1 Additional Problem

1.1 Problem Statement

Suppose you are a company making industrial chemicals. For each unit of chemicals produced, you choose to release pollutant into a nearby system of lakes instead of cleaning up the pollutant. However, there is a regulatory agency making sure that toxins in the lakes do not go above a certain level. They publish the times at which they measure the lakes in advance. Suppose you know when the measurements will take place. How should you maximize pollution, and thus production and profit, across the various lakes while still making sure that the inspectors never measure more than the legal limit at any one time?

1.2 State Variables

Suppose that the lakes are small enough that it takes significantly longer for pollutant to diffuse accross the connections between lakes than it does for the concentrations in the lake to equilibrate. This assumption has the consequence that it doesn't matter where the inspectors measure the concentration of pollutant: they will measure approximately the same value everywhere. The state variables are therefore C_i, the concentration of pollutant in lake i.

1.3 Basic Dynamics

How do the state variables evolve in time?

I make the assumption that the absolute concentrations don't affect the dynamics. The dynamics will only be affected by the difference in concentration between each lake. At first, I don't assume that the pollutant breaks down, but that it stays constant. Therefore, the dynamics of each lake depend only on the connections it has with each neighboring lake.

If the concentration in one lake is higher than the other, then I will expect a linear differential equation to relate them:

$$\frac{dC_i}{dt}_{\text{from } C_j} = r_{ij}\left(C_j - C_i\right)$$

where r_{ij} is a rate of diffusion between lakes i and j. This rate is positive, and in most circumstances, $r_{ij} = r_{ji}$. We won't start assuming that unless we have to, however.

We will need to measure or estimate these rates of diffusion.

Let R be the matrix of rate coefficients, and let C be a vector of concentrations. The matrix of rate coefficients has r_{ij} at the i, j-th position, and the negative sum of all of the row's entries on the diagonal entry. With this layout, equations read:

$$\frac{dC}{dt} = RC$$

We can solve this by diagonalizing R and getting a system of coupled linear differential equations. If λ_k is the k-th eigenvalue with corresponding eigenvector v_k, then the dynamics of an initial concentration $A_k(0)v_k$ are given by:

$$\frac{dA_k(t)}{dt} = A_k(0)e^{\lambda_k t}$$

1.4 Pollutant Breakdown Rates

Suppose the pollutant decomposes at a fixed rate h. The dynamics of pollutant in an isolated lake C_1 would thus be given by:

$$\frac{dC_1}{dt} = -hC_1$$

We can modify the R matrix above by subtracting a decay term. If h_i is the decay rate for the i-th lake, then to add the affect of decay, one should subtract h_i from each i-th diagonal entry.

1.5 Polluting

Now i'm going to add pollution to the model, the thing we're trying to minimize. Suppose that $P(t)$ is a vector which represents the rate of pollution in each lake. The dynamics of the pollution in the lake system now read, in the notation of the previous subsections:

$$\frac{dC}{dt} = RC + P(t)$$

1.6 Equlibrium Concentrations

Immediately, we see an expression for equilibrium concentrations. Assuming that P is constant, we can see that $P = -RC$. Therefore, to solve for C at equilibrium given P, we get:

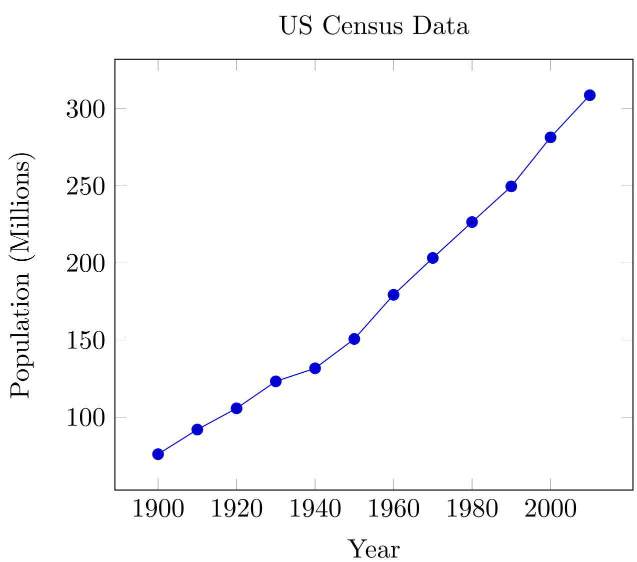
$$C = (-R)^{-1}P$$

Of course, all of the concentrations have to be positive. Therefore, it's not straightforward that there is an equilibrium solution, depending only on the pollutant levels and not initial conditions. There should always be one, assuming that the polutants decay proportional to their concentration and are only replenished in fixed quantitites. However, if the pollutants do not decay, then there should be no equilibrium.

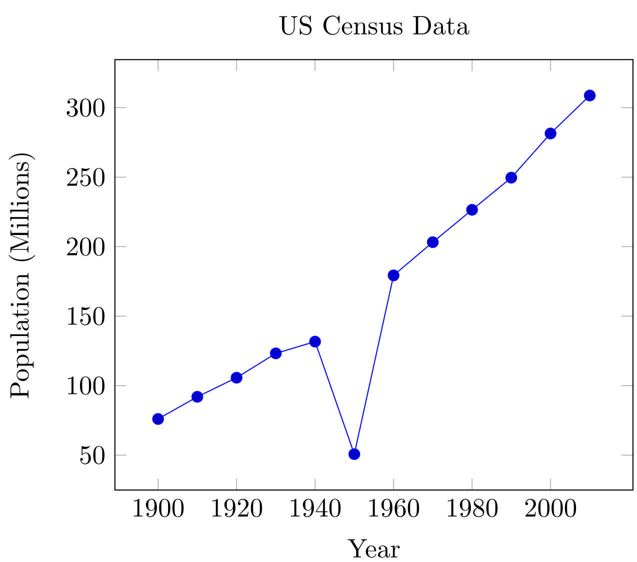
If there is an expression for equilibrium concentration, and if there are threshold concentrations for each lake, then the problem of what constant level of pollution to emit is transformed into a linear programming problem. Suppose the concentrations in each lake must stay below a vector C_m. Further suppose that each plant has a maximum operating capacity P_m. The goal is to maximize the total sum of P while keeping P positive and within the constraints, $(-R)^{-1}C \leq C_m$ and $P \leq P_m$.

