let A be real $n\times n$. Show that the solution $\vec{\mathbf{x}}(t) = e^{At}\vec{\mathbf{x}}(0)$ satisfies $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$.

Matrix Exponential 2×2. Find e^{At} using the formula $e^{At} = \langle\vec{\mathbf{w}}_1|\vec{\mathbf{w}}_2\rangle$ and the corresponding systems $\vec{\mathbf{w}}_1' = A\vec{\mathbf{w}}_1$, $\vec{\mathbf{w}}_1(0) = \begin{pmatrix}1\\0\end{pmatrix}$, $\vec{\mathbf{w}}_2' = A\vec{\mathbf{w}}_2$, $\vec{\mathbf{w}}_2(0) = \begin{pmatrix}0\\1\end{pmatrix}$. In these exercises A is triangular so that first-order methods can solve the systems.

5. $A = \begin{pmatrix}1 & 0\\ 0 & 2\end{pmatrix}$.

6. $A = \begin{pmatrix}-1 & 0\\ 0 & 0\end{pmatrix}$.

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

8. $A = \begin{pmatrix}-1 & 1\\ 0 & 2\end{pmatrix}$.

Matrix Exponential Identities.

9.

10.

11.

12.

13.

14.

Putzer's Spectral Formula.

15.

16.

17.

18.

Spectral Formula 2×2 .

19.

20.

21.

22.

Real Distinct Eigenvalues.

23.

24.

25.

26.

Real Equal Eigenvalues.

27.

28.

29.

30.

Complex Eigenvalues.

31.

32.

33.

34.

How to Remember Putzer's 2×2 Formula.

35.

36.

37.

38.

Spectral Formula $n \times n$.

39.

40.

41.

42.

43.

44.

45.

46.

Proofs of Matrix Exponential Properties.

47.

48.

49.

50.

Computing e^{At}.

51.

52.

53.

54.

11.5 The Eigenanalysis Method

The general solution $\vec{\mathbf{x}}(t) = e^{At}\,\vec{\mathbf{x}}(0)$ of the linear system

$$\frac{d}{dt}\vec{\mathbf{x}}(t) = A\vec{\mathbf{x}}(t)$$

can be obtained entirely by eigenanalysis of the matrix A, which involves finding all eigenpairs. The expected case is when the $n \times n$ matrix A has n independent eigenvectors in its list of eigenpairs

$$(\lambda_1, \vec{\mathbf{v}}_1), \quad (\lambda_2, \vec{\mathbf{v}}_2), \quad \ldots, \quad (\lambda_n, \vec{\mathbf{v}}_n).$$

It is not required that the eigenvalues $\lambda_1, \ldots, \lambda_n$ be distinct. The eigenvalues can be real or complex.

The Eigenanalysis Method for a 2×2 Matrix

Suppose that A is 2×2 real and has eigenpairs

$$(\lambda_1, \vec{\mathbf{v}}_1), \quad (\lambda_2, \vec{\mathbf{v}}_2),$$

with $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$ independent. The eigenvalues λ_1, λ_2 can be both real. Also, they can be a complex conjugate pair $\lambda_1 = \overline{\lambda}_2 = a + ib$ with $b > 0$.

It will be shown that the general solution of $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ can be written as

$$\vec{\mathbf{x}}(t) = c_1 e^{\lambda_1 t}\vec{\mathbf{v}}_1 + c_2 e^{\lambda_2 t}\vec{\mathbf{v}}_2.$$

The details:

$$\begin{aligned}
\vec{\mathbf{x}}' &= c_1(e^{\lambda_1 t})'\vec{\mathbf{v}}_1 + c_2(e^{\lambda_2 t})'\vec{\mathbf{v}}_2 && \textsf{Differentiate the formula for } \vec{\mathbf{x}}. \\
&= c_1 e^{\lambda_1 t}\lambda_1\vec{\mathbf{v}}_1 + c_2 e^{\lambda_2 t}\lambda_2\vec{\mathbf{v}}_2 && \\
&= c_1 e^{\lambda_1 t}A\vec{\mathbf{v}}_1 + c_2 e^{\lambda_2 t}A\vec{\mathbf{v}}_2 && \textsf{Use } \lambda_1\vec{\mathbf{v}}_1 = A\vec{\mathbf{v}}_1,\ \lambda_2\vec{\mathbf{v}}_2 = A\vec{\mathbf{v}}_2. \\
&= A\left(c_1 e^{\lambda_1 t}\vec{\mathbf{v}}_1 + c_2 e^{\lambda_2 t}\vec{\mathbf{v}}_2\right) && \textsf{Factor } A \textsf{ left.} \\
&= A\vec{\mathbf{x}} && \textsf{Definition of } \vec{\mathbf{x}}.
\end{aligned}$$

Let's rewrite the solution $\vec{\mathbf{x}}$ in the vector-matrix form

$$\vec{\mathbf{x}}(t) = \langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 \rangle \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

Because eigenvectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$ are assumed independent, then $\langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 \rangle$ is invertible and setting $t = 0$ in the previous display gives

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 \rangle^{-1}\vec{\mathbf{x}}(0).$$

Because c_1, c_2 can be chosen to produce any initial condition $\vec{\mathbf{x}}(0)$, then $\vec{\mathbf{x}}(t)$ is the *general solution* of the system $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$.

The general solution expressed as $\vec{\mathbf{x}}(t) = e^{At}\,\vec{\mathbf{x}}(0)$ leads to the exponential matrix relation

$$e^{At} = \langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 \rangle \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix} \langle \vec{\mathbf{v}}_1 | \vec{\mathbf{v}}_2 \rangle^{-1}.$$

The formula is immediately useful when the eigenpairs are real.

Complex conjugate eigenvalues. Assume $\lambda_2 = \overline{\lambda}_1$ and λ_1 not real. Eigenpair $(\lambda_2, \vec{\mathbf{v}}_2)$ is never computed or used, because $A\vec{\mathbf{v}}_1 = \lambda_1 \vec{\mathbf{v}}_1$ implies $A\overline{\vec{\mathbf{v}}}_1 = \overline{\lambda}_1 \overline{\vec{\mathbf{v}}}_1$, which implies λ_2 $(= \overline{\lambda}_1)$ has eigenvector $\vec{\mathbf{v}}_2 = \overline{\vec{\mathbf{v}}}_1$.

If A is real, then e^{At} is real, and taking real parts across the formula for e^{At} will give a real formula. Due to the unpleasantness of the complex algebra, we will report the answer found, which is *real*, and then justify it with minimal use of complex numbers.

Define for eigenpair $(\lambda_1, \vec{\mathbf{v}}_1)$ symbols a, b, P as follows:

$$\lambda_1 = a + ib, \quad b > 0, \quad P = \langle\, \mathcal{R}\mathrm{e}(\vec{\mathbf{v}}_1) | \mathcal{I}\mathrm{m}(\vec{\mathbf{v}}_1) \rangle.$$

Then

$$e^{At} = e^{at} P \begin{pmatrix} \cos bt & \sin bt \\ -\sin bt & \cos bt \end{pmatrix} P^{-1}. \tag{1}$$

Justification of (1). The formula is established by showing that the matrix $\Phi(t)$ on the right satisfies $\Phi(0) = I$ and $\Phi' = A\Phi$. Then by definition, $e^{At} = \Phi(t)$. For exposition, let

$$R(t) = e^{at} \begin{pmatrix} \cos bt & \sin bt \\ -\sin bt & \cos bt \end{pmatrix}, \quad \Phi(t) = PR(t)P^{-1}.$$

The identity $\Phi(0) = I$ verified as follows.

$$\begin{aligned} \Phi(0) &= PR(0)P^{-1} \\ &= Pe^0 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} P^{-1} \\ &= I \end{aligned}$$

Write $\lambda_1 = a + ib$ and $\vec{\mathbf{v}}_1 = \mathcal{R}\mathrm{e}(\vec{\mathbf{v}}_1) + i\,\mathcal{I}\mathrm{m}(\vec{\mathbf{v}}_1)$. The expansion of eigenpair relation $A\vec{\mathbf{v}}_1 = \lambda_1 \vec{\mathbf{v}}_1$ into real and imaginary parts gives the relation

$$A\left(\mathcal{R}\mathrm{e}(\vec{\mathbf{v}}_1) + i\,\mathcal{I}\mathrm{m}(\vec{\mathbf{v}}_1)\right) = (a + ib)\left(\mathcal{R}\mathrm{e}(\vec{\mathbf{v}}_1) + i\,\mathcal{I}\mathrm{m}(\vec{\mathbf{v}}_1)\right),$$

which shows that

$$AP = P \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Then

$$\begin{aligned}\Phi'(t)\Phi^{-1}(t) &= PR'(t)P^{-1}PR^{-1}(t)P^{-1}\\ &= PR'(t)R^{-1}(t)P^{-1}\\ &= P\left(aI+\begin{pmatrix}0 & b\\ -b & 0\end{pmatrix}\right)P^{-1}\\ &= P\begin{pmatrix}a & b\\ -b & a\end{pmatrix}P^{-1}\\ &= A\end{aligned}$$

The proof of $\Phi'(t) = A\Phi(t)$ is complete.

The formula for e^{At} implies that the general solution in this special case is

$$\vec{\mathbf{x}}(t) = e^{at}\langle\,\mathcal{R}\mathrm{e}(\vec{\mathbf{v}}_1)|\,\mathcal{I}\mathrm{m}(\vec{\mathbf{v}}_1)\rangle\begin{pmatrix}\cos bt & \sin bt\\ -\sin bt & \cos bt\end{pmatrix}\begin{pmatrix}c_1\\ c_2\end{pmatrix}.$$

The values c_1, c_2 are related to the initial condition $\vec{\mathbf{x}}(0)$ by the matrix identity

$$\begin{pmatrix}c_1\\ c_2\end{pmatrix} = \langle\,\mathcal{R}\mathrm{e}(\vec{\mathbf{v}}_1)|\,\mathcal{I}\mathrm{m}(\vec{\mathbf{v}}_1))^{-1}\vec{\mathbf{x}}(0\rangle.$$

The Eigenanalysis Method for a 3×3 Matrix

Suppose that A is 3×3 real and has eigenpairs

$$(\lambda_1, \vec{\mathbf{v}}_1), \quad (\lambda_2, \vec{\mathbf{v}}_2), \quad (\lambda_3, \vec{\mathbf{v}}_3),$$

with $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ independent. The eigenvalues λ_1, λ_2, λ_3 can be all real. Also, there can be one real eigenvalue λ_3 and a complex conjugate pair of eigenvalues $\lambda_1 = \overline{\lambda}_2 = a + ib$ with $b > 0$.

The general solution of $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ can be written as

$$\vec{\mathbf{x}}(t) = c_1 e^{\lambda_1 t}\vec{\mathbf{v}}_1 + c_2 e^{\lambda_2 t}\vec{\mathbf{v}}_2 + c_3 e^{\lambda_3 t}\vec{\mathbf{v}}_3.$$

The details, which parallel the 2×2 details, are left as an exercise for the reader.

The solution $\vec{\mathbf{x}}$ is written in vector-matrix form

$$\vec{\mathbf{x}}(t) = \langle\vec{\mathbf{v}}_1|\vec{\mathbf{v}}_2, \vec{\mathbf{v}}_3\rangle\begin{pmatrix}e^{\lambda_1 t} & 0 & 0\\ 0 & e^{\lambda_2 t} & 0\\ 0 & 0 & e^{\lambda_3 t}\end{pmatrix}\begin{pmatrix}c_1\\ c_2\\ c_3\end{pmatrix}.$$

Because the three eigenvectors $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$, $\vec{\mathbf{v}}_3$ are assumed independent, then $\langle\vec{\mathbf{v}}_1|\vec{\mathbf{v}}_2|\vec{\mathbf{v}}_3\rangle$ is invertible. Setting $t = 0$ in the previous display gives

$$\begin{pmatrix}c_1\\ c_2\\ c_2\end{pmatrix} = \langle\vec{\mathbf{v}}_1|\vec{\mathbf{v}}_2|\vec{\mathbf{v}}_3\rangle^{-1}\vec{\mathbf{x}}(0).$$

Constants c_1, c_2, c_3 can be chosen to produce any initial condition $\vec{\mathbf{x}}(0)$, therefore $\vec{\mathbf{x}}(t)$ is the *general solution* of the 3×3 system $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$. There is a corresponding exponential matrix relation

$$e^{At} = \langle \vec{\mathbf{v}}_1|\vec{\mathbf{v}}_2|\vec{\mathbf{v}}_3\rangle \begin{pmatrix} e^{\lambda_1 t} & 0 & 0 \\ 0 & e^{\lambda_2 t} & 0 \\ 0 & 0 & e^{\lambda_3 t} \end{pmatrix} \langle \vec{\mathbf{v}}_1|\vec{\mathbf{v}}_2|\vec{\mathbf{v}}_3\rangle^{-1}.$$

This formula is normally used when the eigenpairs are real. When there is a complex conjugate pair of eigenvalues $\lambda_1 = \overline{\lambda}_2 = a + ib$, $b > 0$, then as was shown in the 2×2 case it is possible to extract a real solution $\vec{\mathbf{x}}$ from the complex formula and report a real form for the exponential matrix:

$$\begin{aligned} e^{At} &= P \begin{pmatrix} e^{at}\cos bt & e^{at}\sin bt & 0 \\ -e^{at}\sin bt & e^{at}\cos bt & 0 \\ 0 & 0 & e^{\lambda_3 t} \end{pmatrix} P^{-1}, \\ P &= \langle\, \mathcal{R}\mathrm{e}(\vec{\mathbf{v}}_1)|\mathcal{I}\mathrm{m}(\vec{\mathbf{v}}_1)|\vec{\mathbf{v}}_3\rangle. \end{aligned}$$

The Eigenanalysis Method for an $n \times n$ Matrix

The general solution formula and the formula for e^{At} generalize easily from the 2×2 and 3×3 cases to the general case of an $n \times n$ matrix.

