Applications

This learning module covers the following ideas. When you make your lesson plan, it should explain and contain examples of the following:

1. Use Gaussian elimination to find least degree interpolating polynomials, compute partial fraction decompositions, and solve for electrical currents in electrical systems involving batteries and resistors.

2. Explain how to generalize the derivative to a matrix. Use this generalization to approximate changes in a function and locate optimal values of the function using the second derivative test.

3. Describe a Markov Process, and connect it to powers of a square matrix. Explain how eigenvectors are related to limits of Markov Processes.

4. Find bases of vectors for the row and column spaces of a matrix. Explain how to decompose a matrix into the product of matrices involving these bases, and how to each row and column of your matrix can be expressed as a linear combination of these bases.

5. Use the transpose and inverse of a matrix to solve the least squares regression problem of fitting a line to a set of data. Describe and illustrate Cramer's rule, and use it to obtain a complete solution to the least squares problem.

Following you will find preparation assignments and suggested homework, and then my best attempt at condensing the information we are learning into a concise set of notes. For this unit, these are intended to be a complete set of notes. The homework for this unit is found in the accompany applications problems. As you read my notes, please do the examples yourself and then find examples like them in the homework. With these notes I am trying follow the model found in "Preach My Gospel," where the gospel is taught in a complete concise manner, with suggestions provided for additional study and practice.

1 Preparation and Suggested Homework

As a group of 4, each of you should select one of the following problems. To be prepared for class, you should complete the problem you selected and come to class prepared to show your group how that problem is solved. Occasionally you will be asked to teach this problem to your group, but usually we will attempt similar problems in class. By coming prepared, you will serve as a mini expert group in your group. If you have spent 20-30 minutes on the problem, looked for and tried similar problems (including the online solutions), and are still stuck, then count yourself as prepared. I strongly suggest that you attempt all the others problems (5 minutes with each is sufficient) before coming to class. If you do not understand a problem, that is OK; it means you are learning and have questions that we can answer in class. This preparation activity should take no more than 30 minutes of your time. Please use the rest of your time to solve other homework problems related to what we are learning, and complete corrections as needed.

	Preparation Problems
Day 1	2 , 7, 13, 14
Day 2	23, 29, 37, 41
Day 3	49, 51, 54, 65

The homework problems for this unit come from the accompanying applications problems list. I strongly suggest that you do 2-3 of each type of problem. The technology introduction contains step-by-step solutions for many of the problems, so that you can check your work. I will be spending the next few days creating handwritten solutions for many of the problems.

Minimum Homework Goal: 28 - Minimum Technology Goal: 8 (You should do at least 28 problems, at least 8 of which you solved using Maple. You will report what you have done via a stewardship report.)

2 Solving Systems

2.1 Fitting a Polynomial

Through any two points (with different x values) there is a unique line of the form $y = mx + b$. If you know two points, then you can use them to find the values m and b. Through any 3 points (with different x values) there is a unique parabola of the form ($y = ax^2 + bx + c$), and you can use the 3 points to find the values a, b, c. As you increase the number of points, there is still a unique polynomial with degree one less than the number of points, and you can use the points to find the coefficients of your polynomial. In this section we will illustrate how this is done, and how the solution results in a linear system.

In order to organize our work, we first standardize the notation. Rather than writing $y = mx+b$, we write $y = a_0 + a_1 x$ (where $a_0 = b$ and $a_1 = m$). For a parabola, we write $y = a_0 + a_1 x + a_2 x^2 = \sum_{k=0}^{2} a_k x^k$. This allows us to write any degree polynomial in the form $y = a_0 + a_1 x + \cdots + a_n x^n = \sum_{k=0}^{n} a_k x^k$ (by standardizing what we call the coefficients, we can use summation notation to express any degree polynomial by changing the n on the top of the summation sign).

Now that our notation is organized, let's find a degree 2 polynomial through the three points $(0, 1), (2, 3), (4, 7)$. The polynomial is $y = a_0 + a_1 x + a_2 x^2$, where our job is to find the three constants a_0, a_1, a_2. Since we have three points, put those points into the equation to obtain the three equations

$$a_0 = 1 \qquad a_0 + 2\,a_1 + 4\,a_2 = 3 \qquad a_0 + 4\,a_1 + 16\,a_2 = 7$$

This is a system with 3 equations and 3 unknowns. Write the system in matrix form and reduce it.

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 1 & 2 & 4 & 3 \\ 1 & 4 & 16 & 7 \end{array}\right] \xrightarrow{\text{rref}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & 1/4 \end{array}\right]$$

The solution is $a_0 = 1, a_1 = 1/2, a_2 = 1/4$, or $y = 1 + \frac{1}{2}x + \frac{1}{4}x^2$.

In the solution above, notice that powers of the x values appear as the coefficients of our coefficient matrix, and we augment that matrix by the y values. This is the general pattern used in finding an interpolating polynomial. If there are 4 points, then you just add one row and one column to the matrix. The figure below shows the general method for quadratics and cubics.

$$(x_1, y_1), (x_2, y_2), (x_3, y_3)$$
$$\left[\begin{array}{ccc|c} x_1^0 & x_1^1 & x_1^2 & y_1 \\ x_2^0 & x_2^1 & x_2^2 & y_2 \\ x_3^0 & x_3^1 & x_3^2 & y_3 \end{array}\right] \xrightarrow{\text{rref}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & a_0 \\ 0 & 1 & 0 & a_1 \\ 0 & 0 & 1 & a_2 \end{array}\right]$$
$$y = a_0 + a_1 x + a_2 x^2$$

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$$
$$\left[\begin{array}{cccc|c} x_1^0 & x_1^1 & x_1^2 & x_1^3 & y_1 \\ x_2^0 & x_2^1 & x_2^2 & x_2^3 & y_2 \\ x_3^0 & x_3^1 & x_3^2 & x_3^3 & y_3 \\ x_4^0 & x_4^1 & x_4^2 & x_4^3 & y_4 \end{array}\right] \xrightarrow{\text{rref}} \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & a_0 \\ 0 & 1 & 0 & 0 & a_1 \\ 0 & 0 & 1 & 0 & a_2 \\ 0 & 0 & 0 & 1 & a_3 \end{array}\right]$$
$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

2.2 Partial Fraction Decomposition

A partial fraction decomposition is a method of breaking a complex rational function up into the sum of smaller simpler functions to work with. Some places you will see partial fraction decompositions are in techniques of integration (in math 113) and differential equations (math 316 or math 371). To illustrate their value, let's start with an example.