2 Outliers

The total population of the US, as determined extremely precisely by the US census, is given below:



In this problem, we modify the population in 1950 from 150.697 million to 50.697 million, as if we forgot to type in the one when entering the data. It now looks like this:



We try to answer the question: Which of the models are the most and least affected by this outlier? A model which handles the outlier well might ignore it and try to fit the rest of the data, or it might alert the user that something might be wrong with the data. Alerting the user makes good sense in this case, especially since the "story" behind this outlier is from data entry error. Models which handle this kind of an outlier poorly are ones for which the presence of an outlier significantly changes their fit, and makes their lines leave the data and try to find some middle ground. This is not the desired behavior.

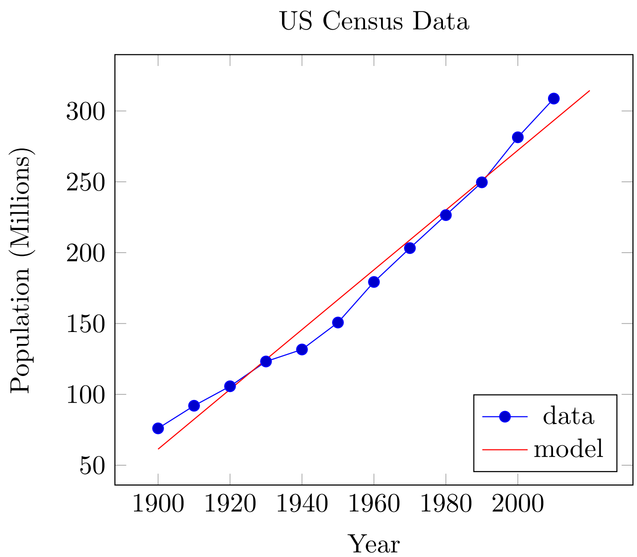
The models we are going to evaluate are:

1. Polynomial of degrees 1 to 4
2. PCHIP
3. Spline
4. Exponential

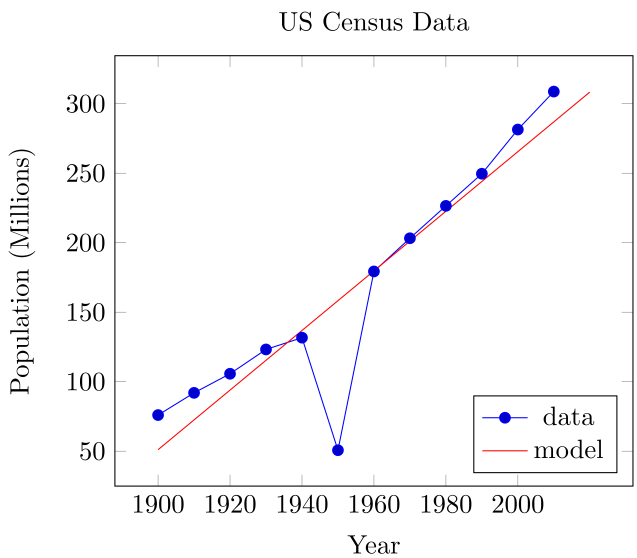
2.1 Polynomial Models

2.1.1 Linear Models

We can try to fit a linear model with least squares. First, to the original data:



It's clear from the graph that the linear model isn't a good fit. Next, to the outlier data:

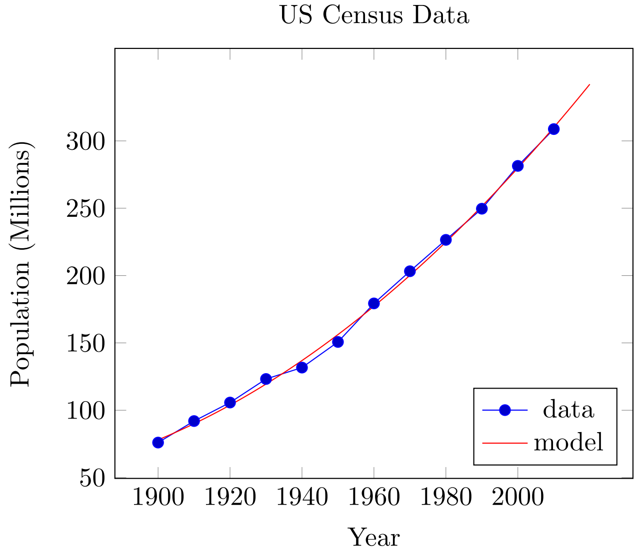


The model appears relatively resilient to the outlier.

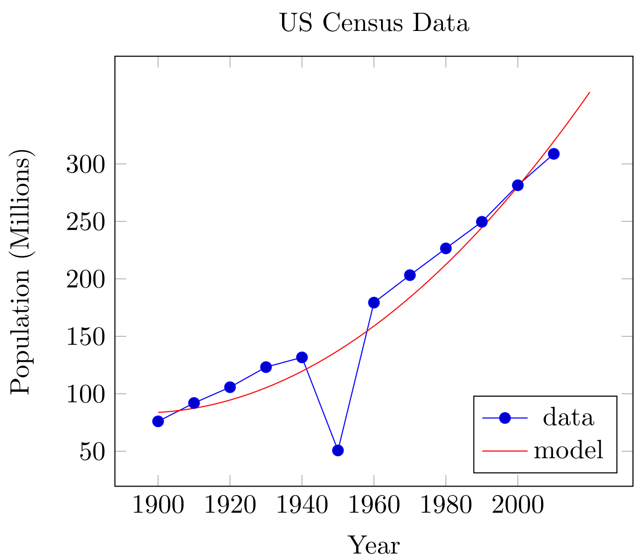
The model moves slightly downwards, but the effect isn't extremely significant, and the outlier decreases the model's prediction for 2020 by only a few million people.