Theorem 17 (The Eigenanalysis Method)
Let the $n \times n$ real matrix A have eigenpairs

$$(\lambda_1, \vec{\mathbf{v}}_1), \quad (\lambda_2, \vec{\mathbf{v}}_2), \quad \ldots, \quad (\lambda_n, \vec{\mathbf{v}}_n),$$

with n independent eigenvectors $\vec{\mathbf{v}}_1$, ..., $\vec{\mathbf{v}}_n$. Then the general solution of the linear system $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$ is given by

$$\vec{\mathbf{x}}(t) = c_1\vec{\mathbf{v}}_1 e^{\lambda_1 t} + c_2\vec{\mathbf{v}}_2 e^{\lambda_2 t} + \cdots + c_n\vec{\mathbf{v}}_n e^{\lambda_n t}.$$

The vector-matrix form of the general solution is

$$\vec{\mathbf{x}}(t) = \langle \vec{\mathbf{v}}_1|\cdots|\vec{\mathbf{v}}_n\rangle\, \textbf{diag}(e^{\lambda_1 t}, \ldots, e^{\lambda_n t}) \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}.$$

This form is real provided all eigenvalues are real. A real form can be made from a complex form by following the example of a 3×3 matrix A. The plan is to list all complex eigenvalues first, in pairs, λ_1, $\overline{\lambda}_1$, ..., λ_p, $\overline{\lambda}_p$. Then the real eigenvalues $r_1, \ldots, r_q$ are listed, $2p + q = n$. Define

$$P = \langle\, \mathcal{R}\mathrm{e}(\vec{\mathbf{v}}_1)|\mathcal{I}\mathrm{m}(\vec{\mathbf{v}}_1)|\ldots|\, \mathcal{R}\mathrm{e}(\vec{\mathbf{v}}_{2p-1})|\mathcal{I}\mathrm{m}(\vec{\mathbf{v}}_{2p-1})|\vec{\mathbf{v}}_{2p+1}|\cdots|\vec{\mathbf{v}}_n\rangle,$$

$$R_\lambda(t) = e^{at}\begin{pmatrix} \cos bt & \sin bt \\ -\sin bt & \cos bt \end{pmatrix}, \quad \text{where} \quad \lambda + a + ib, \quad b > 0.$$

Then the real vector-matrix form of the general solution is

$$\vec{\mathbf{x}}(t) = P\,\mathbf{diag}(R_{\lambda_1}(t), \ldots, R_{\lambda_p}(t), e^{r_1 t}, \ldots, e^{r_q t})\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$$

and

$$e^{At} = P\,\mathbf{diag}(R_{\lambda_1}(t), \ldots, R_{\lambda_p}(t), e^{r_1 t}, \ldots, e^{r_q t})P^{-1}.$$

Remark on Euler Atoms. If the characteristic equation is $(\lambda-1)^3 = 0$ and there are three independent eigenvectors, then the general solution $\vec{\mathbf{x}}(t) = c_1 e^{\lambda_1 t}\vec{\mathbf{v}}_1 + c_2 e^{\lambda_2 t}\vec{\mathbf{v}}_2 + c_3 e^{\lambda_3 t}\vec{\mathbf{v}}_3$ contains no terms with te^t nor t^2e^t. Our intuition from $(\lambda-1)^3 = 0$ is that solution components should be linear combinations of e^t, te^t, t^2e^t. How is that possible? The answer is contained in the linear combination $2e^t + 0te^t + 0t^2e^t$: it is indeed a linear combination of the Euler atoms.

Spectral Theory Methods

The simplicity of Putzer's spectral method for computing e^{At} is appreciated, but we also recognize that the literature has an algorithm to compute e^{At}, devoid of differential equations, which is of fundamental importance in linear algebra. The parallel algorithm computes e^{At} directly from the eigenvalues λ_j of A and certain products of the nilpotent matrices $A - \lambda_j I$. Called **spectral formulas**, they can be implemented in a numerical laboratory or computer algebra system, in order to efficiently compute e^{At}, even in the case of multiple eigenvalues.

Theorem 18 (Computing e^{At} for Simple Eigenvalues)
Let the $n \times n$ matrix A have n simple eigenvalues λ_1, ..., λ_n (possibly complex) and define constant matrices $\vec{\mathbf{Q}}_1$, ..., $\vec{\mathbf{Q}}_n$ by the formulas

$$\vec{\mathbf{Q}}_j = \Pi_{i \neq j}\frac{A - \lambda_i I}{\lambda_j - \lambda_i}, \quad j = 1, \ldots, n.$$

Then

$$e^{At} = e^{\lambda_1 t}\vec{\mathbf{Q}}_1 + \cdots + e^{\lambda_n t}\vec{\mathbf{Q}}_n.$$

Theorem 19 (Computing e^{At} for Multiple Eigenvalues)
Let the $n \times n$ matrix A have k distinct eigenvalues λ_1, ..., λ_k of algebraic multiplicities m_1, ..., m_k. Let $p(\lambda) = \det(A - \lambda I)$ and define polynomials $a_1(\lambda)$, ..., $a_k(\lambda)$ by the partial fraction identity

$$\frac{1}{p(\lambda)} = \frac{a_1(\lambda)}{(\lambda - \lambda_1)^{m_1}} + \cdots + \frac{a_k(\lambda)}{(\lambda - \lambda_k)^{m_k}}.$$

Define constant matrices $\vec{\mathbf{Q}}_1$, ..., $\vec{\mathbf{Q}}_k$ by the formulas

$$\vec{\mathbf{Q}}_j = a_j(A)\Pi_{i\neq j}(A-\lambda_i I)^{m_i}, \quad j=1,\ldots,k.$$

Then

$$e^{At} = \sum_{i=1}^{k} e^{\lambda_i t}\,\vec{\mathbf{Q}}_i \sum_{j=0}^{m_i-1}(A-\lambda_i I)^j\,\frac{t^j}{j!}. \tag{2}$$

Proof: Let $\vec{\mathbf{N}}_i = \vec{\mathbf{Q}}_i(A-\lambda_i I)$, $1\le i\le k$. We first prove

Lemma 1 (Properties)
1. $\vec{\mathbf{Q}}_1+\cdots+\vec{\mathbf{Q}}_k = I$,
2. $\vec{\mathbf{Q}}_i\vec{\mathbf{Q}}_i = \vec{\mathbf{Q}}_i$,
3. $\vec{\mathbf{Q}}_i\vec{\mathbf{Q}}_j = \vec{\mathbf{0}}$ for $i\neq j$,
4. $\vec{\mathbf{N}}_i\vec{\mathbf{N}}_j = \vec{\mathbf{0}}$ for $i\neq j$,
5. $\vec{\mathbf{N}}_i^{m_i} = \vec{\mathbf{0}}$,
6. $A = \sum_{i=1}^k(\lambda_i\vec{\mathbf{Q}}_i + \vec{\mathbf{N}}_i)$.

The proof of **1** follows from clearing fractions in the partial fraction expansion of $1/p(\lambda)$:

$$1 = \sum_{i=1}^{k} a_i(\lambda)\frac{p(\lambda)}{(\lambda-\lambda_i)^{m_i}}.$$

The **projection property 2** follows by multiplication of identity **1** by $\vec{\mathbf{Q}}_i$ and then using **2**.

The proof of **3** starts by observing that $\vec{\mathbf{Q}}_i$ and $\vec{\mathbf{Q}}_j$ together contain all the factors of $p(A)$, therefore $\vec{\mathbf{Q}}_i\vec{\mathbf{Q}}_j = q(A)p(A)$ for some polynomial q. The Cayley-Hamilton theorem $p(A)=\vec{\mathbf{0}}$ finishes the proof.

To prove **4**, write $\vec{\mathbf{N}}_i\vec{\mathbf{N}}_j = (A-\lambda_i I)(A-\lambda_j I)\vec{\mathbf{Q}}_i\vec{\mathbf{Q}}_j$ and apply **3**.

To prove **5**, use $\vec{\mathbf{Q}}_i^{m_i} = \vec{\mathbf{Q}}_i$ (from **2**) to write $\vec{\mathbf{N}}_i^{m_i} = (A-\lambda_i I)^{m_i}\vec{\mathbf{Q}}_i = p(A) = \vec{\mathbf{0}}$.

To prove **6**, multiply **1** by A and rearrange as follows:

$$\begin{array}{rcl} A &=& \sum_{i=1}^k A\vec{\mathbf{Q}}_i \\ &=& \sum_{i=1}^k \lambda_i\vec{\mathbf{Q}}_i + (A-\lambda_i I)\vec{\mathbf{Q}}_i \\ &=& \sum_{i=1}^k \lambda_i\vec{\mathbf{Q}}_i + \vec{\mathbf{N}}_i \end{array}$$

To prove (2), multiply **1** by e^{At} and compute as follows:

$$\begin{aligned} e^{At} &= \textstyle\sum_{i=1}^k \vec{\mathbf{Q}}_i e^{At} \\ &= \textstyle\sum_{i=1}^k \vec{\mathbf{Q}}_i e^{\lambda_i It + (A-\lambda_i I)t} \\ &= \textstyle\sum_{i=1}^k \vec{\mathbf{Q}}_i e^{\lambda_i t}e^{(A-\lambda_i I)t} \\ &= \textstyle\sum_{i=1}^k \vec{\mathbf{Q}}_i e^{\lambda_i t}e^{\vec{\mathbf{Q}}_i(A-\lambda_i I)t} \\ &= \textstyle\sum_{i=1}^k \vec{\mathbf{Q}}_i e^{\lambda_i t}e^{\vec{\mathbf{N}}_i t} \\ &= \textstyle\sum_{i=1}^k \vec{\mathbf{Q}}_i e^{\lambda_i t}\sum_{j=0}^{m_1-1}(A-\lambda_i I)^j\,\frac{t^j}{j!} \end{aligned}$$

Solving Planar Systems $\vec{\mathbf{x}}'(t) = A\vec{\mathbf{x}}(t)$

A 2×2 real system $\vec{\mathbf{x}}'(t) = A\vec{\mathbf{x}}(t)$ can be solved in terms of the roots of the characteristic equation $\det(A - \lambda I) = 0$ and the real matrix A.

Theorem 20 (Planar System, Putzer's Spectral Formula)
Consider the real planar system $\vec{\mathbf{x}}'(t) = A\vec{\mathbf{x}}(t)$. Let λ_1, λ_2 be the roots of the characteristic equation $\det(A - \lambda I) = 0$. The real general solution $\vec{\mathbf{x}}(t)$ is given by the formula

$$\vec{\mathbf{x}}(t) = e^{At}\vec{\mathbf{x}}(0)$$

where the 2×2 exponential matrix e^{At} is given as follows.