Let's find the integral of the function $f(x) = \dfrac{2x+1}{(x-2)\,(x-3)}$. The denominator is the product of two linear functions. Is it possible to break up the function into two simpler functions, namely can we write

$$\frac{2\,x+1}{(x-2)\,(x-3)} = \frac{A}{x-2} + \frac{B}{x-3}$$

for unknown constants A and B? If we multiply both sides by the original denominator, we obtain (cancel the common factors)

$$2x + 1 = A(x - 3) + B(x - 2).$$

Now expand the right hand side and collect the terms which have the same powers of x,

$$2x + 1 = (A + B)x + (-3A - 2B).$$

Both sides of the equation above represent lines. In order for the two lines to be the same line, they must have the same slope and intercept. This means we can create an equation for each power of x by equating the coefficients on both sides of the equation. This gives us the two equations

$$2 = A + B \qquad 1 = -3A - 2B.$$

This is now a linear system and we can use Gaussian elimination to solve it as follows:

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ -3 & -3 & 1 \end{array}\right] \xrightarrow{\text{rref}} \left[\begin{array}{cc|c} 1 & 0 & -5 \\ 0 & 1 & 7 \end{array}\right]$$

Our solution is $A = -5, B = 7$ and we can write $\frac{2\,x+1}{(x-2)(x-3)} = \frac{-5}{x-2} + \frac{7}{x-3}$. We can now integrate each term separately to obtain

$$\int \frac{2\,x+1}{(x-2)\,(x-3)}dx = \int \frac{-5}{x-2}dx + \int \frac{7}{x-3}dx = -5\ln|x-2| + 7\ln|x-3|.$$

The general process for finding a partial fraction decomposition requires that you start with an appropriate guess for the final form, multiply both sides by the original denominator, collect like powers of x on both sides, and then solve the corresponding linear system. Provided the denominator can be factored, this process can always be used to obtain an integral of a rational function. In math 113 we discuss how to choose an appropriate guess, and in differential equations partial fraction are used to solve differential equations with Laplace transforms. For the problems on the homework, I will provide the guessed form. Your job will be to find the unknowns constants in the form, and then integrate.

As a final example, let's compute $\int \frac{-x^2+2\,x+5}{(x^2+1)(x-3)}dx$, using the form

$$\frac{-x^2 + 2\,x + 5}{(x^2+1)\,(x-3)} = \frac{Ax+B}{x^2+1} + \frac{C}{x-3}$$

In this case the denominator doesn't factor into a product of linear terms, so the quadratic term $x^2 + 1$ has a linear term $Ax + B$ in the numerator. Multiplying both sides by the denominator and collecting powers of x gives

$$-x^2 + 2\,x + 5 = (A + C)\,x^2 + (B - 3\,A)\,x + (C - 3\,B).$$

Equating the coefficients of x on each side gives the three equations

$$5 = C - 3B, 2 = B - 3A, -1 = A + C$$

Rewriting in matrix form and reducing the matrix gives us

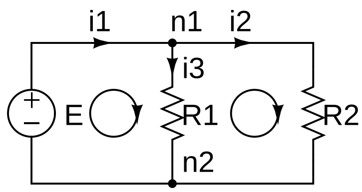
$$\left[\begin{array}{ccc|c} 0 & -3 & 1 & 5 \\ -3 & 1 & 0 & 2 \\ 1 & 0 & 1 & -1 \end{array}\right] \xrightarrow{\text{rref}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & -6/5 \\ 0 & 1 & 0 & -8/5 \\ 0 & 0 & 1 & 1/5 \end{array}\right].$$

We can now integrate using our solution to obtain

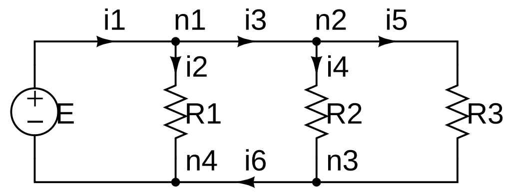
$$\begin{aligned}\int \frac{-x^2+2\,x+5}{(x^2+1)\,(x-3)}dx &= \int \frac{1}{5}\left(\frac{-6x-8}{x^2+1}\right) + \frac{1}{5}\left(\frac{1}{x-3}\right)dx \\ &= -\frac{6}{5}\int \frac{x}{x^2+1}dx - \frac{8}{5}\int \frac{1}{x^2+1}dx + \frac{1}{5}\int \frac{1}{x-3}dx \\ &= -\frac{6}{10}\ln|x^2+1| - \frac{8}{5}\arctan x + \frac{1}{5}\ln|x-3| + C.\end{aligned}$$

2.3 Kirchoff's Electrical Laws

Gustav Kirchoff discovered two laws that pertain the conservation of charge and energy in the study of electricity. In order to describe these laws, we first need to discuss voltage, resistance, and current. Current is the flow of electricity, and often it can be compared to the flow of water. Whenever a current passes across a conductor, it encounters resistance. Ohm's law states that the product of the resistance R and current I across a conductor equals the voltage V, or $RI = V$. If the voltage remains constant, then a large resistance corresponds to a small current. A resistor is an object with high resistance which is placed in an electrical system to slow down the flow (current) of electricity. Resistors are measured in terms of ohms, and the larger the ohms, the smaller the current. The following diagrams illustrate two introductory electrical systems.



Two Loop System



Three Loop System

Wires in this diagram meet at nodes (which are illustrated with a dot, and are labeled with n's). Batteries (or voltage sources) are often depicted with the symbol , and they supply a voltage of E volts. The electrical current on each wire flows away from the positive end of a battery and toward the negative end. At each node the current may change, so the arrows and letters i represent the different currents in the electrical system. Resistors are depicted with the symbol , and the letter R represents the ohms.

Kirchoff discovered two laws. They both help us find current in a system, provided we know the voltage of any batteries, and the resistance of any resistors.

1. Kirchoff's current law states that at every node, the current flowing in equals the current flowing out (at nodes, current in = current out).