2.1.2 Quadratic Models

We can try to fit a quadratic model with least squares. First, to the original data:



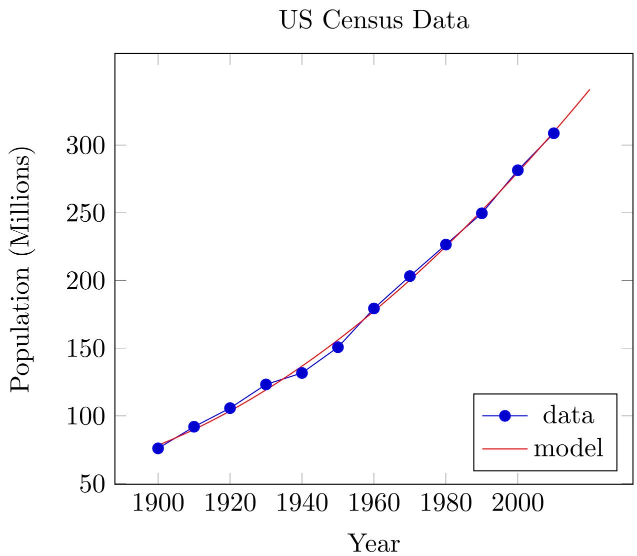
The model fits the data quite well, Next, to the outlier data:



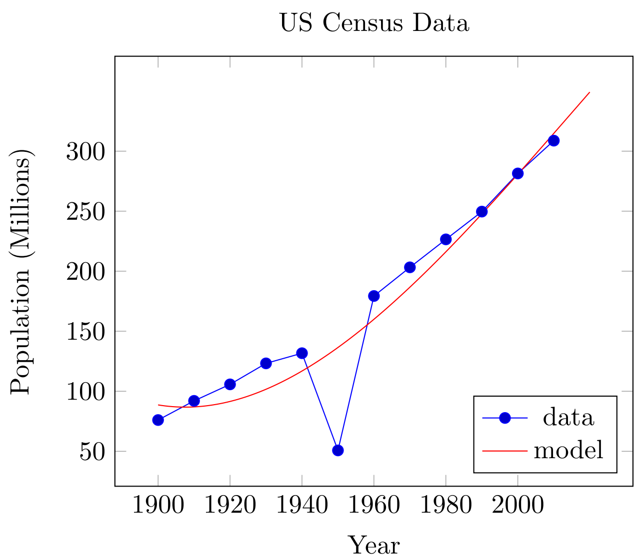
The outlier seems to have a large effect. The model is off the data for most of the way, and its predictions don't seem to make sense anymore. The quadratic model seems to be somewhat sensitive to outliers.

2.1.3 Cubic Models

We can try to fit a cubic model with least squares. First, to the original data fit to a cubic model:



Note a very good correspondence with the data. Next, to the outlier data:



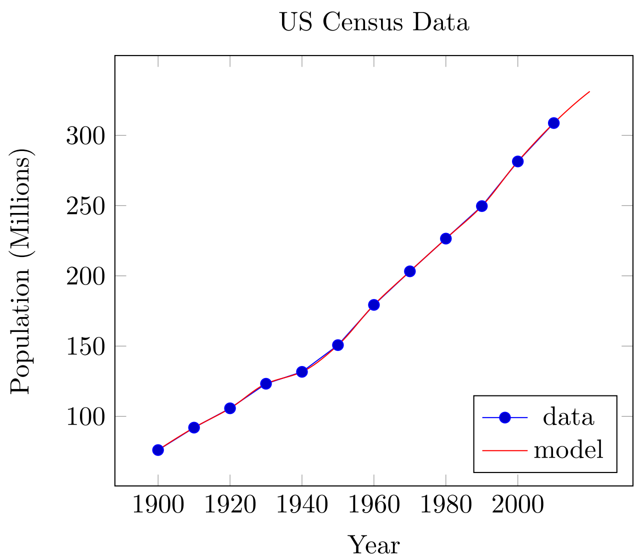
The model doesn't fit the data well here. However, it's interesting to note that the model's predictions seem to be fairly accurate, and aren't far off from the original dataset.

2.1.4 Conclusion

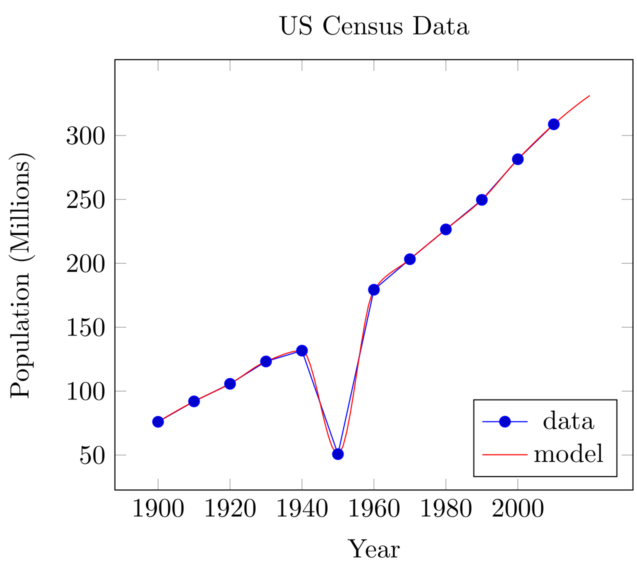
Polynomial Models seem to have different behavior as the degree of the polynomials increases. The higher degree the polynomials go, the worse the behavior near outliers will be. However, the higher degree the polynomials go, the less the outliers affect the predictions of the dataset for the future.