Real $\lambda_1 \neq \lambda_2$ $\qquad$ $e^{At} = e^{\lambda_1 t} I + \dfrac{e^{\lambda_2 t} - e^{\lambda_1 t}}{\lambda_2 - \lambda_1}(A - \lambda_1 I).$

Real $\lambda_1 = \lambda_2$ $\qquad$ $e^{At} = e^{\lambda_1 t} I + t e^{\lambda_1 t}(A - \lambda_1 I).$

Complex $\lambda_1 = \overline{\lambda}_2$,
$\lambda_1 = a + bi$, $b > 0$ $\qquad$ $e^{At} = e^{at}\cos bt\, I + \dfrac{e^{at}\sin(bt)}{b}(A - aI).$

Proof: The formulas are from Putzer's algorithm, or equivalently, from the spectral formulas, with rearranged terms. The complex case is formally the real part of the distinct root case when $\lambda_2 = \overline{\lambda}_1$. The **spectral formula** is the analog of the second order equation formulas, Theorem 1 in Chapter 5.

Illustrations. Typical cases are represented by the following 2×2 matrices A, which correspond to roots λ_1, λ_2 of the characteristic equation $\det(A - \lambda I) = 0$ which are real distinct, real double or complex conjugate. The solution $\vec{\mathbf{x}}(t) = e^{At}\vec{\mathbf{x}}(0)$ is given here in two forms, by writing e^{At} using $\boxed{1}$ a **spectral formula** and $\boxed{2}$ Putzer's **spectral formula**.

$\lambda_1 = 5$, $\lambda_2 = 2$

$$A = \begin{pmatrix} -1 & 3 \\ -6 & 8 \end{pmatrix}$$

Real distinct roots.

$$\boxed{1}\ e^{At} = \frac{e^{5t}}{3}\begin{pmatrix} -3 & 3 \\ -6 & 6 \end{pmatrix} + \frac{e^{2t}}{-3}\begin{pmatrix} -6 & 3 \\ -6 & 3 \end{pmatrix}$$

$$\boxed{2}\ e^{At} = e^{5t} I + \frac{e^{2t} - e^{5t}}{2 - 5}\begin{pmatrix} -6 & 3 \\ -6 & 3 \end{pmatrix}$$

$\lambda_1 = \lambda_2 = 3$

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix}$$

Real double root.

$$\boxed{1}\ e^{At} = e^{3t}\left(I + t\begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}\right)$$

$$\boxed{2}\ e^{At} = e^{3t} I + t e^{3t}\begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}$$

$\lambda_1 = \overline{\lambda}_2 = 2 + 3i$ Complex conjugate roots.

$A = \begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}$

$$\boxed{\mathbf{1}}\ e^{At} = 2\,\mathcal{R}\mathrm{e}\left(\frac{e^{2t+3it}}{2(3i)}\begin{pmatrix} 3i & 3 \\ -3 & 3i \end{pmatrix}\right)$$

$$\boxed{\mathbf{2}}\ e^{At} = e^{2t}\cos 3tI + \frac{e^{2t}\sin 3t}{3}\begin{pmatrix} 0 & 3 \\ -3 & 0 \end{pmatrix}$$

The complex example is typical for real $n \times n$ matrices A with a complex conjugate pair of eigenvalues $\lambda_1 = \overline{\lambda}_2$. Then $\vec{\mathbf{Q}}_2 = \overline{\vec{\mathbf{Q}}}_1$. The result is that λ_2 is not used and we write instead a simpler expression using the college algebra equality $z + \overline{z} = 2\,\mathcal{R}\mathrm{e}(z)$:

$$e^{\lambda_1 t}\vec{\mathbf{Q}}_1 + e^{\lambda_2 t}\vec{\mathbf{Q}}_2 = 2\,\mathcal{R}\mathrm{e}\left(e^{\lambda_1 t}\vec{\mathbf{Q}}_1\right).$$

This observation explains why e^{At} is real when A is real, by pairing complex conjugate eigenvalues in the spectral formula.

11.6 Jordan Form and Eigenanalysis

Generalized Eigenanalysis

The main result is **Jordan's decomposition**

$$A = PJP^{-1},$$

valid for any real or complex square matrix A. We describe here how to compute the invertible matrix P of generalized eigenvectors and the upper triangular matrix J, called a **Jordan form** of A.

Jordan block. An $m \times m$ upper triangular matrix $B(\lambda, m)$ is called a **Jordan block** provided all m diagonal elements are the same eigenvalue λ and all super-diagonal elements are one:

$$B(\lambda, m) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix} \quad (m \times m \text{ matrix})$$

Jordan form. Given an $n \times n$ matrix A, a **Jordan form** J for A is a block diagonal matrix

$$J = \textbf{diag}(B(\lambda_1, m_1), B(\lambda_2, m_2), \ldots, B(\lambda_k, m_k)),$$

where $\lambda_1, \ldots, \lambda_k$ are eigenvalues of A (duplicates possible) and $m_1 + \cdots + m_k = n$. The eigenvalues of J are on the diagonal of J and J has exactly k eigenpairs. If $k < n$, then J is non-diagonalizable. Relation $AP = PJ$ implies A has exactly k eigenpairs and A fails to be diagonalizable for $k < n$.

The relation $A = PJP^{-1}$ is called a **Jordan decomposition** of A. Invertible matrix P is called the **matrix of generalized eigenvectors** of A. It defines a coordinate system $\vec{\mathbf{x}} = P\vec{\mathbf{y}}$ in which the vector function $\vec{\mathbf{x}} \to A\vec{\mathbf{x}}$ is transformed to the simpler vector function $\vec{\mathbf{y}} \to J\vec{\mathbf{y}}$.

If equal eigenvalues are adjacent in J, then Jordan blocks with equal diagonal entries will be adjacent. Zeros can appear on the super-diagonal of J, because adjacent Jordan blocks join on the super-diagonal with a zero. A complete specification of how to build J from A appears below.

Decoding a Jordan Decomposition $A = PJP^{-1}$. If J is a single Jordan block, $J = B(\lambda, m)$, then $P = \langle \vec{\mathbf{v}}_1 | \ldots | \vec{\mathbf{v}}_m \rangle$ and $AP = PJ$

means

$$\begin{array}{rcl} A\vec{\mathbf{v}}_1 & = & \lambda\vec{\mathbf{v}}_1, \\ A\vec{\mathbf{v}}_2 & = & \lambda\vec{\mathbf{v}}_2 + \vec{\mathbf{v}}_1, \\ \vdots & \vdots & \quad\vdots \\ A\vec{\mathbf{v}}_m & = & \lambda\vec{\mathbf{v}}_m + \vec{\mathbf{v}}_{m-1}. \end{array}$$

The exploded view of the relation $AP = PB(\lambda, m)$ is called a **Jordan chain**. The formulas can be compacted via matrix $N = A - \lambda I$ into the recursion

$$N\vec{\mathbf{v}}_1 = \vec{\mathbf{0}}, \quad N\vec{\mathbf{v}}_2 = \vec{\mathbf{v}}_1, \ldots, N\vec{\mathbf{v}}_m = \vec{\mathbf{v}}_{m-1}.$$

The first vector $\vec{\mathbf{v}}_1$ is an eigenvector. The remaining vectors $\vec{\mathbf{v}}_2$, ..., $\vec{\mathbf{v}}_m$ are **not eigenvectors**, they are called **generalized eigenvectors**. A similar formula can be written for each distinct eigenvalue of a matrix A. The collection of formulas are called **Jordan chain relations**. A given eigenvalue may appear multiple times in the chain relations, due to the appearance of two or more Jordan blocks with the same eigenvalue.

Theorem 21 (Jordan Decomposition)
Every $n \times n$ matrix A has a Jordan decomposition $A = PJP^{-1}$.

Proof: The result holds by default for 1×1 matrices. Assume the result holds for all $k \times k$ matrices, $k < n$. The proof proceeds by induction on n.

The induction assumes that for any $k \times k$ matrix A, there is a Jordan decomposition $A = PJP^{-1}$. Then the columns of P satisfy Jordan chain relations

$$A\vec{\mathbf{x}}_i^j = \lambda_i\vec{\mathbf{x}}_i^j + \vec{\mathbf{x}}_i^{j-1}, \quad j > 1, \quad A\vec{\mathbf{x}}_i^1 = \lambda_i\vec{\mathbf{x}}_i^1.$$

Conversely, if the Jordan chain relations are satisfied for k independent vectors $\{\vec{\mathbf{x}}_i^j\}$, then the vectors form the columns of an invertible matrix P such that $A = PJP^{-1}$ with J in Jordan form. The induction step centers upon producing the chain relations and proving that the n vectors are independent.

Let B be $n \times n$ and λ_0 an eigenvalue of B. The Jordan chain relations hold for $A = B$ if and only if they hold for $A = B - \lambda_0 I$. Without loss of generality, we can assume 0 is an eigenvalue of B.

Because B has 0 as an eigenvalue, then $p = \dim(\mathbf{kernel}(B)) > 0$ and $k = \dim(\mathbf{Image}(B)) < n$, with $p + k = n$. If $k = 0$, then $B = 0$, which is a Jordan form, and there is nothing to prove. Assume henceforth p and k positive. Let $S = \langle\, \mathbf{col}(B, i_1) | \ldots | \,\mathbf{col}(B, i_k)\rangle$ denote the matrix of pivot columns $i_1, \ldots, i_k$ of B. The pivot columns are known to span $\mathbf{Image}(B)$. Let A be the $k \times k$ basis representation matrix defined by the equation $BS = SA$, or equivalently, $B\,\mathbf{col}(S, j) = \sum_{i=1}^k a_{ij}\,\mathbf{col}(S, i)$. The induction hypothesis applied to A implies there is a basis of k-vectors satisfying Jordan chain relations

$$A\vec{\mathbf{x}}_i^j = \lambda_i\vec{\mathbf{x}}_i^j + \vec{\mathbf{x}}_i^{j-1}, \quad j > 1, \quad A\vec{\mathbf{x}}_i^1 = \lambda_i\vec{\mathbf{x}}_i^1.$$

The values λ_i, $i = 1, \ldots, p$, are the distinct eigenvalues of A. Apply S to these equations to obtain for the n-vectors $\vec{\mathbf{y}}_i^j = S\vec{\mathbf{x}}_i^j$ the Jordan chain relations

$$B\vec{\mathbf{y}}_i^j = \lambda_i\vec{\mathbf{y}}_i^j + \vec{\mathbf{y}}_i^{j-1}, \quad j > 1, \quad B\vec{\mathbf{y}}_i^1 = \lambda_i\vec{\mathbf{y}}_i^1.$$

Because S has independent columns and the k-vectors $\vec{\mathbf{x}}_i^j$ are independent, then the n-vectors $\vec{\mathbf{y}}_i^j$ are independent.

The **plan** is to isolate the chains for eigenvalue zero, then extend these chains by one vector. Then 1-chains will be constructed from eigenpairs for eigenvalue zero to make n generalized eigenvectors.

Suppose q values of i satisfy $\lambda_i = 0$. We allow $q = 0$. For simplicity, assume such values i are $i = 1, \ldots, q$. The key formula $\vec{\mathbf{y}}_i^j = S\vec{\mathbf{x}}_i^j$ implies $\vec{\mathbf{y}}_i^j$ is in $\mathbf{Image}(B)$, while $B\vec{\mathbf{y}}_i^1 = \lambda_i \vec{\mathbf{y}}_i^1$ implies $\vec{\mathbf{y}}_1^1, \ldots, \vec{\mathbf{y}}_q^1$ are in $\mathbf{kernel}(B)$. Each eigenvector $\vec{\mathbf{y}}_i^1$ starts a Jordan chain ending in $\vec{\mathbf{y}}_i^{m(i)}$. Then[6] the equation $B\vec{\mathbf{u}} = \vec{\mathbf{y}}_i^{m(i)}$ has an n-vector solution $\vec{\mathbf{u}}$. We label $\vec{\mathbf{u}} = \vec{\mathbf{y}}_i^{m(i)+1}$. Because $\lambda_i = 0$, then $B\vec{\mathbf{u}} = \lambda_i \vec{\mathbf{u}} + \vec{\mathbf{y}}_i^{m(i)}$ results in an extended Jordan chain

$$\begin{array}{lcllcl}
B\vec{\mathbf{y}}_i^1 & = & \lambda_i \vec{\mathbf{y}}_i^1 & & \\
B\vec{\mathbf{y}}_i^2 & = & \lambda_i \vec{\mathbf{y}}_i^2 & + & \vec{\mathbf{y}}_i^1 \\
 & \vdots & & & \\
B\vec{\mathbf{y}}_i^{m(i)} & = & \lambda_i \vec{\mathbf{y}}_i^{m(i)} & + & \vec{\mathbf{y}}_i^{m(i)-1} \\
B\vec{\mathbf{y}}_i^{m(i)+1} & = & \lambda_i \vec{\mathbf{y}}_i^{m(i)+1} & + & \vec{\mathbf{y}}_i^{m(i)}
\end{array}$$

Let's extend the independent set $\{\vec{\mathbf{y}}_i^1\}_{i=1}^q$ to a basis of $\mathbf{kernel}(B)$ by adding $s = n - k - q$ additional independent vectors $\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_s$. This basis consists of eigenvectors of B for eigenvalue 0. Then the set of n vectors $\vec{\mathbf{v}}_r$, $\vec{\mathbf{y}}_i^j$ for $1 \le r \le s$, $1 \le i \le p$, $1 \le j \le m(i) + 1$ consists of eigenvectors of B and vectors that satisfy Jordan chain relations. These vectors are columns of a matrix $\mathcal{P}$ that satisfies $B\mathcal{P} = \mathcal{P}\mathcal{J}$ where $\mathcal{J}$ is a Jordan form.