2. Kirchoff's voltage law states that on any loop in the system, the directed sum of voltages supplied equals the directed sum of voltage drops (in loops, voltage in = voltage out).

Let's use Kirchoff's laws to generate a system of equations for the two loop system. At the first node (n_1), current i_1 flows in while i_2 and i_3 flow out. Kirchoff's current law states that $i_1 = i_2 + i_3$ or $i_1 - i_2 - i_3 = 0$. At the second node, both i_2 and i_3 are flowing in while i_1 flows out. This means that $i_2 + i_3 = i_1$ or $-i_1 + i_2 + i_3 = 0$. This second equation is the same multiplying both sides of the first by -1. We now look at Kirchoff's voltage law. Pick a loop and work your way around the loop in a clockwise fashion. Each time you encounter a battery or resistor, include a term for the voltage supplied E on the left side of an equation, and the voltage drop (resistance times current Ri) on the right. If you encounter a battery or resistor as you work against the current, then times that term by a negative 1. The left loop has a battery with voltage E and the resistor contributes a drop in voltage of $R_1 i_2$ volts. An equation for the first loop is $E = R_1 i_2$. On the right loop we encounter along the current i_3 a resistor with resistance R_2 ohms. While working our way up i_2 (against the flow), we encounter a R_1 ohm resistor. There are no batteries. This gives us the equation $0 = -R_1 i_2 + R_2 i_3$. We can now write a system of equations involving the unknowns i_1, i_2, i_3, put it in matrix form and solve

$$\begin{cases} i_1 - i_2 - i_3 = 0 \\ -i_1 + i_2 + i_3 = 0 \\ R_1 i_2 = E \\ -R_1 i_2 + R_2 i_3 = 0 \end{cases} \xrightarrow{\text{matrix form}} \left[\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ -1 & 1 & 1 & 0 \\ 0 & R_1 & 0 & E \\ 0 & -R_1 & R_2 & 0 \end{array}\right] \xrightarrow{\text{rref}} \left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{E}{R_1} + \frac{E}{R_2} \\ 0 & 1 & 0 & \frac{E}{R_1} \\ 0 & 0 & 1 & \frac{E}{R_2} \\ 0 & 0 & 0 & 0 \end{array}\right].$$

The reason we have a row of zeros at the bottom of or system is because the two rows corresponding to the nodes are linearly dependent. Hence, when we reduce the matrix that dependence relation becomes a row of zeros.

A similar computation can be done for the three loop system. There are 6 unknown currents, 4 nodes, and 3 loops. This will give us 7 equations with 6 unknowns. The 4 equations from the nodes will again contribute rows which are linearly dependent. Reduction will give a unique solution. In the homework, you are asked to setup systems of equations for various electrical systems, and then solve them.

3 Calculus

3.1 Differentials and Approximation

3.1.1 Differential Notation Review

The derivative $f'(x)$ is often interpreted as slope at a point. The slope between two points is the rise over the run, the change in y over the change in x, or symbolically we write the slope as $\frac{\Delta y}{\Delta x}$ The notation $\frac{dy}{dx}$ conveys the idea that the derivative is really slope at a point. The notation $dy = f'dx$ reminds us that an actual change Δy in a function can be found using the approximation $\Delta y \approx dy = f'dx$.

For example, when constructing a circle of radius 3 inches, the area $A = \pi r^2$ would increase if a manufacturing process instead creates a circle of radius 3.1 inches. By about how much would the area of the circle increase? The differential $dA = A'dr = 2\pi r dr = 2\pi(2)(.1) = .6\pi \approx 1.88$ gives an approximate increase in the area. The actual increase in the area is $\Delta A = (3.1)^2\pi - (3)^2\pi \approx 1.92$. Notice that $dA \approx \Delta A$, namely the differential of the area $dA \approx 1.88$ is approximately the same as the actual change in the area $\Delta A \approx 1.92$.

3.1.2 Functions of the form $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$

Differential notation is used to study functions with any number of inputs or outputs. The word function requires that when we write $y = f(x)$, there must be one output y for every input x. In first semester calculus, you study what happens if your inputs and outputs are both numbers. The fact that the input and output are both single numbers is expressed by the notation $f : \mathbb{R} \to \mathbb{R}$, where $\mathbb{R}$ means all real numbers. In this class and beyond, we study what happens if the inputs and outputs are vectors, not just numbers. The notation $\mathbb{R}^2$ represents the vectors in the plane, while $\mathbb{R}^3$ represents the vectors in space and $\mathbb{R}^n$ represents vectors $(x_1, x_2, x_3, \ldots, x_n)$ with n components.

The function $f(x, y) = 9 - x^2 - y^2$ has a vector (x, y) of two inputs, and returns a number as its output. We can write $f : \mathbb{R}^2 \to \mathbb{R}^1$ to explicitly remind us about the size of our inputs and outputs. The functions $x = \cos t, y = \sin t$ can be combined into one function $\vec{r}(t) = (\cos t, \sin t)$ where we input one variable t and get out a vector $(x, y) = (\cos t, \sin t)$ with two components. The notation $\vec{r}(t) : \mathbb{R}^1 \to \mathbb{R}^2$ reminds us of the size of the inputs and outputs.

In multivariable calculus, we study functions of the form $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$ for $n, m \leq 3$, and learn many uses of these functions. In linear algebra, we will study "linear" functions of the form $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$ for any finite n, m.

3.1.3 Partial Derivatives

Recall the derivative of a function with one input is defined to be $\dfrac{df}{dx} = \lim\limits_{h\to 0} \dfrac{f(x+h) - f(x)}{h}$. The derivative gives the best possible linear approximation to changes in a function. This idea is represented by the equation $\Delta y \approx f'(x)\Delta x$, or in differential form we write $dy = f'(x)dx$. An equation of a tangent line is found by noticing that a change in y is approximately $y - f(c)$ when the change in x is $x - c$, so the differential form $dy = f'dx$ becomes $(y - f(c)) = f'(c)(x - c)$.