2.2 PCHIP

PCHIP stands for Piecewise Cubic Hermite Interpolating Polynomial. This method of interpolation attempts to make a piecewise polynomial which interpolates each data point. Below is the PCHIP for the data with no outlier:



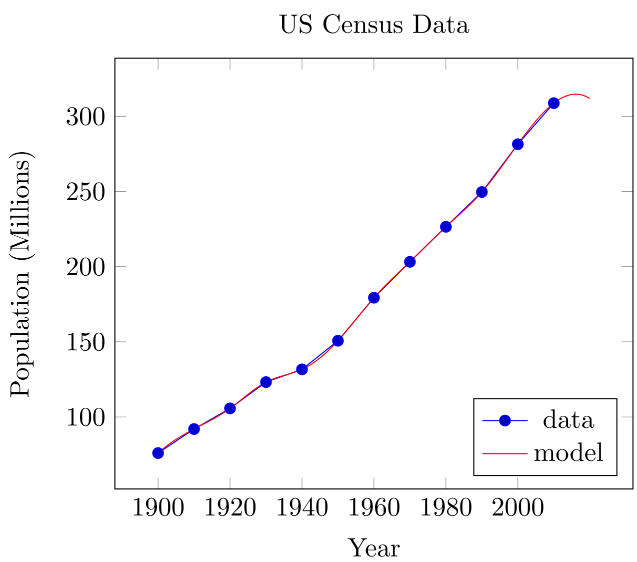
And the PCHIP for the data with the outlier:



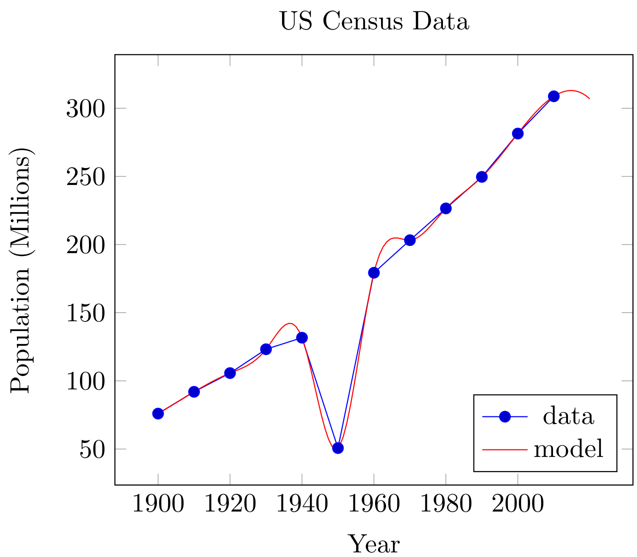
The outlier hardle affects the prediction for 2020 at all. However, the model happily accomodates the data point, moving down towards it. This behavior is bad because the method doesn't recognize the data point as an outlier.

2.3 Spline

Cubic Splines are another interpolating method, similar to PCHIP.



We see that the splines closely track the data, just as before.

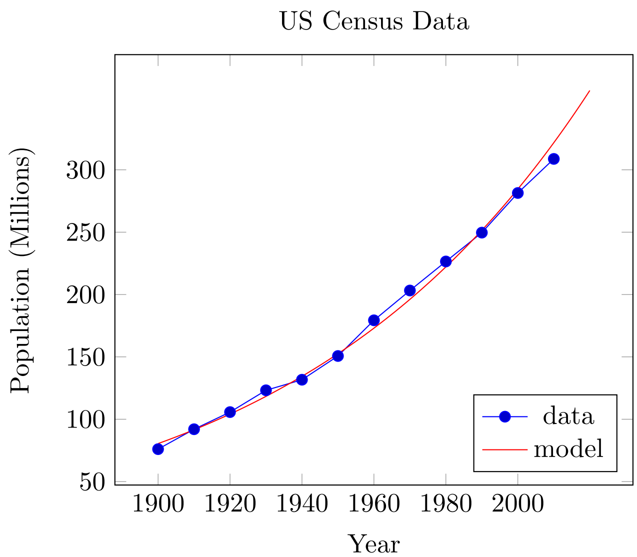


When the outlier is introduced, however, the splines have some “overshoot” near the outlier. The region in which the outlier distorts the model, compared to the true data, is much larger.

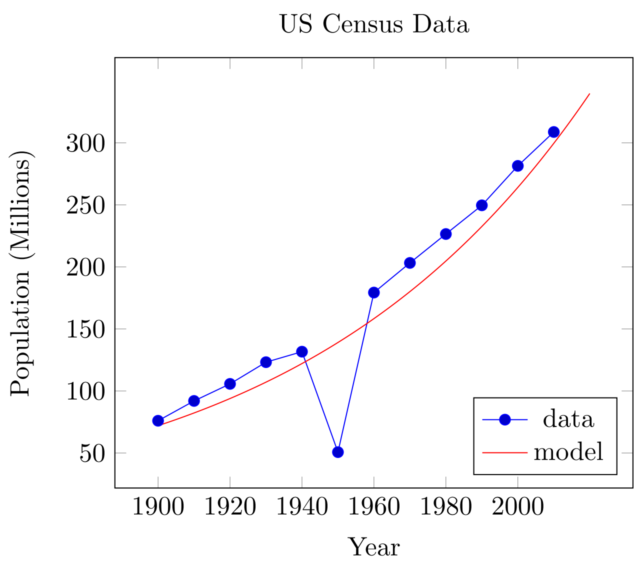
The predictions are the same.

2.4 Exponential Models

We can make an exponential model by taking the log of both axes and then fitting with a linear model. We might expect the linear model to be tolerant to outliers as before in “log-space”, so I expect this model to be relatively tolerant.



This model actually doesn't fit the data as well as I anticipated. I expected the exponential growth law to fit well, because it was the first part of a population situation. However, the model seems to be diverging from the data, especially towards the end.



The outlier pulls the model out significantly.

2.5 Conclusion

In summary, polynomial models are more resilient to outliers the lower degree they are, although higher degree polynomial models seem to be able to recover from outliers early in the data set to preserve their predictive value. PCHIP and Spline models try to fit the data exactly, since they're interpolating models. The PCHIP does this better, because it minimizes the overshoot when coming back to the normal data that we saw with the spline model. Exponential Models seem to be easily distorted by outliers.

3 Rabbits and Foxes

3.1 Discussion of the model

We have already justified the logistic model for a single species of animal. How can we judge interactions between two species?