To prove $\mathcal{P}$ invertible, assume a linear combination of the columns of $\mathcal{P}$ is zero:

$$\sum_{i=q+1}^{p} \sum_{j=1}^{m(i)} b_i^j \vec{\mathbf{y}}_i^j + \sum_{i=1}^{q} \sum_{j=1}^{m(i)+1} b_i^j \vec{\mathbf{y}}_i^j + \sum_{i=1}^{s} c_i \vec{\mathbf{v}}_i = \vec{\mathbf{0}}.$$

Apply B to this equation. Because $B\vec{\mathbf{w}} = \vec{\mathbf{0}}$ for any $\vec{\mathbf{w}}$ in $\mathbf{kernel}(B)$, then

$$\sum_{i=q+1}^{p} \sum_{j=1}^{m(i)} b_i^j B\vec{\mathbf{y}}_i^j + \sum_{i=1}^{q} \sum_{j=2}^{m(i)+1} b_i^j B\vec{\mathbf{y}}_i^j = \vec{\mathbf{0}}.$$

The Jordan chain relations imply that the k vectors $B\vec{\mathbf{y}}_i^j$ in the linear combination consist of $\lambda_i \vec{\mathbf{y}}_i^j + \vec{\mathbf{y}}_i^{j-1}$, $\lambda_i \vec{\mathbf{y}}_i^1$, $i = q + 1, \ldots, p$, $j = 2, \ldots, m(i)$, plus the vectors $\vec{\mathbf{y}}_i^j$, $1 \le i \le q$, $1 \le j \le m(i)$. Independence of the original k vectors $\{\vec{\mathbf{y}}_i^j\}$ plus $\lambda_i \ne 0$ for $i > q$ implies this new set is independent. Then all coefficients in the linear combination are zero.

The first linear combination then reduces to $\sum_{i=1}^q b_i^1 \vec{\mathbf{y}}_i^1 + \sum_{i=1}^s c_i \vec{\mathbf{v}}_i = \vec{\mathbf{0}}$. Independence of the constructed basis for $\mathbf{kernel}(B)$ implies $b_i^1 = 0$ for $1 \le i \le q$ and $c_i = 0$ for $1 \le i \le s$. Therefore, the columns of $\mathcal{P}$ are independent. The induction is complete.

[6] The n-vector $\vec{\mathbf{u}}$ is constructed by setting $\vec{\mathbf{u}} = \vec{\mathbf{0}}$, then copy components of k-vector $\vec{\mathbf{x}}_i^{m(i)}$ into pivot locations: $\mathbf{row}(\vec{\mathbf{u}}, i_j) = \mathbf{row}(\vec{\mathbf{x}}_i^{m(i)}, j)$, $j = 1, \ldots, k$.

Geometric and algebraic multiplicity. The **geometric multiplicity** is defined by $\textsf{\textbf{GeoMult}}(\lambda) = \dim(\mathbf{kernel}(A - \lambda I))$, which is the number of basis vectors in a solution to $(A - \lambda I)\vec{\mathbf{x}} = \vec{\mathbf{0}}$, or, equivalently, the number of free variables. The **algebraic multiplicity** is the integer $k = \textsf{\textbf{AlgMult}}(\lambda)$ such that $(r - \lambda)^k$ divides the characteristic polynomial $\det(A - \lambda I)$, but larger powers do not.

Theorem 22 (Algebraic and Geometric Multiplicity)
Let A be a square real or complex matrix. Then

$$1 \le \textsf{\textbf{GeoMult}}(\lambda) \le \textsf{\textbf{AlgMult}}(\lambda). \tag{1}$$

In addition, there are the following relationships between the Jordan form J and algebraic and geometric multiplicities.

GeoMult(λ)	Equals the number of Jordan blocks in J with eigenvalue λ,
AlgMult(λ)	Equals the number of times λ is repeated along the diagonal of J.

Proof: Let $d = \textsf{\textbf{GeoMult}}(\lambda_0)$. Construct a basis $v_1, \ldots, v_n$ of $\mathcal{R}^n$ such that $v_1, \ldots, v_d$ is a basis for $\mathbf{kernel}(A - \lambda_0 I)$. Define $S = \langle v_1 | \ldots | v_n \rangle$ and $B = S^{-1}AS$. The first d columns of AS are $\lambda_0 v_1, \ldots, \lambda_0 v_d$. Then $B = \left(\begin{array}{c|c} \lambda_0 I & C \\ \hline 0 & D \end{array} \right)$ for some matrices C and D. Cofactor expansion implies some polynomial g satisfies

$$\det(A - \lambda I) = \det(S(B - \lambda I)S^{-1}) = \det(B - \lambda I) = (\lambda - \lambda_0)^d g(\lambda)$$

and therefore $d \le \textsf{\textbf{AlgMult}}(\lambda_0)$. Other details of proof are left to the reader.

Chains of generalized eigenvectors. Given an eigenvalue λ of the matrix A, the topic of generalized eigenanalysis determines a Jordan block $B(\lambda, m)$ in J by finding an m-**chain** of generalized eigenvectors $\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_m$, which appear as columns of P in the relation $A = PJP^{-1}$. The very first vector $\vec{\mathbf{v}}_1$ of the chain is an eigenvector, $(A - \lambda I)\vec{\mathbf{v}}_1 = \vec{\mathbf{0}}$. The others $\vec{\mathbf{v}}_2, \ldots, \vec{\mathbf{v}}_k$ are not eigenvectors but satisfy

$$(A - \lambda I)\vec{\mathbf{v}}_2 = \vec{\mathbf{v}}_1, \quad \ldots \quad , \quad (A - \lambda I)\vec{\mathbf{v}}_m = \vec{\mathbf{v}}_{m-1}.$$

Implied by the term m-**chain** is insolvability of $(A - \lambda I)\vec{\mathbf{x}} = \vec{\mathbf{v}}_m$. The chain size m is subject to the inequality $1 \le m \le \textsf{\textbf{AlgMult}}(\lambda)$.

The Jordan form J may contain several Jordan blocks for one eigenvalue λ. To illustrate, if J has only one eigenvalue λ and $\textsf{\textbf{AlgMult}}(\lambda) = 3$, then J might be constructed as follows:

$$\begin{aligned}
J = \textbf{diag}(B(\lambda,1), B(\lambda,1), B(\lambda,1)) &= \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}, \\
J = \textbf{diag}(B(\lambda,1), B(\lambda,2)) &= \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \\
J = B(\lambda,3) &= \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}.
\end{aligned}$$

The three generalized eigenvectors for this example correspond to

$$\begin{aligned}
J = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} \quad &\leftrightarrow \quad \text{Three 1-chains,} \\
J = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \quad &\leftrightarrow \quad \text{One 1-chain and one 2-chain,} \\
J = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \quad &\leftrightarrow \quad \text{One 3-chain.}
\end{aligned}$$

Computing m-chains. Let us fix the discussion to an eigenvalue λ of A. Define $N = A - \lambda I$ and $p = \textsf{AlgMult}(\lambda)$.

To compute an m-chain, start with an eigenvector $\vec{\mathbf{v}}_1$ and solve recursively by `rref` methods $N\vec{\mathbf{v}}_{j+1} = \vec{\mathbf{v}}_j$ until there fails to be a solution. This must seemingly be done for *all possible choices* of $\vec{\mathbf{v}}_1$! The search for m-chains terminates when p independent generalized eigenvectors have been calculated.

If A has an essentially unique eigenpair $(\lambda, \vec{\mathbf{v}}_1)$, then this process terminates immediately with an m-chain where $m = p$. The chain produces one Jordan block $B(\lambda, m)$ and the generalized eigenvectors $\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_m$ are recorded into the matrix P.

If $\vec{\mathbf{u}}_1$, $\vec{\mathbf{u}}_2$ form a basis for the eigenvectors of A corresponding to λ, then the problem $N\vec{\mathbf{x}} = \vec{\mathbf{0}}$ has 2 free variables. Therefore, we seek to find an m_1-chain and an m_2-chain such that $m_1 + m_2 = p$, corresponding to two Jordan blocks $B(\lambda, m_1)$ and $B(\lambda, m_2)$.

To understand the logic applied here, the reader should verify that for $\mathcal{N} = \textbf{diag}(B(0,m_1), B(0,m_2), \ldots, B(0,m_k))$ the problem $\mathcal{N}\vec{\mathbf{x}} = \vec{\mathbf{0}}$ has k free variables, because $\mathcal{N}$ is already in `rref` form. These remarks imply that a k-dimensional basis of eigenvectors of A for eigenvalue λ

causes a search for m_i-chains, $1 \le i \le k$, such that $m_1 + \cdots + m_k = p$, corresponding to k Jordan blocks $B(\lambda, m_1)$, …, $B(\lambda, m_k)$.

A common naive approach for computing generalized eigenvectors can be illustrated by letting

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \vec{\mathbf{u}}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \quad \vec{\mathbf{u}}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

Matrix A has one eigenvalue $\lambda = 1$ and two eigenpairs $(1, \vec{\mathbf{u}}_1)$, $(1, \vec{\mathbf{u}}_2)$. Starting a chain calculation with $\vec{\mathbf{v}}_1$ equal to either $\vec{\mathbf{u}}_1$ or $\vec{\mathbf{u}}_2$ gives a 1-chain. This naive approach leads to only two independent generalized eigenvectors. However, the calculation must proceed until three independent generalized eigenvectors have been computed. To resolve the trouble, keep a 1-chain, say the one generated by $\vec{\mathbf{u}}_1$, and start a new chain calculation using $\vec{\mathbf{v}}_1 = a_1\vec{\mathbf{u}}_1 + a_2\vec{\mathbf{u}}_2$. Adjust the values of a_1, a_2 until a 2-chain has been computed:

$$\langle A - \lambda I | \vec{\mathbf{v}}_1 \rangle = \begin{pmatrix} 0 & 1 & 1 & a_1 \\ 0 & 0 & 0 & -a_1 + a_2 \\ 0 & 0 & 0 & a_1 - a_2 \end{pmatrix} \approx \begin{pmatrix} 0 & 1 & 1 & a_1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

provided $a_1 - a_2 = 0$. Choose $a_1 = a_2 = 1$ to make $\vec{\mathbf{v}}_1 = \vec{\mathbf{u}}_1 + \vec{\mathbf{u}}_2 \ne \vec{\mathbf{0}}$ and solve for $\vec{\mathbf{v}}_2 = \begin{pmatrix} 0, 1, 0 \end{pmatrix}$. Then $\vec{\mathbf{u}}_1$ is a 1-chain and $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$ is a 2-chain. The generalized eigenvectors $\vec{\mathbf{u}}_1$, $\vec{\mathbf{v}}_1$, $\vec{\mathbf{v}}_2$ are independent and form the columns of P while $J = \mathbf{diag}(B(\lambda, 1), B(\lambda, 2))$ (recall $\lambda = 1$). We justify $A = PJP^{-1}$ by testing $AP = PJ$, using the formulas

$$J = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Jordan Decomposition using `maple`

Displayed here is `maple` code which applied to the matrix

$$A = \begin{pmatrix} 4 & -2 & 5 \\ -2 & 4 & -3 \\ 0 & 0 & 2 \end{pmatrix}$$

produces the Jordan decomposition

$$A = PJP^{-1} = \frac{1}{4}\begin{pmatrix} 1 & 4 & -7 \\ -1 & 4 & 1 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 6 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix} \frac{1}{4}\begin{pmatrix} 8 & -8 & 16 \\ 2 & 2 & 3 \\ 0 & 0 & 4 \end{pmatrix}.$$

```
A := Matrix([[4, -2, 5], [-2, 4, -3], [0, 0, 2]]);
factor(LinearAlgebra[CharacteristicPolynomial](A,lambda));
# Answer == (lambda-6)*(lambda-2)^2
J,P:=LinearAlgebra[JordanForm](A,output=[’J’,’Q’]);
zero:=A.P-P.J; # zero matrix expected
```

Number of Jordan Blocks

In calculating generalized eigenvectors of A for eigenvalue λ, it is possible to decide in advance how many Jordan chains of size k should be computed. A practical consequence is to organize the computation for certain chain sizes.