If a function has more than one input variable, then division by the vector $\vec{h}$ is not well-defined, so we run into a problem with generalizing derivatives. Instead, we compute partial derivatives which approximate change in the function if we hold all other variables constant and just differentiate with respect to one variable. For the function $f(x, y)$, we define the partial derivative of f with respect to x as $f_x(x, y) = f_x = \dfrac{\partial f}{\partial x} = \lim\limits_{h\to 0} \dfrac{f(x+h, y) - f(x, y)}{h}$, and the partial derivative of f with respect to y as $f_y(x, y) = f_y = \dfrac{\partial f}{\partial y} = \lim\limits_{k\to 0} \dfrac{f(x, y+k) - f(x, y)}{k}$. Notice that f_x computes a limit as y is held constant and we vary x.

For $f(x,y) = 3x^2+4xy+\cos(xy)+y^3$, we obtain $f_x = 6x+4y-y\sin(xy)+0$ and $f_y = 0+4x-x\sin(xy)+3y^2$. Partial derivatives are found by holding all other variables constant, and then differentiating with respect to the variable in question.

3.1.4 The Derivative

The derivative of a function $\vec{f}:\mathbb{R}^n \to \mathbb{R}^m$ is an $m \times n$ matrix written $D\vec{f}(\vec{x})$, where the columns of the matrix are the partial derivatives of the function with respect to an input variable (the first column is the partial derivative with respect to the first variable, and so on). Some people call this derivative the "total" derivative instead of the derivative, to emphasize that the "total" derivative combines the "partial" derivatives into a matrix. This definition of the derivative gives the best possible linear approximation to changes in a function.

Some examples of functions and their derivative follow. Remember that each input variable corresponds to a column of the matrix.

Function	Derivative
$f(x) = x^2$	$Df(x) = \begin{bmatrix} 2x \end{bmatrix}$
$\vec{r}(t) = \langle 3\cos(t), 2\sin(t)\rangle$	$D\vec{r}(t) = \begin{bmatrix} -3\sin t \\ 2\cos t \end{bmatrix}$
$\vec{r}(t) = \langle \cos(t), \sin(t), t\rangle$	$D\vec{r}(t) = \begin{bmatrix} -\sin t \\ \cos t \\ 1 \end{bmatrix}$
$f(x,y) = 9-x^2-y^2$	$Df(x,y) = \nabla f(x,y) = \begin{bmatrix} -2x & -2y \end{bmatrix}$
$f(x,y,z) = x^2+y+xz^2$	$Df(x,y,z) = \nabla f(x,y,z) = \begin{bmatrix} 2x+z^2 & 1 & 2xz \end{bmatrix}$
$\vec{F}(x,y) = \langle -y, x\rangle$	$D\vec{F}(x,y) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$
$\vec{F}(r,\theta,z) = \langle r\cos\theta, r\sin\theta, z\rangle$	$D\vec{F}(r,\theta,z) = \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$
$\vec{r}(u,v) = \left\langle u, v, 9-u^2-v^2\right\rangle$	$D\vec{r}(u,v) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -2u & -2v \end{bmatrix}$

3.1.5 Approximation

To emphasize that the derivative is the best possible linear approximation to changes in a function, replace f' in the single variable equation $dy = f'dx$ with the derivative $D\vec{f}(\vec{x})$ to obtain $d\vec{y} = D\vec{f}(\vec{x})d\vec{x}$ (d[outputs] $= D\vec{f}d$[inputs]). For a function $z = f(x,y)$ we obtain $dz = Df(x,y)\begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} f_x & f_y \end{bmatrix}\begin{bmatrix} dx \\ dy \end{bmatrix} = f_x dx + f_y dy$ using matrix multiplication.

Suppose a manufacturing company wants to create a cylinder with a 1 inch radius, and 5 inches tall (something like a soda can). If the machine makes cylinders which have a 1.1 inch radius and are 4.8 inches tall, how much will that affect the volume of soda that can be put inside the can? Using differential notation, we write $V = \pi r^2 h, r = 1, h = 5, dr = .1, dh = -.2$. The derivative of volume is the matrix $DV(r,h) = \begin{bmatrix} V_r & V_h \end{bmatrix} = \begin{bmatrix} 2\pi rh & \pi r^2 \end{bmatrix}$. The approximate change in volume is

$$dV = DV\begin{bmatrix} dr \\ dh \end{bmatrix} = \begin{bmatrix} 2\pi rh & \pi r^2 \end{bmatrix}\begin{bmatrix} dr \\ dh \end{bmatrix} = \begin{bmatrix} 10\pi & \pi \end{bmatrix}\begin{bmatrix} .1 \\ -.2 \end{bmatrix} = \pi - .2\pi = .8\pi \approx 2.5133.$$

The actual change in volume is $\pi(1.1)^2(4.8) - \pi(1)^2(5) \approx 2.5384$. The actual change is pretty close to the estimated change.

Let's now look at an example where the number of outputs is more than 1. Consider the equations $x = t+1$ and $y = t^2$, or in vector form $\vec{r}(t) = \left\langle t+1, t^2\right\rangle$. This vector equation models the path of an object as it moves through the plane. At $t = 3$, the object is at $\vec{r}(3) = \left\langle 3+1, 3^2\right\rangle = \langle 4, 9\rangle$. About where will the object be at $t = 3.1$? Differentials say $d\vec{r} = D\vec{r}dt$, or

$$d\vec{r} = \begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} 1 \\ 2t \end{bmatrix} dt = \begin{bmatrix} 1 \\ 6 \end{bmatrix}(.1) = \begin{bmatrix} .1 \\ .6 \end{bmatrix}.$$

This means the x value has increased .1 to 4.1, and the y values increase .6 to 9.6.

3.2 The Second Derivative Test

Let's start with a review from first semester calculus. If a function $y = f(x)$ has a relative extremum at $x = c$, then $f'(c) = 0$ or the derivative is undefined. The places where the derivative is either zero or undefined are called critical values of the function. The first derivative test allows you to check the value of the derivative on both sides of the critical value and then interpret whether that point is a maximum or minimum using increasing/decreasing arguments. The second derivative test requires you to compute the second derivative at $x = c$. If $f''(c) > 0$ (the function is concave upwards), then the function has a minimum at $x = c$. If $f''(c) < 0$ (the function is concave downwards), then the function has a maximum at $x = c$. If $f''(c) = 0$, then the second derivative test fails.