3.1.1 Rabbit Dynamics

Consider two coexisting populations of foxes and rabbits. In the absence of foxes, the rabbits follow a logistic model for growth, with carrying capacity K and growth rate A. Once the foxes are introduced, predation occurs at a rate proportional to both the number of foxes and the number of rabbits. If R is the population of rabbits and F is the population of foxes, the differential equation describing rabbit population should be as follows, with B a proportionality constant:

$$\frac{dR}{dt} = AR\left(1 - \frac{R}{K}\right) - BRF \tag{1}$$

3.1.2 Foxes in Isolation

Since we have introduced a new variable F, representing the number of foxes, we now need to determine its dynamics. Suppose that in the absense of rabbits, the foxes die of starvation and other natural causes at a rate C, and proportional to their population. This means that in the absence of rabbits, we would expect the foxes to die off exponentially.

This model is very simple and somewhat flawed, because past a certain time without food, we wouldn't expect any foxes to be alive, but the exponential model never decreases exactly to zero. Therefore this model won't be predictive when there are very few foxes.

3.1.3 Fox Predation

Suppose that from predating the rabbits, the foxes gain in population at a rate proportional to their population and the rabbits' population. This can be interpreted as the equilibrium limit of the independent chances of any given fox catching any other rabbit. This model obviously ignores many factors, but it's the simplest which captures the essence of predation.

The if D is the predation proportionality constant, then the differential equation encoding the dynamics of the fox population is as follows:

$$\frac{dF}{dt} = (DR - C) F \tag{2}$$

3.2 Equilibria

The full system of coupled differential equations is as follows:

$$\frac{dR}{dt} = AR\left(1 - \frac{R}{K}\right) - BRF \tag{3}$$

$$\frac{dF}{dt} = (DR - C) F \tag{4}$$

If the system is in equlibrium, then $\frac{dF}{dt} = 0$. Thus, either $R = \frac{C}{D}$ or $F = 0$.

If the system is in equilibrium, then $\frac{dR}{dt} = 0$. Suppose $F = 0$. Then, $AR\left(1 - \frac{R}{K}\right) = 0$, and thus either $R = 0$ or $R = K$.

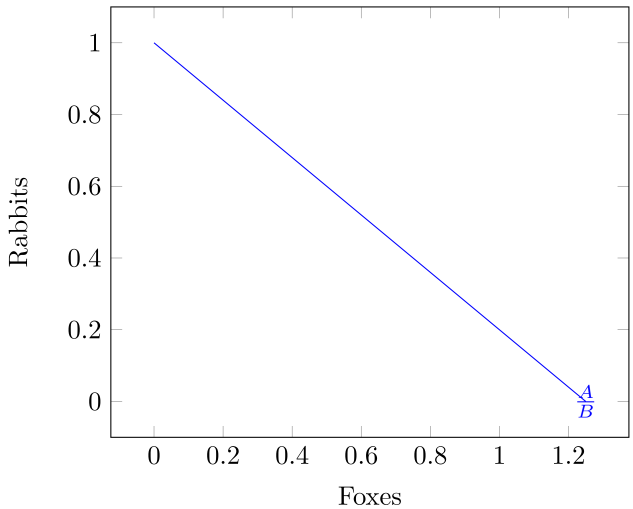
Suppose $R = \frac{C}{D}$. Then, $A\frac{C}{D}\left(1 - \frac{C}{D}\frac{1}{K}\right) - B\frac{C}{D}F = 0$. Therefore, $F = \frac{A}{B}\left(1 - \frac{C}{D}\frac{1}{K}\right)$.

We summarize all of the equilibrium points in a table.

R	F
0	0
K	0
$\frac{C}{D}$	$\frac{A}{B}\left(1 - \frac{C}{D}\frac{1}{K}\right)$

To help visualize the stable point for which the populations of both rabbits and foxes are nonzero, I plotted the following diagram. The position of the stable point is determined by the position of $\frac{C}{D}$ on the foxes axis, where $K = 1$. We see immediately that this stable point isn't even in the state space when $\frac{C}{D} > K$.

Possible Locations of Stable Point



3.2.1 Which ones are additional?

We are treating this model as a modification to a model in which the rabbits are described by an exponential growth model in the absence of foxes. This corresponds to sending the carrying capacity to infinity. Note that if the carrying capacity goes to infinity, then the second equilibrium point disappears, and the $\frac{1}{K}$ in the third one becomes zero. The term with $\frac{1}{K}$ represents the modification to this model from the finite carrying capacity of the rabbits.

3.3 Stability of Equilibria

To determine whether an equilibrium point is stable or unstable, we can treat the dynamics around the equilibrium point as a linear system, even if the overall dynamics are nonlinear. This involves an object called the Jacobian. The jacobian for this system of differential equations is:

$$J(R,F) = \begin{bmatrix} \frac{\partial\left(\frac{dR}{dt}\right)}{\partial R} & \frac{\partial\left(\frac{dF}{dt}\right)}{\partial R} \\ \frac{\partial\left(\frac{dR}{dt}\right)}{\partial F} & \frac{\partial\left(\frac{dF}{dt}\right)}{\partial F} \end{bmatrix} = \begin{bmatrix} (A-BF) - 2A\frac{R}{K} & -BR \\ DF & DR - C \end{bmatrix} \tag{5}$$

The stability in the neighborhood of each equilibrium point is given by the eigenvalues of each jacobian, evaluated at that point.

R	F	λ_1	λ_2
0	0	A	$-C$
K	0	$-A$	$D\left(K - \frac{C}{D}\right)$
$\frac{C}{D}$	$\frac{A}{B}\left(1 - \frac{C}{D}\frac{1}{K}\right)$	$\frac{A}{2}\frac{C}{D}\left(-\frac{1}{K} + \sqrt{\frac{1}{K^2} + \frac{4B}{A}\left(\frac{1}{K} - \frac{D}{C}\right)}\right)$	$\frac{A}{2}\frac{C}{D}\left(-\frac{1}{K} - \sqrt{\frac{1}{K^2} + \frac{4B}{A}\left(\frac{1}{K} - \frac{D}{C}\right)}\right)$

3.3.1 Stability of points on the "Rabbit Axis"

Zero will always be a saddle point, attractive from the foxes side, because foxes die with no rabbits, and repelling from the rabbits side, because without foxes rabbits grow logistically. The rabbits' carrying capacity is sometimes stable, and sometimes unstable. If $K > \frac{C}{D}$, then both eigenvalues are negative, and it is attracting from all sides. However, if $K < \frac{C}{D}$, the rabbits' carrying capacity becomes a saddle, and it repels in the direction of more foxes.