Theorem 23 (Number of Jordan Blocks)
Given eigenvalue λ of A, define $N = A - \lambda I$, $k(j) = \dim(\mathbf{kernel}(N^j))$. Let p be the least integer such that $N^p = N^{p+1}$. Then the Jordan form of A has $2k(j-1) - k(j-2) - k(j)$ Jordan blocks $B(\lambda, j-1)$, $j = 3, \ldots, p$.

The proof of the theorem is in the exercises, where more detail appears for $p = 1$ and $p = 2$. Complete results are in the `maple` code below.

An Illustration. This example is a 5×5 matrix A with one eigenvalue $\lambda = 2$ of multiplicity 5. Let $s(j)$ = number of $j \times j$ Jordan blocks.

$$A = \begin{pmatrix} 3 & -1 & 1 & 0 & 0 \\ 2 & 0 & 1 & 1 & 0 \\ 1 & -1 & 2 & 1 & 0 \\ -1 & 1 & 0 & 2 & 1 \\ -3 & 3 & 0 & -2 & 3 \end{pmatrix}, \quad N = A - 2I = \begin{pmatrix} 1 & -1 & 1 & 0 & 0 \\ 2 & -2 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 & 0 \\ -1 & 1 & 0 & 0 & 1 \\ -3 & 3 & 0 & -2 & 1 \end{pmatrix}.$$

Then $N^3 = N^4 = N^5 = 0$ implies $k(3) = k(4) = k(5) = 5$. Further, $k(2) = 4$, $k(1) = 2$. Then $s(5) = s(4) = 0$, $s(3) = s(2) = 1$, $s(1) = 0$, which implies one block of each size 2 and 3.

Some `maple` code automates the investigation:

```
with(LinearAlgebra):
A := Matrix([
[ 3, -1, 1,  0, 0],[ 2,  0, 1,  1, 0],
[ 1, -1, 2,  1, 0],[-1,  1, 0,  2, 1],
[-3,  3, 0, -2, 3] ]);
lambda:=2;
n:=RowDimension(A);N:=A-lambda*IdentityMatrix(n);
for j from 1 to n do
 k[j]:=n-Rank(N^j); od:
for p from n to 2 by -1 do
```

```
 if(k[p]<>k[p-1])then break; fi: od;
txt:=(j,x)->printf(‘if‘(x=1,
 cat("B(lambda,",j,") occurs 1 time\n"),
 cat("B(lambda,",j,") occurs ",x," times\n"))):
printf("lambda=%d, nilpotency=%d\n",lambda,p);
if(p=1) then txt(1,k[1]); else
 txt(p,k[p]-k[p-1]);
 for j from p to 3 by -1 do
   txt(j-1,2*k[j-1]-k[j-2]-k[j]): od:
 txt(1,2*k[1]-k[2]);
fi:
#lambda=2, nilpotency=3
#B(lambda,3) occurs 1 time
#B(lambda,2) occurs 1 time
#B(lambda,1) occurs 0 times
J,P:=JordanForm(A,output=[’J’,’Q’])}:
# Answer check for the maple code
```

$$J = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}, \quad P = \frac{1}{2}\begin{pmatrix} 0 & 1 & 2 & -1 & 0 \\ -4 & 2 & 2 & -2 & 2 \\ -4 & 1 & 1 & -1 & 1 \\ -4 & -3 & 1 & -1 & 1 \\ 4 & -5 & -3 & 1 & -3 \end{pmatrix}$$

Numerical Instability

The matrix $A = \begin{pmatrix} 1 & 1 \\ \varepsilon & 1 \end{pmatrix}$ has two possible Jordan forms

$$J(\varepsilon) = \begin{cases} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} & \varepsilon = 0, \\ \\ \begin{pmatrix} 1+\sqrt{\varepsilon} & 0 \\ 0 & 1-\sqrt{\varepsilon} \end{pmatrix} & \varepsilon > 0. \end{cases}$$

When $\varepsilon \approx 0$, then numerical algorithms become unstable, unable to lock onto the correct Jordan form. Briefly, $\lim_{\varepsilon \to 0} J(\varepsilon) \neq J(0)$.

The Real Jordan Form of A

Given a real matrix A, generalized eigenanalysis seeks to find a *real* invertible matrix $\mathcal{P}$ and a *real* upper triangular block matrix R such that $A = \mathcal{P}R\mathcal{P}^{-1}$.

If λ is a real eigenvalue of A, then a **real Jordan block** is a matrix

$$B = \textsf{\textbf{diag}}(\lambda, \ldots, \lambda) + N, \quad N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

If $\lambda = a + ib$ is a complex eigenvalue of A, then symbols λ, 1 and 0 are replaced respectively by 2×2 real matrices $\Lambda = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, $\mathcal{I} = \textsf{\textbf{diag}}(1, 1)$ and $\mathcal{O} = \textsf{\textbf{diag}}(0, 0)$. The corresponding $2m\times 2m$ real Jordan block matrix is given by the formula

$$B = \textsf{\textbf{diag}}(\Lambda, \ldots, \Lambda) + \mathcal{N}, \quad \mathcal{N} = \begin{pmatrix} \mathcal{O} & \mathcal{I} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{O} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{I} \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{O} \end{pmatrix}.$$

Direct Sum Decomposition

The **generalized eigenspace** of eigenvalue λ of an $n\times n$ matrix A is the subspace $\mathbf{kernel}((A - \lambda I)^p)$ where $p = \textsf{\textbf{AlgMult}}(\lambda)$. We state without proof the main result for generalized eigenspace bases, because details appear in the exercises. A proof is included for the direct sum decomposition, even though Putzer's spectral theory independently produces the same decomposition.

Theorem 24 (Generalized Eigenspace Basis)
The subspace $\mathbf{kernel}((A - \lambda I)^k)$, $k = \textsf{\textbf{AlgMult}}(\lambda)$ has a k-dimensional basis whose vectors are the columns of P corresponding to blocks $B(\lambda, j)$ of J, in Jordan decomposition $A = PJP^{-1}$.

Theorem 25 (Direct Sum Decomposition)
Given $n\times n$ matrix A and distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, $n_1 = \textsf{\textbf{AlgMult}}(\lambda_i)$, $\ldots$, $n_k = \textsf{\textbf{AlgMult}}(\lambda_i)$, then A induces a direct sum decomposition

$$\mathcal{C}^n = \mathbf{kernel}((A - \lambda_1 I)^{n_1} \oplus \cdots \oplus \mathbf{kernel}((A - \lambda_k I)^{n_k}.$$

This equation means that each complex vector $\vec{\mathbf{x}}$ in $\mathcal{C}^n$ can be uniquely written as

$$\vec{\mathbf{x}} = \vec{\mathbf{x}}_1 + \cdots + \vec{\mathbf{x}}_k$$

where each $\vec{\mathbf{x}}_i$ belongs to $\mathbf{kernel}\,((A - \lambda_i)^{n_i})$, $i = 1, \ldots, k$.

Proof: The previous theorem implies there is a basis of dimension n_i for $E_i \equiv \mathbf{kernel}((A - \lambda_i I)^{n_i})$, $i = 1, \ldots, k$. Because $n_1 + \cdots + n_k = n$, then there are n vectors in the union of these bases. The independence test for these n vectors amounts to showing that $\vec{\mathbf{x}}_1 + \cdots + \vec{\mathbf{x}}_k = \vec{\mathbf{0}}$ with $\vec{\mathbf{x}}_i$ in E_i, $i = 1, \ldots, k$, implies all $\vec{\mathbf{x}}_i = \vec{\mathbf{0}}$. This will be true provided $E_i \cap E_j = \{\vec{\mathbf{0}}\}$ for $i \neq j$.

Let's assume a Jordan decomposition $A = PJP^{-1}$. If $\vec{\mathbf{x}}$ is common to both E_i and E_j, then basis expansion of $\vec{\mathbf{x}}$ in both subspaces implies a linear combination of the columns of P is zero, which by independence of the columns of P implies $\vec{\mathbf{x}} = \vec{\mathbf{0}}$.

The proof is complete.

Computing Exponential Matrices

Discussed here are methods for finding a real exponential matrix e^{At} when A is real. Two formulas are given, one for a real eigenvalue and one for a complex eigenvalue. These formulas supplement the spectral formulas given earlier in the text.

Nilpotent matrices. A matrix N which satisfies $N^p = 0$ for some integer p is called **nilpotent**. The least integer p for which $N^p = 0$ is called the **nilpotency** of N. A nilpotent matrix N has a finite exponential series:

$$e^{Nt} = I + Nt + N^2\frac{t^2}{2!} + \cdots + N^{p-1}\frac{t^{p-1}}{(p-1)!}.$$

If $N = B(\lambda, p) - \lambda I$, then the finite sum has a splendidly simple expression. Due to $e^{\lambda t + Nt} = e^{\lambda t}e^{Nt}$, this proves the following result.

Theorem 26 (Exponential of a Jordan Block Matrix)
If λ is real and

$$B = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix} \quad (p \times p \text{ matrix})$$

then

$$e^{Bt} = e^{\lambda t}\begin{pmatrix} 1 & t & \frac{t^2}{2} & \cdots & \frac{t^{p-2}}{(p-2)!} & \frac{t^{p-1}}{(p-1)!} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & t \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

The equality also holds if λ is a complex number, in which case both sides of the equation are complex.

Real Exponentials for Complex λ. A Jordan decomposition $A = \mathcal{P}J\mathcal{P}^{-1}$, in which A has only real eigenvalues, has real generalized eigenvectors appearing as columns in the matrix $\mathcal{P}$, in the natural order given in J. When $\lambda = a + ib$ is complex, $b > 0$, then the real and imaginary parts of each generalized eigenvector are entered pairwise into $\mathcal{P}$; the conjugate eigenvalue $\overline{\lambda} = a - ib$ is skipped. The complex entry along the diagonal of J is changed into a 2×2 matrix under the correspondence

$$a + ib \leftrightarrow \begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

The result is a *real* matrix $\mathcal{P}$ and a *real* block upper triangular matrix J which satisfy $A = \mathcal{P}J\mathcal{P}^{-1}$.

Theorem 27 (Real Block Diagonal Matrix, Eigenvalue $a + ib$)

Let $\Lambda = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, $\mathcal{I} = \textbf{diag}(1,1)$ and $\mathcal{O} = \textbf{diag}(0,0)$. Consider a real Jordan block matrix of dimension $2m \times 2m$ given by the formula

$$B = \begin{pmatrix} \Lambda & \mathcal{I} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{O} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \Lambda & \mathcal{I} \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{O} & \Lambda \end{pmatrix}.$$

If $\mathcal{R} = \begin{pmatrix} \cos bt & \sin bt \\ -\sin bt & \cos bt \end{pmatrix}$, then

$$e^{Bt} = e^{at} \begin{pmatrix} \mathcal{R} & t\mathcal{R} & \frac{t^2}{2}\mathcal{R} & \cdots & \frac{t^{m-2}}{(m-2)!}\mathcal{R} & \frac{t^{m-1}}{(m-1)!}\mathcal{R} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{R} & t\mathcal{R} \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{R} \end{pmatrix}.$$

Solving $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$. The solution $\vec{\mathbf{x}}(t) = e^{At}\vec{\mathbf{x}}(0)$ must be real if A is real. The real solution can be expressed as $\vec{\mathbf{x}}(t) = \mathcal{P}\vec{\mathbf{y}}(t)$ where $\vec{\mathbf{y}}'(t) = R\vec{\mathbf{y}}(t)$ and R is a real Jordan form of A, containing real Jordan blocks $B_1, \ldots, B_k$ down its diagonal. Theorems above provide explicit formulas for the block matrices $e^{B_i t}$ in the relation

$$e^{Rt} = \textbf{diag}\left(e^{B_1 t}, \ldots, e^{B_k t}\right).$$

The resulting formula

$$\vec{\mathbf{x}}(t) = \mathcal{P}e^{Rt}\mathcal{P}^{-1}\vec{\mathbf{x}}(0)$$

contains only real numbers, real exponentials, plus sine and cosine terms, which are possibly multiplied by polynomials in t.

Exercises 11.6

Jordan block. Write out explicitly.

1.

2.

3.

4.

Jordan form. Which are Jordan forms and which are not? Explain.

5.

6.

7.