The function $f(x) = x^3 - 3x$ has derivatives $f' = 3x^2 - 3$ and $f'' = 6x$. The first derivative is zero when $3(x^2 - 1) = 3(x - 1)(x + 1) = 0$, or $x = \pm 1$. The second derivative at $x = 1$ is 6 (concave upwards), so there is a minimum at $x = 1$. The second derivative at $x = -1$ is -6 (concave downwards), so there is a maximum at that point.

We're now ready to extend this idea to all functions of the form $f : \mathbb{R}^n \to \mathbb{R}$ (the output is 1 dimensional, so that it makes sense to talk about a largest or smallest number). We will only consider the case $n = 2$, as it simplifies the computations and provides all that is needed to extend to all dimensions. The first derivative test breaks down in every dimension past the first, because there are more than 2 ways to approach a point of the domain (you can't just look at the left side or right side). However, at a local extremum, the derivative is still zero, which often results in solving a system of equations. In higher dimensions, there are three classifications of critical points: maximum, minimum, or saddle point (a point where the tangent plane is horizontal, but in some directions you increase and in other directions you decrease).

The second derivative test does not break down. Consider the function $z = f(x, y)$. Its derivative $Df(x, y) = \begin{bmatrix} f_x & f_y \end{bmatrix}$ is a function with two inputs and two outputs. The second derivative $D^2f(x, y) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$ is a 2×2 square matrix called the Hessian of f. This matrix will always be symmetric, in that the transpose of the matrix equals itself. The eigenvalues of D^2f give the "directional" second derivative in the direction of a corresponding eigenvector. The largest eigenvalue is the largest possible value of the second derivative in any direction. The smallest eigenvalue is the smallest possible value of the second derivative in any direction. The **second derivative test** is the following. Start by finding all the critical points (places where the derivative is zero). Then find the eigenvalues of the second derivative.

1. If the eigenvalues are all positive at a critical point, then in every direction the function is concave upwards. The function has a minimum at that critical point.

2. If the eigenvalues are all negative at a critical point, then in every direction the function is concave downwards. The function has a maximum there.

3. If there is a positive eigenvalue and a negative eigenvalue, the function has a saddle point there.

4. If either the largest or smallest eigenvalue is zero, then the second derivative test fails.

Eigenvalues are the key numbers needed to generalize optimization to all dimensions. A proof of this fact is beyond the scope of this class.

For the function $f(x, y) = x^2 + xy + y^2$, the gradient is $Df = \begin{bmatrix} 2x + y & x + 2y \end{bmatrix}$, which is zero only at $x = 0, y = 0$ (solve the system of equations $2x + y = 0, x + 2y = 0$). The Hessian is $D^2f = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. The eigenvalues are found by solving $0 = \det \begin{bmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{bmatrix} = (2 - \lambda)^2 - 1 = 4 - 4\lambda + \lambda^2 - 1 = (\lambda - 3)(\lambda - 1)$, so $\lambda = 3, 1$ are the eigenvalues. Since both eigenvalues are positive, the function is concave upwards in all directions, so there is a minimum at $(0, 0)$.

For the function $f(x, y) = x^3 - 3x + y^2 - 4y$, the gradient is $Df = \begin{bmatrix} 3x^2 - 3 & 2y - 4 \end{bmatrix}$, which is zero at $x = 1, y = 2$ or $x = -1, y = 2$. Hence there are two critical points, so we have to find two sets of eigenvalues.

The Hessian is $D^2 f = \begin{bmatrix} 6x & 0 \\ 0 & 2 \end{bmatrix}$. When $x = -1, y = 2$, the eigenvalues of $\begin{bmatrix} -6 & 0 \\ 0 & 2 \end{bmatrix}$ are $\lambda = -6, 2$. Since one is positive and one is negative, there is a saddle point at $(-1, 2)$. When $x = 1, y = 2$, the eigenvalues of $\begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}$ are $\lambda = 6, 2$. Since both are positive, there is a minimum at $(-1, 2)$ (as in every direction the function is concave upwards).

3.3 Markov Process

Matrices can be used to model a process called a Markov Process. To fit this kind of model, a process must have specific states, and the matrix which models the process is a transition matrix which specifies how each state will change through a given transition. An example of a set of states is “open” or “closed” in an electrical circuit, or “working properly” and “working improperly” for operation of machinery at a manufacturing facility. A car rental company which rents vehicles in different locations can use a Markov Process to keep track of where their inventory of cars will be in the future. Stock market analysts use Markov processes and a generalization called stochastic processes to make predictions about future stock values.

We'll illustrate a Markov Process related to classifying land in some region as “Residential,” “Commercial,” or “Industrial.” Suppose in a given region over a 5 year time span that 80% of residential land will remain residential, 10% becomes commercial, and 10% becomes industrial. For commerical land, 70% remains commercial, 20% becomes residential, and 10% becomes industrial. For industrial land, 70% remains industrial, 30% becomes commercial, and 0% becomes residential. The “transition” matrix for a Markov process (the matrix on the left below) is a matrix where each column relates to one of the “states,” and the number in each row is the proportion of the column state that will change to the row state through the transition (where the ordering on row and column states is the same). We can calculate the next “state” by multiplying a current state by this transition matrix. In the land use model described above, the transition matrix is shown below. If current land use is about 50% residential, 30% commercial, and 20% industrial, then 5 years later the land use is calculated below.

$$\begin{array}{c|c} \begin{array}{rc} & \begin{array}{ccc} R & C & I \end{array} \\ \begin{array}{r} \text{to } R \\ \text{to } C \\ \text{to } I \end{array} & \begin{bmatrix} .8 & .2 & 0 \\ .1 & .7 & .3 \\ .1 & .1 & .7 \end{bmatrix} \end{array} & \begin{bmatrix} .8 & .2 & 0 \\ .1 & .7 & .3 \\ .1 & .1 & .7 \end{bmatrix} \begin{bmatrix} 50 \\ 30 \\ 20 \end{bmatrix} = \begin{bmatrix} 46 \\ 32 \\ 22 \end{bmatrix} . \\ \text{Transition Matrix} & \text{5 Year Projection} \end{array}$$