3.3.2 Stability of center point

The center point is algebraically more tricky to deal with. We can notice the $\frac{1}{K^2}$ term in the square root. If the other term were zero, than in λ_1, the value of the square root would exactly cancel out the leading $-\frac{1}{K}$, and the eigenvalue would be 0. Therefore, if the other term within the square root is greater than zero, then we would expect λ_1 to be positive, and if it is less than zero, we would expect λ_1 to be negative. The term in question is $\frac{4B}{A}\left(\frac{1}{K} - \frac{D}{C}\right)$.

If $K > \frac{C}{D}$, then the term is positive, and λ_1 is positive, and therefore repelling. This might be significant, but if we plug in the position for the stable point itself, we find that it is in fact not in the positive quadrant, and we can remove it from consideration.

The other eigenvalue is still negative, and therefore, if $K > \frac{C}{D}$, the internal stable point is a saddle. If, however, $K < \frac{C}{D}$, then the term is negative, and the total value of λ_1 is thus negative. Then the internal stable point is attractive.

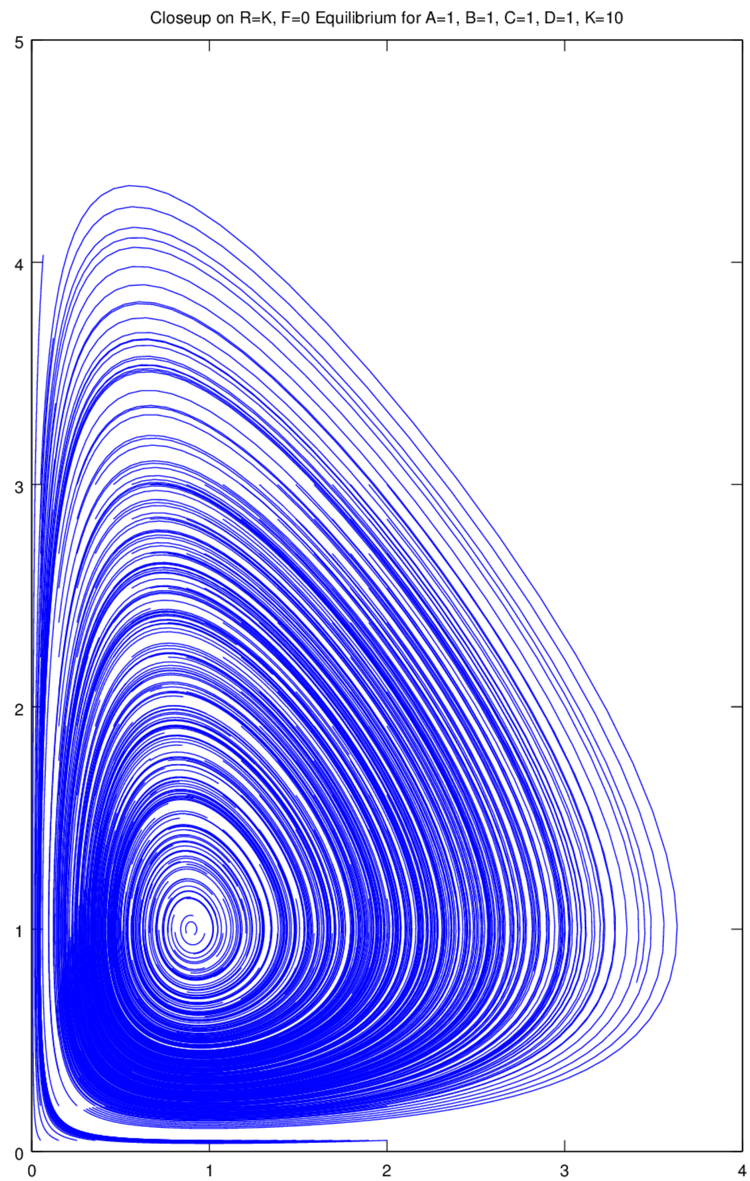
Note that if $K = \frac{C}{D}$, then the two nonzero fixed points are actually on top of each other.