8.

Decoding $A = PJP^{-1}$. Decode $A = PJP^{-1}$ in each case, displaying explicitly the Jordan chain relations.

9.

10.

11.

12.

Geometric multiplicity. Determine the geometric multiplicity **GeoMult**(λ).

13.

14.

15.

16.

Algebraic multiplicity. Determine the algebraic multiplicity **AlgMult**(λ).

17.

18.

19.

20.

Generalized eigenvectors. Find all generalized eigenvectors and represent $A = PJP^{-1}$.

21.

22.

23.

24.

25.

26.

27.

28.

29.

30.

31.

32.

Computing m-chains. Find the Jordan chains for the given eigenvalue.

33.

34.

35.

36.

37.

38.

39.

40.

Jordan Decomposition. Use `maple` to find the Jordan decomposition.

41.

42.

43.

44.

45.

46.

47.

48.

Number of Jordan Blocks. Outlined here is the derivation of

$$s(j) = 2k(j-1) - k(j-2) - k(j).$$

Definitions:

- $s(j)$= number of blocks $B(\lambda, j)$
- $N = A - \lambda I$
- $k(j) = \dim(\mathbf{kernel}(N^j))$
- $L_j = \mathbf{kernel}(N^{j-1})^{\perp}$ relative to $\mathbf{kernel}(N^j)$
- $\ell(j) = \dim(L_j)$
- p minimizes $\mathbf{kernel}(N^p) = \mathbf{kernel}(N^{p+1})$

49. Verify $k(j) \le k(j+1)$ from

$$\mathbf{kernel}(N^j) \subset \mathbf{kernel}(N^{j+1}).$$

50. Verify the direct sum formula

$$\mathbf{kernel}(N^j) = \mathbf{kernel}(N^{j-1}) \oplus L_j.$$

Then $k(j) = k(j-1) + \ell(j)$.

51. Given $N^j\vec{\mathbf{v}} = \vec{\mathbf{0}}$, $N^{j-1}\vec{\mathbf{v}} \ne \vec{\mathbf{0}}$, define $\vec{\mathbf{v}}_i = N^{j-i}\vec{\mathbf{v}}$, $i = 1, \ldots, j$. Show that these are independent vectors satisfying Jordan chain relations $N\vec{\mathbf{v}}_1 = \vec{\mathbf{0}}$, $N\vec{\mathbf{v}}_{i+i} = \vec{\mathbf{v}}_i$.

52. A block $B(\lambda, p)$ corresponds to a Jordan chain $\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_p$ constructed from the Jordan decomposition. Use $N^{j-1}\vec{\mathbf{v}}_j = \vec{\mathbf{v}}_1$ and $\mathbf{kernel}(N^p) = \mathbf{kernel}(N^{p+1})$ to show that the number of such blocks $B(\lambda, p)$ is $\ell(p)$. Then for $p > 1$, $s(p) = k(p) - k(p-1)$.

53. Show that $\ell(j-1) - \ell(j)$ is the number of blocks $B(\lambda, j)$ for $2 < j < p$. Then

$$s(j) = 2k(j-1) - k(j) - k(j-2).$$

54. Test the formulas above on the special matrices

$$A = \mathsf{\mathbf{diag}}(B(\lambda,1), B(\lambda,1), B(\lambda,1)),$$
$$A = \mathsf{\mathbf{diag}}(B(\lambda,1), B(\lambda,2), B(\lambda,3)),$$
$$A = \mathsf{\mathbf{diag}}(B(\lambda,1), B(\lambda,3), B(\lambda,3)),$$
$$A = \mathsf{\mathbf{diag}}(B(\lambda,1), B(\lambda,1), B(\lambda,3)),$$

Generalized Eigenspace Basis.

Let A be $n \times n$ with distinct eigenvalues λ_i, $n_i = \mathsf{\mathbf{AlgMult}}(\lambda_i)$ and $E_i = \mathbf{kernel}((A - \lambda_i I)^{n_i})$, $i = 1, \ldots, k$. Assume a Jordan decomposition $A = PJP^{-1}$.

55. Let Jordan block $B(\lambda, j)$ appear in J. Prove that a Jordan chain corresponding to this block is a set of j independent columns of P.

56. Let $\mathcal{B}_\lambda$ be the union of all columns of P originating from Jordan chains associated with Jordan blocks $B(\lambda, j)$. Prove that $\mathcal{B}_\lambda$ is an independent set.

57. Verify that $\mathcal{B}_\lambda$ has $\mathsf{\mathbf{AlgMult}}(\lambda)$ basis elements.

58. Prove that $E_i = \mathbf{span}(\mathcal{B}_{\lambda_i})$ and $\dim(E_i) = n_i$, $i = 1, \ldots, k$.

Numerical Instability. Show directly that $\lim_{\epsilon \to 0} J(\epsilon) \ne J(0)$.

59.

60.

61.

62.

Direct Sum Decomposition. Display the direct sum decomposition.

63.

64.

65.

66.

67.

68.

69.

70.

Exponential Matrices. Compute the exponential matrix on paper and then check the answer using `maple`.

71.

72.

73.

74.

75.

76.

77.

78.

Nilpotent matrices. Find the nilpotency of N.

79.

80.

81.

82.

Real Exponentials. Compute the real exponential e^{At} on paper. Check the answer in `maple`.

83.

84.

85.

86.

Real Jordan Form. Find the real Jordan form.

87.

88.

89.

90.

Solving $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$. Solve the differential equation.

91.

92.

93.

94.

11.7 Nonhomogeneous Linear Systems

Variation of Parameters

The **method of variation of parameters** is a general method for solving a linear nonhomogeneous system

$$\vec{\mathbf{x}}' = A\vec{\mathbf{x}} + \vec{\mathbf{F}}(t).$$

Historically, it was a trial solution method, whereby the nonhomogeneous system is solved using a trial solution of the form

$$\vec{\mathbf{x}}(t) = e^{At}\vec{\mathbf{x}}_0(t).$$

In this formula, $\vec{\mathbf{x}}_0(t)$ is a vector function to be determined. The method is imagined to originate by varying $\vec{\mathbf{x}}_0$ in the general solution $\vec{\mathbf{x}}(t) = e^{At}\vec{\mathbf{x}}_0$ of the linear homogenous system $\vec{\mathbf{x}}' = A\vec{\mathbf{x}}$. Hence was coined the names *variation of parameters* and *variation of constants.*

Modern use of variation of parameters is through a formula, memorized for routine use.

Theorem 28 (Variation of Parameters for Systems)
Let A be a constant $n \times n$ matrix and $\vec{\mathbf{F}}(t)$ a continuous function near $t = t_0$. The unique solution $\vec{\mathbf{x}}(t)$ of the matrix initial value problem

$$\vec{\mathbf{x}}'(t) = A\vec{\mathbf{x}}(t) + \vec{\mathbf{F}}(t), \quad \vec{\mathbf{x}}(t_0) = \vec{\mathbf{x}}_0,$$

is given by the **variation of parameters formula**

$$\vec{\mathbf{x}}(t) = e^{At}\vec{\mathbf{x}}_0 + e^{At}\int_{t_0}^{t} e^{-rA}\vec{\mathbf{F}}(r)dr. \tag{1}$$

Proof of (1). Define

$$\vec{\mathbf{u}}(t) = \vec{\mathbf{x}}_0 + \int_{t_0}^{t} e^{-rA}\vec{\mathbf{F}}(r)dr.$$

To show (1) holds, we must verify $\vec{\mathbf{x}}(t) = e^{At}\vec{\mathbf{u}}(t)$. First, the function $\vec{\mathbf{u}}(t)$ is differentiable with continuous derivative $e^{-tA}\vec{\mathbf{F}}(t)$, by the fundamental theorem of calculus applied to each of its components. The product rule of calculus applies to give

$$\begin{aligned}\vec{\mathbf{x}}'(t) &= \left(e^{At}\right)'\vec{\mathbf{u}}(t) + e^{At}\vec{\mathbf{u}}'(t)\\ &= Ae^{At}\vec{\mathbf{u}}(t) + e^{At}e^{-At}\vec{\mathbf{F}}(t)\\ &= A\vec{\mathbf{x}}(t) + \vec{\mathbf{F}}(t).\end{aligned}$$

Therefore, $\vec{\mathbf{x}}(t)$ satisfies the differential equation $\vec{\mathbf{x}}' = A\vec{\mathbf{x}} + \vec{\mathbf{F}}(t)$. Because $\vec{\mathbf{u}}(t_0) = \vec{\mathbf{x}}_0$, then $\vec{\mathbf{x}}(t_0) = \vec{\mathbf{x}}_0$, which shows the initial condition is also satisfied. The proof is complete.

Undetermined Coefficients

The trial solution method known as the method of undetermined coefficients can be applied to vector-matrix systems $\vec{\mathbf{x}}' = A\vec{\mathbf{x}} + \vec{\mathbf{F}}(t)$ when the components of $\vec{\mathbf{F}}$ are sums of terms of the form

$$(\text{polynomial in } t)e^{at}(\cos(bt) \text{ or } \sin(bt)).$$

Such terms are known as **Euler solution atoms**. It is usually efficient to write $\vec{\mathbf{F}}$ in terms of the columns $\vec{\mathbf{e}}_1, \ldots, \vec{\mathbf{e}}_n$ of the $n \times n$ identity matrix I, as the combination

$$\vec{\mathbf{F}}(t) = \sum_{j=1}^{n} F_j(t)\vec{\mathbf{e}}_j.$$

Then

$$\vec{\mathbf{x}}(t) = \sum_{j=1}^{n} \vec{\mathbf{x}}_j(t),$$

where $\vec{\mathbf{x}}_j(t)$ is a particular solution of the simpler equation

$$\vec{\mathbf{x}}'(t) = A\vec{\mathbf{x}}(t) + f(t)\vec{\mathbf{c}}, \quad f = F_j, \quad \vec{\mathbf{c}} = \vec{\mathbf{e}}_j.$$

An initial trial solution $\vec{\mathbf{x}}(t)$ for $\vec{\mathbf{x}}'(t) = A\vec{\mathbf{x}}(t) + f(t)\vec{\mathbf{c}}$ can be determined from the following **initial trial solution rule**:

> Let $f(t)$ be a sum of Euler solution atoms. Identify independent functions whose linear combinations give all derivatives of $f(t)$. The initial trial solution is a linear combination of these functions with undetermined vector coefficients $\{\vec{\mathbf{c}}_j\}$.

In the well-known scalar case, the trial solution must be modified if its terms contain any portion of the general solution to the homogeneous equation. In the vector case, if $f(t)$ is a polynomial, then the *correction rule* for the initial trial solution is avoided by assuming the matrix A is invertible. This assumption means that $r = 0$ is not a root of $\det(A - rI) = 0$, which prevents the homogenous solution from having any polynomial terms.

The initial vector trial solution is substituted into the differential equation to find the undetermined coefficients $\{\vec{\mathbf{c}}_j\}$, hence finding a particular solution.