If the same transitions in land use continue, we multiply the previous projection (state) by the transition matrix to obtain a 10, 15, or 20 year projection for land use as shown below:

$$\underset{\text{10 Year Projection}}{\begin{bmatrix} .8 & .2 & 0 \\ .1 & .7 & .3 \\ .1 & .1 & .7 \end{bmatrix} \begin{bmatrix} 46 \\ 32 \\ 22 \end{bmatrix} = \begin{bmatrix} 43.2 \\ 33.6 \\ 23.2 \end{bmatrix}} \quad \underset{\text{15 Year Projection}}{\begin{bmatrix} .8 & .2 & 0 \\ .1 & .7 & .3 \\ .1 & .1 & .7 \end{bmatrix} \begin{bmatrix} 43.2 \\ 33.6 \\ 23.2 \end{bmatrix} = \begin{bmatrix} 41.28 \\ 34.8 \\ 23.92 \end{bmatrix}} \quad \underset{\text{20 Year Projection}}{\begin{bmatrix} .8 & .2 & 0 \\ .1 & .7 & .3 \\ .1 & .1 & .7 \end{bmatrix} \begin{bmatrix} 41.28 \\ 34.8 \\ 23.92 \end{bmatrix} = \begin{bmatrix} 39.984 \\ 35.664 \\ 24.352 \end{bmatrix}}$$

As we continue to multiply on the left by our transition matrix, each time we add 5 more years to our projection. This projection is valid as long as the same trends continue. Will the projections eventually stabilize, reaching a steady state?

If $\vec{x}_0$ is our initial state, then the product $A^n \vec{x}_0 = \vec{x}_n$ gives us the state after n transitions. What happens in the limit as $n \to \infty$? If the current trends in land use continue, what portion of the land will eventually be in each state. Can we reach a state $\vec{x} = (R, C, I)$ such that $A\vec{x} = \vec{x}$, the next state is the same as the current? If this occurs, then any future transitions will not change the state either. This state $\vec{x}$ is called a steady-state, since it does not change when multiplied by the transition matrix. Finding a steady state is an eigenvalue problem, as we are looking for a solution to the equation $A\vec{x} = 1\vec{x}$ (where the eigenvalue is 1). For any Markov process (where the columns of the matrix sum to 1), the number 1 will always be an eigenvalue. The solution to the problem $\lim\limits_{n\to\infty} A^n \vec{x}_0$ is a steady state. For our Markov process, an eigenvector (using technology) corresponding to 1 is $\begin{bmatrix} \frac{3}{2} & \frac{3}{2} & 1 \end{bmatrix}^T$. Multiplying by 2 we have $\begin{bmatrix} \frac{3}{2} & \frac{3}{2} & 1 \end{bmatrix}^T$. This means that the ratio of land will be 3 acres residential to 3 acres commercial to 2 acres industrial. To write this in terms of percentages, divide each component by 8 (the sum 3+3+2) to obtain 3/8 : 3/8 : 2/8 or 37.5% : 37.5% : 25%. This is the long term percentages of land use.

4 Rows and Columns

4.1 Row Space and Column Space and Rank

There are many connections between the rows and columns of a matrix. This section will illustrate how you can decompose a matrix A into the product of two matrices C and R which tell you how to reconstruct the original matrix A as linear combinations or columns of C or rows of R. We'll start with an example, and then describe the general process. Then we will add on some new vocabulary (span, dimension, and basis).

Consider the following matrix and its reduced row echelon form:

$$\underset{\text{Reduction}}{\begin{bmatrix}1&2&0\\2&1&3\\0&4&-4\end{bmatrix}\xrightarrow{\text{rref}}\begin{bmatrix}1&0&2\\0&1&-1\\0&0&0\end{bmatrix}}\qquad \underset{\text{Interpretation}}{2\begin{bmatrix}1\\2\\0\end{bmatrix}-1\begin{bmatrix}2\\1\\4\end{bmatrix}=\begin{bmatrix}0\\3\\-4\end{bmatrix}}$$

Because there is a column which is not a pivot column, we say that the columns of A are linearly dependent. The third column is not a pivot column, so we can use the number in the reduced row echelon form to write the third column as a linear combination of the first two columns as shown above. The reduced row echelon form tells us precisely how to linearly combine the pivot columns of our matrix to construct the original matrix.

Let C be the matrix whose columns are the pivot columns of A. Let R be the matrix whose rows are the nonzero rows of the reduced echelon form of A. Then we have $A = CR$. For the matrix in the previous example, we have

$$A=\begin{bmatrix}1&2&0\\2&1&3\\0&4&-4\end{bmatrix}\quad C=\begin{bmatrix}1&2\\2&1\\0&4\end{bmatrix}\quad R=\begin{bmatrix}1&0&2\\0&1&-1\end{bmatrix}\quad \begin{bmatrix}1&2\\2&1\\0&4\end{bmatrix}\begin{bmatrix}1&0&2\\0&1&-1\end{bmatrix}=\begin{bmatrix}1&2&0\\2&1&3\\0&4&-4\end{bmatrix}.$$

Some key observations to make are:

1. Every column of A is a linear combination of the pivot columns.

2. The columns of R are the coefficients needed to obtain the columns of A as linear combinations of columns of C. In other words, multiplying by R on the right of C gives us linear combinations of columns of C.

3. Every row of A is a linear combination of the rows of R.

4. The rows of C are the coefficients needed to obtain the rows of A as linear combinations of the rows of R. In other words, multiplying by C on the left of R gives us linear combinations of rows of R.