3.3.3 Conclusion

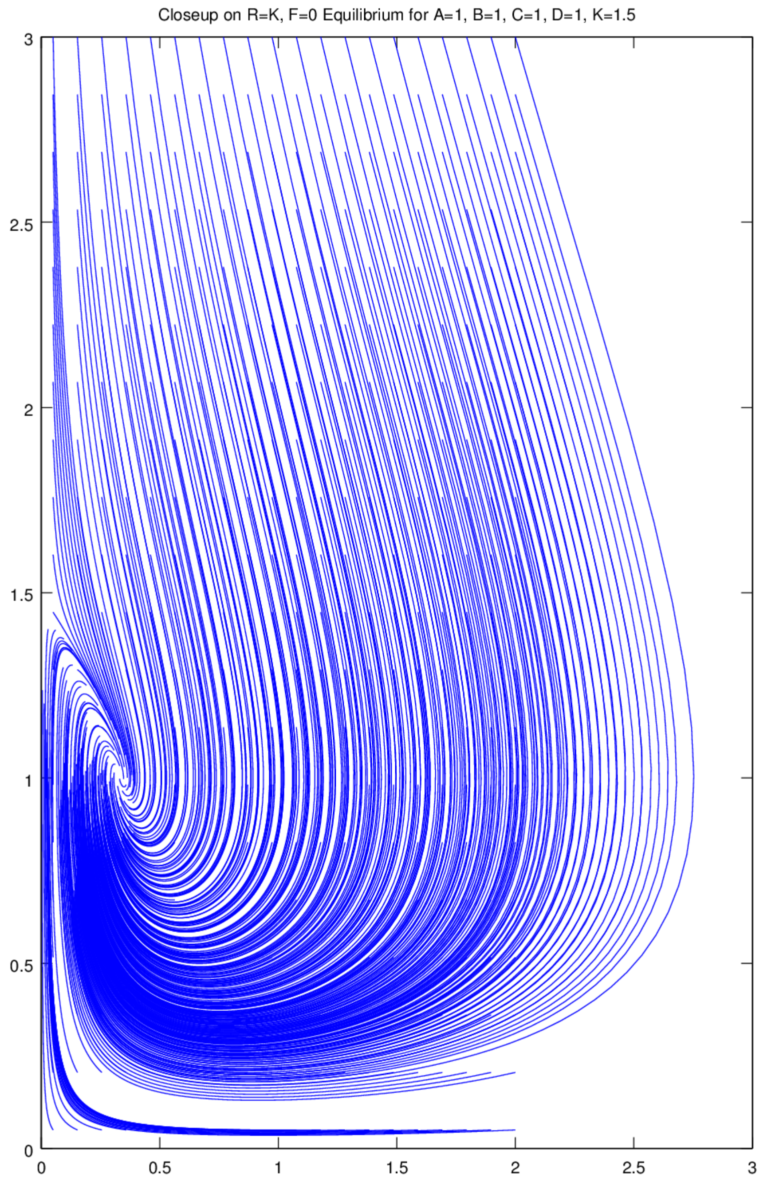
To sum up, when $K > \frac{C}{D}$, the foxes die out and the rabbits reach their carrying capacity. However, if $K < \frac{C}{D}$, then the foxes survive and the rabbit-fox population reaches an equilibrium with $\frac{C}{D}$ rabbits and a nonzero amount of foxes. For a predator species, the key factor for survival is the ratio between how much they can get from predation, D, and the rate at which they starve, C. If they can push the ratio $\frac{C}{D}$ above the carrying capacity of a prey species, then they can safely predate them. However, if they starve too fast, or derive too little nutrients from the prey, then they will die off.

3.4 Phase Portraits

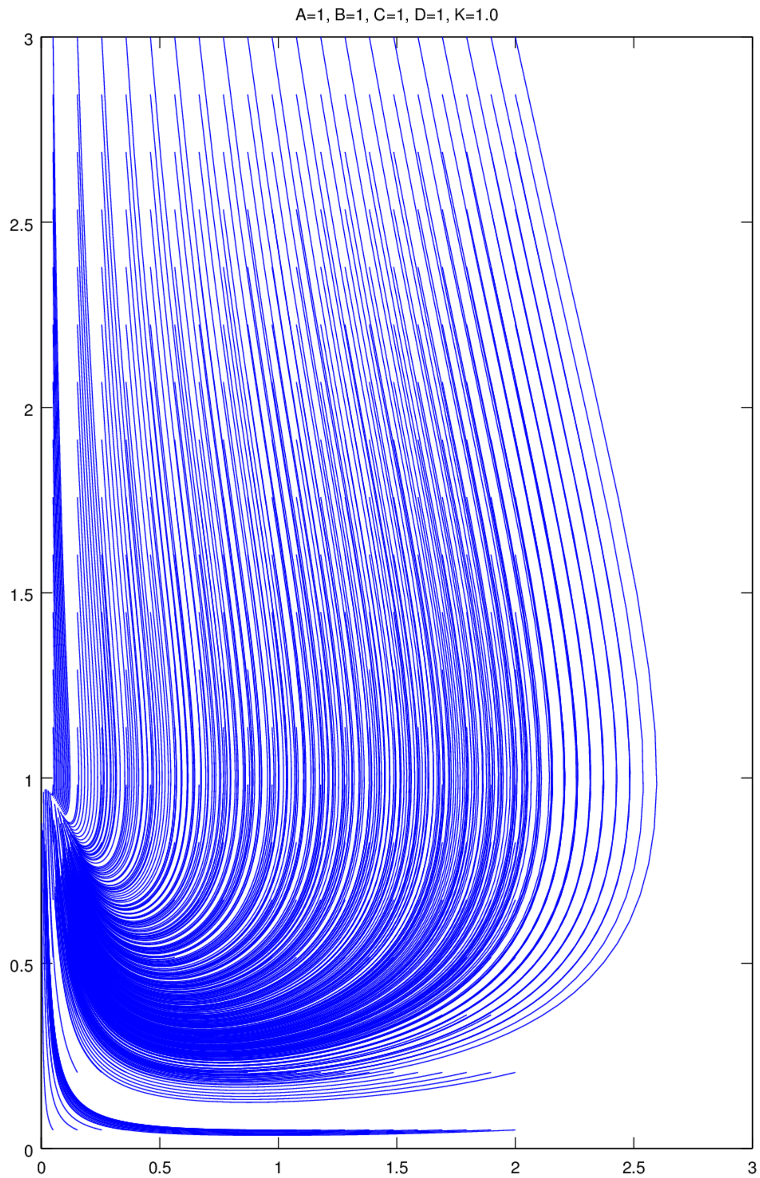
This is the case where the carrying capacity K is much greater than $\frac{C}{D}$. The population periodically oscillates around a central point, which is an attracting center.