Theorem 29 (Polynomial solutions)
Let $f(t) = \sum_{j=0}^{k} p_j \frac{t^j}{j!}$ be a polynomial of degree k. Assume A is an $n \times n$ constant invertible matrix. Then $\vec{\mathbf{u}}' = A\vec{\mathbf{u}} + f(t)\vec{\mathbf{c}}$ has a polynomial solution $\vec{\mathbf{u}}(t) = \sum_{j=0}^{k} \vec{\mathbf{c}}_j \frac{t^j}{j!}$ of degree k with vector coefficients $\{\vec{\mathbf{c}}_j\}$ given by the relations

$$\vec{\mathbf{c}}_j = -\sum_{i=j}^{k} p_i A^{j-i-1}\vec{\mathbf{c}}, \quad 0 \le j \le k.$$

Theorem 30 (Polynomial × exponential solutions)
Let $g(t) = \sum_{j=0}^{k} p_j \frac{t^j}{j!}$ be a polynomial of degree k. Assume A is an $n \times n$ constant matrix and $B = A - aI$ is invertible. Then $\vec{\mathbf{u}}\,' = A\vec{\mathbf{u}} + e^{at}g(t)\vec{\mathbf{c}}$ has a polynomial-exponential solution $\vec{\mathbf{u}}(t) = e^{at}\sum_{j=0}^{k} \vec{\mathbf{c}}_j \frac{t^j}{j!}$ with vector coefficients $\{\vec{\mathbf{c}}_j\}$ given by the relations

$$\vec{\mathbf{c}}_j = -\sum_{i=j}^{k} p_i B^{j-i-1}\vec{\mathbf{c}}, \quad 0 \le j \le k.$$

Proof of Theorem 29. Substitute $\vec{\mathbf{u}}(t) = \sum_{j=0}^{k} \vec{\mathbf{c}}_j \frac{t^j}{j!}$ into the differential equation, then

$$\sum_{j=0}^{k-1} \vec{\mathbf{c}}_{j+1}\frac{t^j}{j!} = A\sum_{j=0}^{k} \vec{\mathbf{c}}_j \frac{t^j}{j!} + \sum_{j=0}^{k} p_j \frac{t^j}{j!}\vec{\mathbf{c}}.$$

Then terms on the right for $j = k$ must add to zero and the others match the left side coefficients of $t^j/j!$, giving the relations

$$A\vec{\mathbf{c}}_k + p_k\vec{\mathbf{c}} = \vec{\mathbf{0}}, \quad \vec{\mathbf{c}}_{j+1} = A\vec{\mathbf{c}}_j + p_j\vec{\mathbf{c}}.$$

Solving these relations recursively gives the formulas

$$\begin{array}{lcl} \vec{\mathbf{c}}_k & = & -p_k A^{-1}\vec{\mathbf{c}}, \\ \vec{\mathbf{c}}_{k-1} & = & -\left(p_{k-1}A^{-1} + p_k A^{-2}\right)\vec{\mathbf{c}}, \\ & \vdots & \\ \vec{\mathbf{c}}_0 & = & -\left(p_0 A^{-1} + \cdots + p_k A^{-k-1}\right)\vec{\mathbf{c}}. \end{array}$$

The relations above can be summarized by the formula

$$\vec{\mathbf{c}}_j = -\sum_{i=j}^{k} p_i A^{j-i-1}\vec{\mathbf{c}}, \quad 0 \le j \le k.$$

The calculation shows that if $\vec{\mathbf{u}}(t) = \sum_{j=0}^{k} \vec{\mathbf{c}}_j \frac{t^j}{j!}$ and $\vec{\mathbf{c}}_j$ is given by the last formula, then $\vec{\mathbf{u}}(t)$ substituted into the differential equation gives matching LHS and RHS. The proof is complete.

Proof of Theorem 30. Let $\vec{\mathbf{u}}(t) = e^{at}\vec{\mathbf{v}}(t)$. Then $\vec{\mathbf{u}}\,' = A\vec{\mathbf{u}} + e^{at}g(t)\vec{\mathbf{c}}$ implies $\vec{\mathbf{v}}\,' = (A - aI)\vec{\mathbf{v}} + g(t)\vec{\mathbf{c}}$. Apply Theorem 29 to $\vec{\mathbf{v}}\,' = B\vec{\mathbf{v}} + g(t)\vec{\mathbf{c}}$. The proof is complete.

11.8 Second-order Systems

A model problem for second order systems is the system of three masses coupled by springs studied in section 11.1, equation (6):

$$\begin{aligned} m_1 x_1''(t) &= -k_1 x_1(t) + k_2[x_2(t) - x_1(t)], \\ m_2 x_2''(t) &= -k_2[x_2(t) - x_1(t)] + k_3[x_3(t) - x_2(t)], \\ m_3 x_3''(t) &= -k_3[x_3(t) - x_2(t)] - k_4 x_3(t). \end{aligned} \tag{1}$$

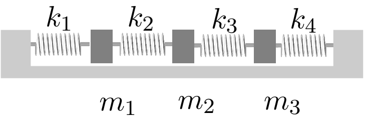

Figure 22. Three masses connected by springs. The masses slide on a frictionless surface.

In vector-matrix form, this system is a **second order system**

$$M\vec{\mathbf{x}}''(t) = K\vec{\mathbf{x}}(t)$$

where the **displacement** $\vec{\mathbf{x}}$, **mass matrix** M and **stiffness matrix** K are defined by the formulas

$$\vec{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad M = \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix}, \quad K = \begin{pmatrix} -k_1 - k_2 & k_2 & 0 \\ k_2 & -k_2 - k_3 & k_3 \\ 0 & k_3 & -k_3 - k_4 \end{pmatrix}.$$

Because M is invertible, the system can always be written as

$$\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}, \quad A = M^{-1}K.$$

Converting $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$ to $\vec{\mathbf{u}}' = C\vec{\mathbf{u}}$

Given a second order $n \times n$ system $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$, define the variable $\vec{\mathbf{u}}$ and the $2n \times 2n$ block matrix C as follows.

$$\vec{\mathbf{u}} = \begin{pmatrix} \vec{\mathbf{x}} \\ \vec{\mathbf{x}}' \end{pmatrix}, \quad C = \left(\begin{array}{c|c} 0 & I \\ \hline A & 0 \end{array}\right). \tag{2}$$

Then each solution $\vec{\mathbf{x}}$ of the second order system $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$ produces a corresponding solution $\vec{\mathbf{u}}$ of the first order system $\vec{\mathbf{u}}' = C\vec{\mathbf{u}}$. Similarly, each solution $\vec{\mathbf{u}}$ of $\vec{\mathbf{u}}' = C\vec{\mathbf{u}}$ gives a solution $\vec{\mathbf{x}}$ of $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$ by the formula $\vec{\mathbf{x}} = \textbf{diag}(I, 0)\vec{\mathbf{u}}$.

Euler's Substitution $\vec{\mathbf{x}} = e^{\lambda t}\vec{\mathbf{v}}$

The fundamental substitution of L. Euler applies to vector-matrix differential systems. In particular, for $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$, the equation $\vec{\mathbf{x}} = e^{\lambda t}\vec{\mathbf{v}}$ produces the **characteristic equation**

$$\det(A - \lambda^2 I) = 0,$$

and the **eigenpair equation**

$$A\vec{\mathbf{v}} = \lambda^2\vec{\mathbf{v}}, \quad \vec{\mathbf{v}} \neq \vec{\mathbf{0}},$$

which means that $(\lambda^2, \vec{\mathbf{v}})$ is an eigenpair of the matrix A.

Negative eigenvalues of A produce complex conjugate values for λ. For instance, $\lambda^2 = -4$ implies $\lambda = \pm 2i$, and then, even though vector $\vec{\mathbf{v}}$ has real components, the solution $\vec{\mathbf{x}}(t) = e^{\lambda t}\vec{\mathbf{v}}$ is a vector with complex entries: $\vec{\mathbf{x}}(t) = e^{2it}\vec{\mathbf{v}} = \cos(2t)\vec{\mathbf{v}} + i\sin(2t)\vec{\mathbf{v}}$.

Details. Compute $\vec{\mathbf{x}}' = \frac{d}{dt}e^{\lambda t}\vec{\mathbf{v}} = \lambda e^{\lambda t}\vec{\mathbf{v}} = \lambda\vec{\mathbf{x}}$. Then $\vec{\mathbf{x}}'' = \lambda^2\vec{\mathbf{x}}$. If $\vec{\mathbf{x}} = e^{\lambda t}\vec{\mathbf{v}}$ is a nonzero solution of $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$, then $\lambda^2\vec{\mathbf{x}} = A\vec{\mathbf{x}}$ holds, which is equivalent to $\lambda^2\vec{\mathbf{v}} = A\vec{\mathbf{v}}$. Then $(\lambda^2, \vec{\mathbf{v}})$ is an eigenpair of A. Conversely, if $(\lambda^2, \vec{\mathbf{v}})$ is an eigenpair of A, then the steps reverse to obtain $\lambda^2\vec{\mathbf{x}} = A\vec{\mathbf{x}}$, which means that $\vec{\mathbf{x}} = e^{\lambda t}\vec{\mathbf{v}}$ is a nonzero solution of $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$.

By linear algebra, the equation $A\vec{\mathbf{v}} = \lambda^2\vec{\mathbf{v}}$ has a solution $\vec{\mathbf{v}} \neq \vec{\mathbf{0}}$ if and only if the homogeneous problem $(A - \lambda^2 I)\vec{\mathbf{v}} = \vec{\mathbf{0}}$ has infinitely many solutions. Cramer's Rule implies this event happens exactly when $\det(A - \lambda^2 I) = 0$.

Characteristic Equation for $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$

The characteristic equation for the $n \times n$ second order system $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$ will be derived anew from the corresponding $2n \times 2n$ first order system $\vec{\mathbf{u}}' = C\vec{\mathbf{u}}$. We will prove the following identity.

Theorem 31 (Characteristic Equation)
Let $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$ be given with $n \times n$ constant matrix A. Let $\vec{\mathbf{u}}' = C\vec{\mathbf{u}}$ be its corresponding first order system, where

$$\vec{\mathbf{u}} = \begin{pmatrix} \vec{\mathbf{x}} \\ \vec{\mathbf{x}}' \end{pmatrix}, \quad C = \left(\begin{array}{c|c} 0 & I \\ \hline A & 0 \end{array}\right).$$

Then

$$\det(C - \lambda I) = (-1)^n \det(A - \lambda^2 I). \tag{3}$$

Proof: The method of proof is to verify the product formula

$$\left(\begin{array}{r|r} -\lambda I & I \\ \hline A & -\lambda I \end{array}\right)\left(\begin{array}{r|r} I & 0 \\ \hline \lambda I & I \end{array}\right) = \left(\begin{array}{r|r} 0 & I \\ \hline A - \lambda^2 I & -\lambda I \end{array}\right).$$

Then the determinant product formula applies to give

$$\det(C - \lambda I)\det\left(\begin{array}{r|r} I & 0 \\ \hline \lambda I & I \end{array}\right) = \det\left(\begin{array}{r|r} 0 & I \\ \hline A - \lambda^2 I & -\lambda I \end{array}\right). \tag{4}$$

Cofactor expansion is applied to give the two identities

$$\det\left(\begin{array}{r|r} I & 0 \\ \hline \lambda I & I \end{array}\right) = 1, \quad \det\left(\begin{array}{r|r} 0 & I \\ \hline A - \lambda^2 I & -\lambda I \end{array}\right) = (-1)^n \det(A - \lambda^2 I).$$

Then (4) implies (3). The proof is complete.

Solving $\vec{\mathbf{u}}' = C\vec{\mathbf{u}}$ and $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$

Consider the $n \times n$ second order system $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$ and its corresponding $2n \times 2n$ first order system $\vec{\mathbf{u}}' = C\vec{\mathbf{u}}$, where

$$(5) \qquad C = \left(\begin{array}{c|c} 0 & I \\ \hline A & 0 \end{array}\right), \quad \vec{\mathbf{u}} = \begin{pmatrix} \vec{\mathbf{x}} \\ \vec{\mathbf{x}}' \end{pmatrix}.$$

Theorem 32 (Eigenanalysis of A and C)
Let A be a given $n \times n$ constant matrix and define the corresponding $2n \times 2n$ system by

$$\vec{\mathbf{u}}' = C\vec{\mathbf{u}}, \quad C = \left(\begin{array}{c|c} 0 & I \\ \hline A & 0 \end{array}\right), \quad \vec{\mathbf{u}} = \begin{pmatrix} \vec{\mathbf{x}} \\ \vec{\mathbf{x}}' \end{pmatrix}.$$

Then

$$(6) \qquad (C - \lambda I)\begin{pmatrix} \vec{\mathbf{w}} \\ \vec{\mathbf{z}} \end{pmatrix} = \vec{\mathbf{0}} \quad \text{if and only if} \quad \left\{ \begin{array}{rcl} A\vec{\mathbf{w}} & = & \lambda^2 \vec{\mathbf{w}}, \\ \vec{\mathbf{z}} & = & \lambda \vec{\mathbf{w}}. \end{array} \right.$$

Proof: The result is obtained by block multiplication, because

$$C - \lambda I = \left(\begin{array}{c|c} -\lambda I & I \\ \hline A & -\lambda I \end{array}\right).$$

Theorem 33 (General Solutions of $\vec{\mathbf{u}}' = C\vec{\mathbf{u}}$ and $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$)
Let A be a given $n \times n$ constant matrix and define the corresponding $2n \times 2n$ system by

$$\vec{\mathbf{u}}' = C\vec{\mathbf{u}}, \quad C = \left(\begin{array}{c|c} 0 & I \\ \hline A & 0 \end{array}\right), \quad \vec{\mathbf{u}} = \begin{pmatrix} \vec{\mathbf{x}} \\ \vec{\mathbf{x}}' \end{pmatrix}.$$