The following product illustrates the first 2 key observations with regards to columns. Notice how information from the columns of R is used to combine the columns of C.

$$\begin{aligned}\begin{bmatrix}1&2\\2&1\\0&4\end{bmatrix}\begin{bmatrix}1&0&2\\0&1&-1\end{bmatrix} &= \begin{bmatrix}\begin{bmatrix}1&2\\2&1\\0&4\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} & \begin{bmatrix}1&2\\2&1\\0&4\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix} & \begin{bmatrix}1&2\\2&1\\0&4\end{bmatrix}\begin{bmatrix}2\\-1\end{bmatrix}\end{bmatrix}\\ &= \begin{bmatrix}1\begin{bmatrix}1\\2\\0\end{bmatrix}+0\begin{bmatrix}2\\1\\4\end{bmatrix} & 0\begin{bmatrix}1\\2\\0\end{bmatrix}+1\begin{bmatrix}2\\1\\4\end{bmatrix} & 2\begin{bmatrix}1\\2\\0\end{bmatrix}-1\begin{bmatrix}2\\1\\4\end{bmatrix}\end{bmatrix}\end{aligned}$$

The following product illustrates the last two key observations with regard to rows. Notice how information from the rows of C is used to combine the rows of R.

$$\begin{bmatrix}1 & 2\\2 & 1\\0 & 4\end{bmatrix}\begin{bmatrix}1 & 0 & 2\\0 & 1 & -1\end{bmatrix} = \begin{bmatrix}\begin{bmatrix}1 & 2\end{bmatrix}\begin{bmatrix}1 & 0 & 2\\0 & 1 & -1\end{bmatrix}\\ \begin{bmatrix}2 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 2\\0 & 1 & -1\end{bmatrix}\\ \begin{bmatrix}0 & 4\end{bmatrix}\begin{bmatrix}1 & 0 & 2\\0 & 1 & -1\end{bmatrix}\end{bmatrix} = \begin{bmatrix}1\begin{bmatrix}1 & 0 & 2\end{bmatrix}+2\begin{bmatrix}0 & 1 & -1\end{bmatrix}\\ 2\begin{bmatrix}1 & 0 & 2\end{bmatrix}+1\begin{bmatrix}0 & 1 & -1\end{bmatrix}\\ 0\begin{bmatrix}1 & 0 & 2\end{bmatrix}+4\begin{bmatrix}0 & 1 & -1\end{bmatrix}\end{bmatrix}.$$

In summary, matrix multiplication is a way of creating new matrices as linear combinations of other matrices. It can be viewed either by looking at rows, or by looking at columns.

In the above example, the first two columns formed a linearly independent set, and using them we could obtain every other column of A. The set of all linear combinations of columns of A is called the column space of A. Recall that the span of a set of vectors is all possible linear combinations of those vectors. So the column space of A is the span of the columns of A. Since every non pivot column of A is a linear combination of A, we can say that the column space is the span of the pivot columns of A. This introduces an idea called a basis. A basis for a vector space is a set of linearly independent vectors whose span is the space we are interested in. So a basis for the column space is the set of pivot columns. The dimension of a vector space is the number of vectors in a basis. The dimension of our column space is 2.

In a similar fashion, we define the row space to be the span of the row vectors. If we reduce a matrix A to reduced row echelon form, then every row of A is a linear combination of the nonzero rows of A. This means that the row space is spanned by the nonzero rows of the reduced row echelon form. Those vectors are linearly independent and hence form a basis for the row space. The dimension of the row space is hence 2 as well.

In general, the rank of a matrix always equals the dimension of both the column space and row space. So we could define the rank of matrix to be the dimension of the column space, or the dimension of the row space, or the number of pivots, they are all the same. A basis for the column space is the pivot columns of A. A basis for the row space is the nonzero rows of the reduced row echelon form of A. Using these bases, we can reconstruct the matrix A as a product CR where C's columns are a basis for the column space, and R's rows are a basis for the row space. We will spend much more time with this idea in future units, as span, basis, and dimension are fundamental ideas in Linear Algebra.

The ideas above work with any size matrix (square or not). Here is a decomposition for a 5 by 7 matrix. The pivots are in columns 1, 2, 3, and 5. The rref form is the right most matrix (after dropping the row of zeros). The rank is 4.

$$\begin{bmatrix}2 & 1 & 0 & 3 & -3 & 3 & 2\\1 & 0 & 3 & 4 & 2 & 4 & 4\\0 & 4 & 0 & 4 & 0 & 0 & 4\\0 & 0 & 0 & 0 & 1 & -1 & 1\\0 & 1 & 3 & 4 & 0 & 6 & 1\end{bmatrix} = \begin{bmatrix}2 & 1 & 0 & -3\\1 & 0 & 3 & 2\\0 & 4 & 0 & 0\\0 & 0 & 0 & 1\\0 & 1 & 3 & 0\end{bmatrix}\begin{bmatrix}1 & 0 & 0 & 1 & 0 & 0 & 2\\0 & 1 & 0 & 1 & 0 & 0 & 1\\0 & 0 & 1 & 1 & 0 & 2 & 0\\0 & 0 & 0 & 0 & 1 & -1 & 1\end{bmatrix}$$

Using the above decomposition, I can write the 7th column as 2 times the first plus 1 times the second plus 1 times the 5th. I can write the 2nd row of the original matrix as the following linear combination of the reduced row echelon form

$$1\begin{bmatrix}1 & 0 & 0 & 1 & 0 & 0 & 2\end{bmatrix}+0\begin{bmatrix}0 & 1 & 0 & 1 & 0 & 0 & 1\end{bmatrix}+3\begin{bmatrix}0 & 0 & 1 & 1 & 0 & 2 & 0\end{bmatrix}+2\begin{bmatrix}0 & 0 & 0 & 0 & 1 & -1 & 1\end{bmatrix}.$$

4.2 Linear Regression

Our final application ask the following question: when a system has no solution, how do you find an approximate solution? In particular, if you have 3 or more points, how do you find a line that is "closest" to passing through these points? This line, called the least squares regression line, can be used to find trends in many branches of science. Statistics builds upon this idea to provide powerful tools for predicting the future.

Suppose we want to find a line that is closest to passing through the three points $(0,1),(2,3),(4,6)$. Since the points are not collinear, there is no such line. Suppose for a moment that there were a line of the form $y=mx+b$ that did pass through the points. Plugging our 3 points into the equation $mx+b=y$ gives the system of equations

$$\begin{cases} b=1 \\ 2m+b=3 \\ 4m+b=6 \end{cases} \xrightarrow{\text{augmented matrix}} \left[\begin{array}{cc|c} 0 & 1 & 1 \\ 2 & 1 & 3 \\ 4 & 1 & 6 \end{array}\right] \xrightarrow{\text{matrix form}} \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}.$$

Notice that the system can be written in matrix form $A\vec{x}=\vec{b}$ where A contains a column of x values and 1's and $\vec{b}$ is a column of y values. If you try to reduce this matrix, you will discover the system is inconsistent (has no solution). There are more equations than variables, so we call the system over determined.