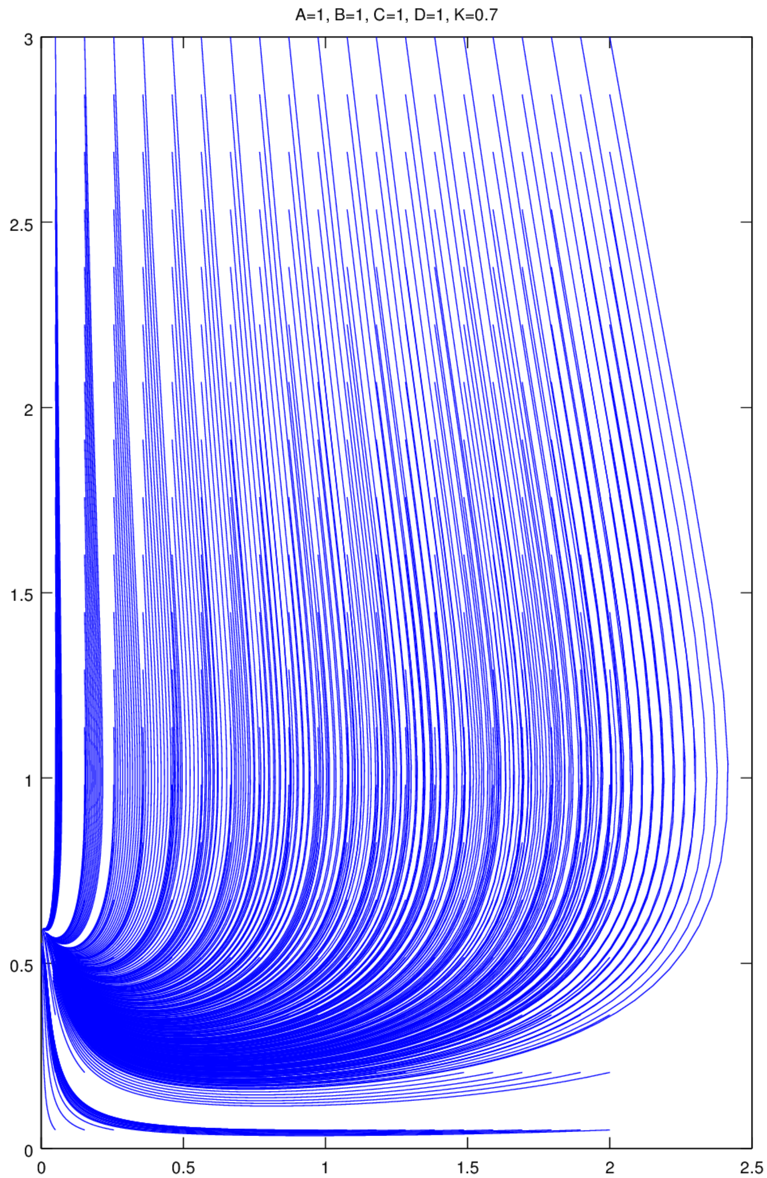
As we bring K down to just above $\frac{C}{D}$, the cycle stretches as it gets near K on the rabbits axis.



When $\frac{C}{D} = K$, the total field is attracted slowly.



The whole state space is attracted more quickly to the only-rabbits state.


A=1, B=1, C=1, D=1, K=0.7
3
2.5
2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5

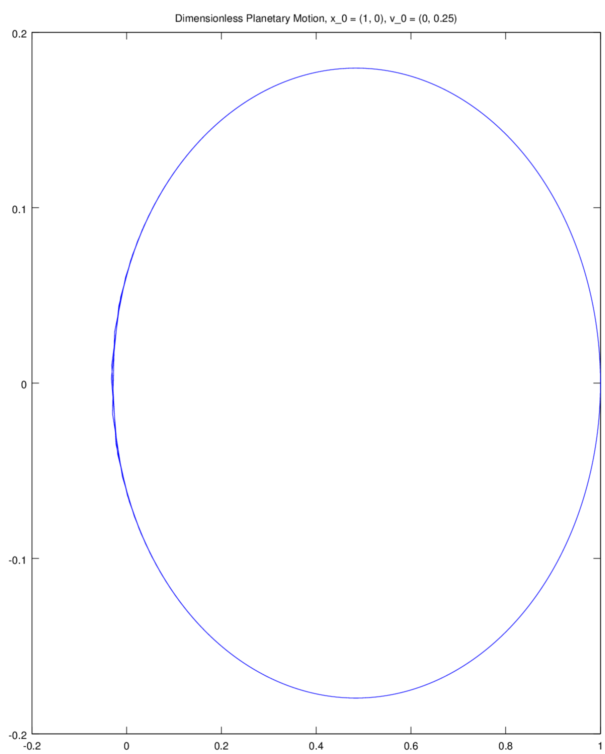
4 ODE Solvers

For this exercise, I'm integrating an ODE forwards in time. The ODE models planetary motion, and is given below:

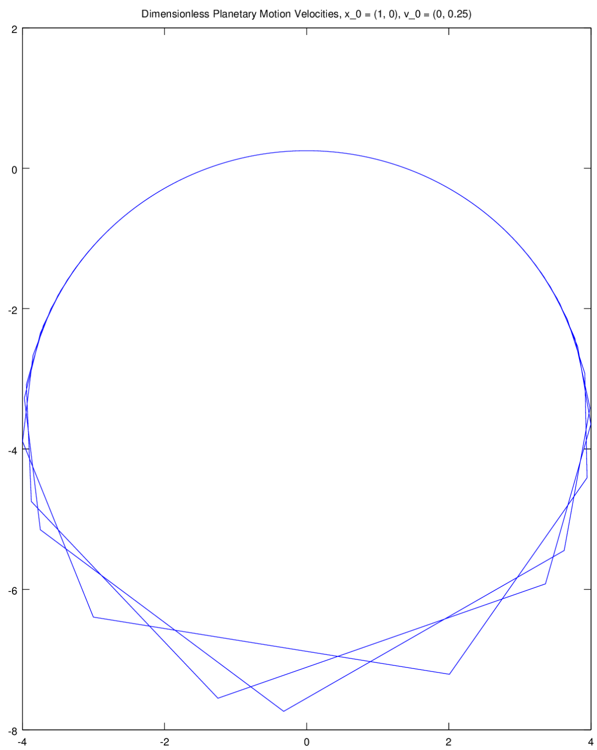
$$\frac{dx}{dt} = -\frac{x}{\left(x^2+y^2\right)^3} \tag{6}$$

$$\frac{dy}{dt} = -\frac{y}{\left(x^2+y^2\right)^3} \tag{7}$$

When I integrate this from time 0 to 2π, I get the following plot of position-space:



And velocity-space:



I would expect the bottom plot to follow a smooth trajectory, but it is jagged. This may be because near the origin, the derivatives in the model become very large. Since these initial conditions send the planet very close to the origin, we see the individual steps that the ODE solver takes in the plots of velocity.