Assume C has eigenpairs $\{(\lambda_j, \vec{\mathbf{y}}_j)\}_{j=1}^{2n}$ and $\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_{2n}$ are independent. Let I denote the $n \times n$ identity and define $\vec{\mathbf{w}}_j = \mathbf{diag}(I, 0)\vec{\mathbf{y}}_j$, $j = 1, \ldots, 2n$. Then $\vec{\mathbf{u}}' = C\vec{\mathbf{u}}$ and $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$ have general solutions

$$\begin{array}{rcll} \vec{\mathbf{u}}(t) & = & c_1 e^{\lambda_1 t}\vec{\mathbf{y}}_1 + \cdots + c_{2n} e^{\lambda_{2n} t}\vec{\mathbf{y}}_{2n} & (2n \times 1), \\ \vec{\mathbf{x}}(t) & = & c_1 e^{\lambda_1 t}\vec{\mathbf{w}}_1 + \cdots + c_{2n} e^{\lambda_{2n} t}\vec{\mathbf{w}}_{2n} & (n \times 1). \end{array}$$

Proof: Let $\vec{\mathbf{x}}_j(t) = e^{\lambda_j t}\vec{\mathbf{w}}_j$, $j = 1, \ldots, 2n$. Then $\vec{\mathbf{x}}_j$ is a solution of $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$, because $\vec{\mathbf{x}}_j''(t) = e^{\lambda_j t}(\lambda_j)^2\vec{\mathbf{w}}_j = A\vec{\mathbf{x}}_j(t)$, by Theorem 32. To be verified is the independence of the solutions $\{\vec{\mathbf{x}}_j\}_{j=1}^{2n}$. Let $\vec{\mathbf{z}}_j = \lambda_j \vec{\mathbf{w}}_j$ and apply Theorem 32 to write $\vec{\mathbf{y}}_j = \begin{pmatrix} \vec{\mathbf{w}}_j \\ \vec{\mathbf{z}}_j \end{pmatrix}$, $A\vec{\mathbf{w}}_j = \lambda_j^2 \vec{\mathbf{w}}_j$. Suppose constants $a_1, \ldots, a_{2n}$ are given such that $\sum_{j=1}^{2n} a_k \vec{\mathbf{x}}_j = 0$. Differentiate this relation to give $\sum_{j=1}^{2n} a_k e^{\lambda_j t}\vec{\mathbf{z}}_j = 0$ for all t. Set $t = 0$ in the last summation and combine to obtain $\sum_{j=1}^{2n} a_k \vec{\mathbf{y}}_j = 0$. Independence of $\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_{2n}$ implies that $a_1 = \cdots = a_{2n} = 0$. The proof is complete.

Eigenanalysis when A has Negative Eigenvalues. If all eigenvalues μ of A are negative or zero, then, for some $\omega \geq 0$, eigenvalue μ is related to an eigenvalue λ of C by the relation $\mu = -\omega^2 = \lambda^2$. Then $\lambda = \pm\omega i$ and $\omega = \sqrt{|\mu|}$. Consider an eigenpair $(-\omega^2, \vec{\mathbf{v}})$ of the real $n \times n$ matrix A with $\omega \geq 0$ and let

$$u(t) = \begin{cases} c_1 \cos\omega t + c_2 \sin\omega t & \omega > 0, \\ c_1 + c_2 t & \omega = 0. \end{cases}$$

Then $u''(t) = -\omega^2 u(t)$ (both sides are zero for $\omega = 0$). It follows that $\vec{\mathbf{x}}(t) = u(t)\vec{\mathbf{v}}$ satisfies $\vec{\mathbf{x}}''(t) = -\omega^2\vec{\mathbf{x}}(t)$ and $A\vec{\mathbf{x}}(t) = u(t)A\vec{\mathbf{v}} = -\omega^2\vec{\mathbf{x}}(t)$. Therefore, $\vec{\mathbf{x}}(t) = u(t)\vec{\mathbf{v}}$ satisfies $\vec{\mathbf{x}}''(t) = A\vec{\mathbf{x}}(t)$.

Theorem 34 (Eigenanalysis Solution of $\vec{\mathbf{x}}'' = A\vec{\mathbf{x}}$)
Let the $n \times n$ real matrix A have eigenpairs $\{(\mu_j, \vec{\mathbf{v}}_j)\}_{j=1}^n$. Assume $\mu_j = -\omega_j^2$ with $\omega_j \geq 0$, $j = 1, \ldots, n$. Assume that $\vec{\mathbf{v}}_1$, ..., $\vec{\mathbf{v}}_n$ are linearly independent. Then the general solution of $\vec{\mathbf{x}}''(t) = A\vec{\mathbf{x}}(t)$ is given in terms of $2n$ arbitrary constants a_1, ..., a_n, b_1, ..., b_n by the formula

$$\vec{\mathbf{x}}(t) = \sum_{j=1}^n \left(a_j \cos\omega_j t + b_j \frac{\sin\omega_j t}{\omega_j} \right) \vec{\mathbf{v}}_j \tag{7}$$

This expression uses the limit convention $\left.\dfrac{\sin\omega t}{\omega}\right|_{\omega=0} = t$.

Proof: The text preceding the theorem and superposition establish that $\vec{\mathbf{x}}(t)$ is a solution. It only remains to prove that it is the general solution, meaning that the arbitrary constants can be assigned to allow any possible initial condition $\vec{\mathbf{x}}(0) = \vec{\mathbf{x}}_0$, $\vec{\mathbf{x}}'(0) = \vec{\mathbf{y}}_0$. Define the constants uniquely by the relations

$$\begin{aligned} \vec{\mathbf{x}}_0 &= \textstyle\sum_{j=1}^n a_j \vec{\mathbf{v}}_j, \\ \vec{\mathbf{y}}_0 &= \textstyle\sum_{j=1}^n b_j \vec{\mathbf{v}}_j, \end{aligned}$$

which is possible by the assumed independence of the vectors $\{\vec{\mathbf{v}}_j\}_{j=1}^n$. Then equation (7) implies $\vec{\mathbf{x}}(0) = \sum_{j=1}^n a_j\vec{\mathbf{v}}_j = \vec{\mathbf{x}}_0$ and $\vec{\mathbf{x}}'(0) = \sum_{j=1}^n b_j\vec{\mathbf{v}}_j = \vec{\mathbf{y}}_0$. The proof is complete.

11.9 Numerical Methods for Systems

An initial value problem for a system of two differential equations is given by the equations

$$
\begin{array}{rcl}
x'(t) & = & f(t, x(t), y(t)), \\
y'(t) & = & g(t, x(t), y(t)), \\
x(t_0) & = & x_0, \\
y(t_0) & = & y_0.
\end{array}
\tag{1}
$$

A **numerical method** for (1) is an algorithm that computes an approximate dot table with first line t_0, x_0, y_0. Generally, the dot table has equally spaced t-values, two consecutive t-values differing by a constant value $h \neq 0$, called the **step size**. To illustrate, if $t_0 = 2$, $x_0 = 5$, $y_0 = 100$, then a typical dot table for step size $h = 0.1$ might look like

t	x	y
2.0	5.00	100.00
2.1	5.57	103.07
2.2	5.62	104.10
2.3	5.77	102.15
2.4	5.82	101.88
2.5	5.96	100.55

Graphics. The dot table represents the data needed to plot a solution curve to system (1) in three dimensions (t, x, y) or in two dimensions, using a tx-scene or a ty-scene. In all cases, the plot is a simple connect-the-dots graphic.

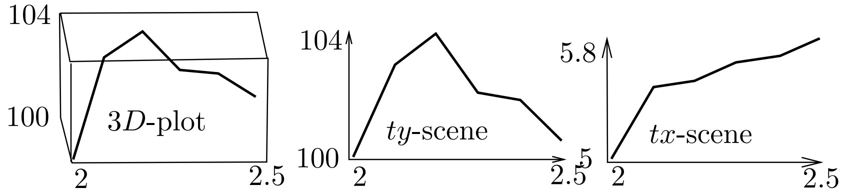

Figure 23. Dot table plots.
The three dimensional plot is a space curve made directly from the dot table. The tx-scene and the ty-scene are made from the same dot table using corresponding data columns.

Myopic Algorithms. All of the popular algorithms for generating a numerical dot table for system (1) are **near-sighted**, because they predict the next line in the dot table from the current dot table line, ignoring effects and errors for all other preceding dot table lines. Among such algorithms are **Euler's method**, **Heun's method** and the **RK4 method**.

Numerical Algorithms: Planar Case

Stated here without proof are three numerical algorithms for solving two-dimensional initial value problems (1). Justification of the formulas is obtained from the vector relations in the next subsection.

Notation. Let t_0, x_0, y_0 denote the entries of the dot table on a particular line. Let h be the increment for the dot table and let $t_0 + h$, x, y stand for the dot table entries on the next line.

Planar Euler Method.

$$\begin{aligned} x &= x_0 + hf(t_0, x_0, y_0), \\ y &= y_0 + hg(t_0, x_0, y_0). \end{aligned}$$

Planar Heun Method.

$$\begin{aligned} x_1 &= x_0 + hf(t_0, x_0, y_0), \\ y_1 &= y_0 + hg(t_0, x_0, y_0), \\ x &= x_0 + h(f(t_0, x_0, y_0) + f(t_0 + h, x_1, y_1))/2 \\ y &= y_0 + h(g(t_0, x_0, y_0) + g(t_0 + h, x_1, y_1))/2. \end{aligned}$$

Planar RK4 Method.

$$\begin{aligned} k_1 &= hf(t_0, x_0, y_0), \\ m_1 &= hg(t_0, x_0, y_0), \\ k_2 &= hf(t_0 + h/2, x_0 + k_1/2, y_0 + m_1/2), \\ m_2 &= hg(t_0 + h/2, x_0 + k_1/2, y_0 + m_1/2), \\ k_3 &= hf(t_0 + h/2, x_0 + k_2/2, y_0 + m_2/2), \\ m_3 &= hg(t_0 + h/2, x_0 + k_2/2, y_0 + m_2/2), \\ k_4 &= hf(t_0 + h, x_0 + k_3, y_0 + m_3), \\ m_4 &= hg(t_0 + h, x_0 + k_3, y_0 + m_3), \\ x &= x_0 + \frac{1}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right), \\ y &= y_0 + \frac{1}{6}\left(m_1 + 2m_2 + 2m_3 + m_4\right). \end{aligned}$$

Numerical Algorithms: General Case

Consider a vector initial value problem

$$\vec{\mathbf{u}}'(t) = \vec{\mathbf{F}}(t, \vec{\mathbf{u}}(t)), \quad \vec{\mathbf{u}}(t_0) = \vec{\mathbf{u}}_0.$$

Described here are the vector formulas for Euler, Heun and RK4 methods. These myopic algorithms predict the next table dot $t_0 + h$, $\vec{\mathbf{u}}$ from the current dot t_0, $\vec{\mathbf{u}}_0$. The number of scalar values in a table dot is $1 + n$, where n is the dimension of the vectors $\vec{\mathbf{u}}$ and $\vec{\mathbf{F}}$.

Vector Euler Method.

$$\vec{\mathbf{u}} = \vec{\mathbf{u}}_0 + h\vec{\mathbf{F}}(t_0, \vec{\mathbf{u}}_0)$$

Vector Heun Method.

$$\vec{\mathbf{w}} = \vec{\mathbf{u}}_0 + h\vec{\mathbf{F}}(t_0, \vec{\mathbf{u}}_0), \quad \vec{\mathbf{u}} = \vec{\mathbf{u}}_0 + \frac{h}{2}\left(\vec{\mathbf{F}}(t_0, \vec{\mathbf{u}}_0) + \vec{\mathbf{F}}(t_0 + h, \vec{\mathbf{w}})\right)$$

Vector RK4 Method.

$$\begin{array}{rcl}
\vec{\mathbf{k}}_1 & = & h\vec{\mathbf{F}}(t_0, \vec{\mathbf{u}}_0), \\
\vec{\mathbf{k}}_1 & = & h\vec{\mathbf{F}}(t_0 + h/2, \vec{\mathbf{u}}_0 + \vec{\mathbf{k}}_1/2), \\
\vec{\mathbf{k}}_1 & = & h\vec{\mathbf{F}}(t_0 + h/2, \vec{\mathbf{u}}_0 + \vec{\mathbf{k}}_2/2), \\
\vec{\mathbf{k}}_1 & = & h\vec{\mathbf{F}}(t_0 + h, \vec{\mathbf{u}}_0 + \vec{\mathbf{k}}_3), \\
\vec{\mathbf{u}} & = & \vec{\mathbf{u}}_0 + \frac{1}{6}\left(\vec{\mathbf{k}}_1 + 2\vec{\mathbf{k}}_2 + 2\vec{\mathbf{k}}_3 + \vec{\mathbf{k}}_4\right).
\end{array}$$