Is there a way to reduce the number of rows in our system by using linear combinations of the rows we have, so that the resulting system has only 2 rows? If we multiply on the left by a 2 by 3 matrix, we would obtain a system with 2 rows instead of 3, and the rows of the new matrix would be linear combinations of the rows of our original matrix. The only 2 by 3 matrix in this problem is the transpose of A. So lets multiply both sides by the transpose of A:

$$A=\begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix}, A^T=\begin{bmatrix} 0 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix}, A^TA=\begin{bmatrix} 0 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix}=\begin{bmatrix} 20 & 6 \\ 6 & 3 \end{bmatrix}, A^T\vec{b}=\begin{bmatrix} 0 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}=\begin{bmatrix} 30 \\ 10 \end{bmatrix}.$$

The equation $A\vec{x}=\vec{b}$ becomes the equation $A^TA\vec{x}=A^T\vec{b}$, or $\begin{bmatrix} 20 & 6 \\ 6 & 3 \end{bmatrix}\begin{bmatrix} m \\ b \end{bmatrix}=\begin{bmatrix} 30 \\ 10 \end{bmatrix}$. The solution can then be obtained by reducing $\left[\begin{array}{cc|c} 20 & 6 & 30 \\ 6 & 3 & 10 \end{array}\right]$ to $\left[\begin{array}{cc|c} 1 & 0 & 5/4 \\ 0 & 1 & 5/6 \end{array}\right]$, which means the solution is $y=\frac{5}{4}x+\frac{5}{6}$. Later we will prove that A^TA has an inverse, which means we can write $\vec{x}=(A^TA)^{-1}A^T\vec{b}=\begin{bmatrix} 1/8 & -1/4 \\ -1/4 & 5/6 \end{bmatrix}\begin{bmatrix} 30 \\ 10 \end{bmatrix}=\begin{bmatrix} 5/4 \\ 5/6 \end{bmatrix}$ as the solution to our system.

In general, a least squares regression problem is solved by (1) assuming the form of a solution ($y=mx+b$), (2) putting your values of x and y into the system to get a matrix equation $A\vec{x}=\vec{b}$, (3) multiply both sides by A^T, and (4) solving the simplified system (using reduction, inverse, or Cramer's rule - the next section).

4.3 Cramer's Rule

Cramer's rule is a theoretical tool which gives the solution to any linear system $A\vec{x}=\vec{b}$ with n equations and n unknowns, provided that there is a unique solution. Let $D=\det(A)$. Let D_i be the determinant of the matrx formed by replacing the ith column of A with $\vec{b}$. Then Cramer's rule states that $x_1=\frac{D_1}{D}, x_2=\frac{D_2}{D},\ldots,x_n=\frac{D_n}{D}$. We will prove it in class with pictures which connect determinants to area. This method of solving a system of equations is doable for 2 by 2 and 3 by 3 systems, but quickly becomes computationally inefficient beyond (as computing determinants is ugly on large matrices). For large systems, it is much faster to use Gaussian elimination. However, Cramer's rule is a very powerful theoretical idea, and can simplify theoretical computations. We'll look at an example with numbers, and then use it in two theoretical instances.

Let's solve $\begin{bmatrix} 1 & 2 & 0 \\ -2 & 0 & 1 \\ 0 & 3 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}=\begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$ using Cramer's rule. We compute

$$D=\begin{vmatrix} 1 & 2 & 0 \\ -2 & 0 & 1 \\ 0 & 3 & -2 \end{vmatrix} - 11, D_1=\begin{vmatrix} 2 & 2 & 0 \\ -2 & 0 & 1 \\ 1 & 3 & -2 \end{vmatrix}=-12, D_2=\begin{vmatrix} 1 & 2 & 0 \\ -2 & -2 & 1 \\ 0 & 1 & -2 \end{vmatrix}=-5, D_3=\begin{vmatrix} 1 & 2 & 2 \\ -2 & 0 & -2 \\ 0 & 3 & 1 \end{vmatrix}=-2$$

By Cramer's Rule we have, $x_1=12/11, x_2=5/11, x_3=2/11$. Using Gaussian Elimination, we obtain the same solution

$$\left[\begin{array}{ccc|c} 1 & 2 & 0 & 2 \\ -2 & 0 & 1 & -2 \\ 0 & 3 & -2 & 1 \end{array}\right] \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & 12/11 \\ 0 & 1 & 0 & 5/11 \\ 0 & 0 & 1 & 2/11 \end{bmatrix}.$$

To find the inverse of a 2 by 2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we need to solve $\left[\begin{array}{cc|c} a & b & 1 \\ c & d & 0 \end{array}\right]$ and $\left[\begin{array}{cc|c} a & b & 0 \\ c & d & 1 \end{array}\right]$. Cramer's rule gives us the formula

$$A^{-1} = \begin{bmatrix} \begin{vmatrix} 1 & b \\ 0 & d \end{vmatrix} / |A| & \begin{vmatrix} 0 & b \\ 1 & d \end{vmatrix} / |A| \\ \\ \begin{vmatrix} a & 1 \\ c & 0 \end{vmatrix} / |A| & \begin{vmatrix} a & 0 \\ c & 1 \end{vmatrix} / |A| \end{bmatrix} = \frac{1}{|A|} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

This provides a quick way to compute determinants of 2 by 2 matrices. Interchange the diagonal entries, change the sign on the others, and divide by the determinant.

To solve the least square regression problem, we need to solve the problem $A^T A\vec{x} = A^T b$. We write the matrices on both sides as

$$A^T A = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} = \begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{bmatrix} \qquad A^T b = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix}.$$

Cramer's rule then says that our solution is

$$m = \frac{\begin{bmatrix} \sum x_i y_i & \sum x_i \\ \sum y_i & n \end{bmatrix}}{\begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{bmatrix}} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2} \qquad b = \frac{\begin{bmatrix} \sum x_i^2 & \sum x_i y_i \\ \sum x_i & \sum y_i \end{bmatrix}}{\begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{bmatrix}} = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2}.$$

At this point a computer program or spread sheet can quickly compute the coefficients without needing Gaussian elimination. Finding this solution without Cramer's rule is more difficult.