Problem 1

I adjusted the control term following the logic presented by Moon et al in the supplemental material to "Genetic programs constructed from layered logic gates in single cells". I decided that there were three possible states-gene, open gene, and open gene with polymerase. I assumed that transcription could only occur when the open gene and polymerase were present. Therefore, I wrote a control term as

$$u_j = \frac{k_2 f_{TL}}{1 + k_2 f_{TL} + k_3 f_T} \tag{1}$$

where f_{TL} gives the fraction of the time where the polymerase is bound to the open gene, and f_T is $1 - f_{TL}$. The ks represent weights. Using this control term, and the formulation given in lecture for transcription, I produced a graph showing transcription rates for the two cases, Figure

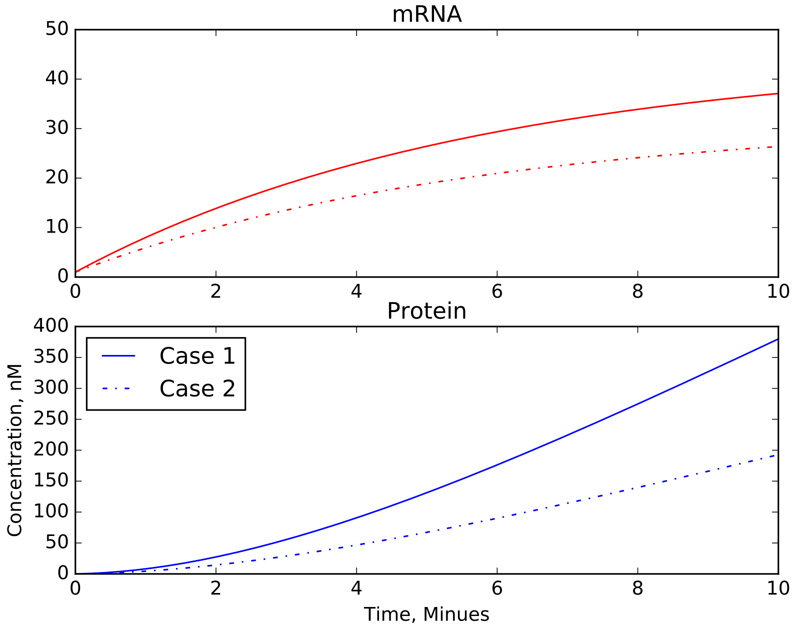

Figure 1: Amounts of mRNA and protein produced with modified control term. To differentiate the two cases, when the gene and RNAP are incubated together (case 1), I made k_2 large, to heavily weight the state of bound open gene complex to promoter and in the case where no gene was present (case 2), I made k_3 large to account for a large population of the unbound state.

In my formulation, since gene length is not accounted for in the control term, it appears in the denominator of the expression for the transcription rate.

$$r_{Tj} = \bar{r_{Tj}} u_j, \bar{r_{Tj}} = k_T R_T \frac{L_T}{L_{Tj}} \frac{G_j}{K_{Tj} + G_j} \tag{2}$$

So, as gene length increases, the rate of transcription decreases, as does the rate of translation due to the longer mRNA produced by longer genes.

Problem 2

At steady state, the protein levels are largely controlled by the weights in the control terms, as shown in Figure 2. When the system is induced, the rate constants for transcription and translation as well as the rnapII concentration and ribosome concentration play a large role in the system response, as shown in Figure 3.

As expected, when only the concentrations of P3 and mRNA are measured, the number of measurable species is fewer than in the case when all species are measured, as shown in Figure 4. Additionally, as measuring frequency decreases (time between measurements increases), so does the number of estimable parameters. The details of which parameters are estimable at $\epsilon = .001$ are given in the supplementary material.

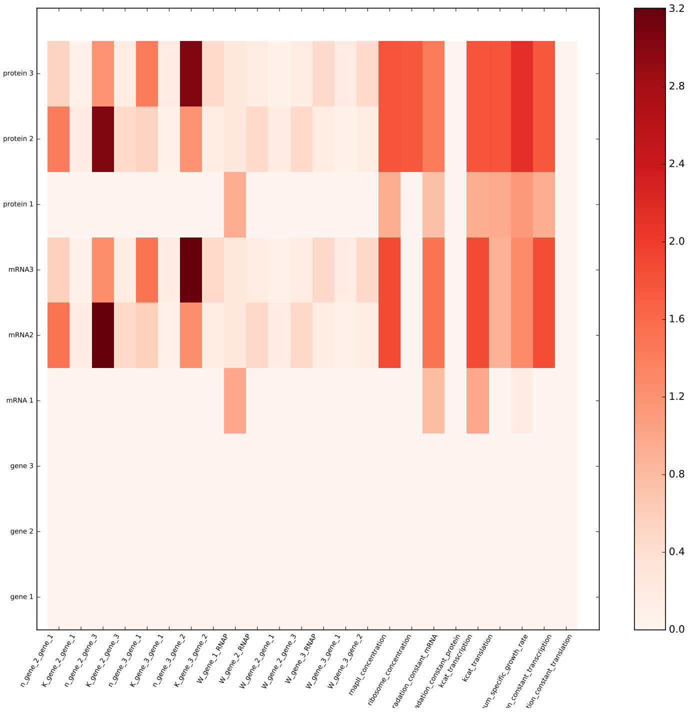

Figure 2: Heat map produced from singular value decomposition of three gene network at steady state

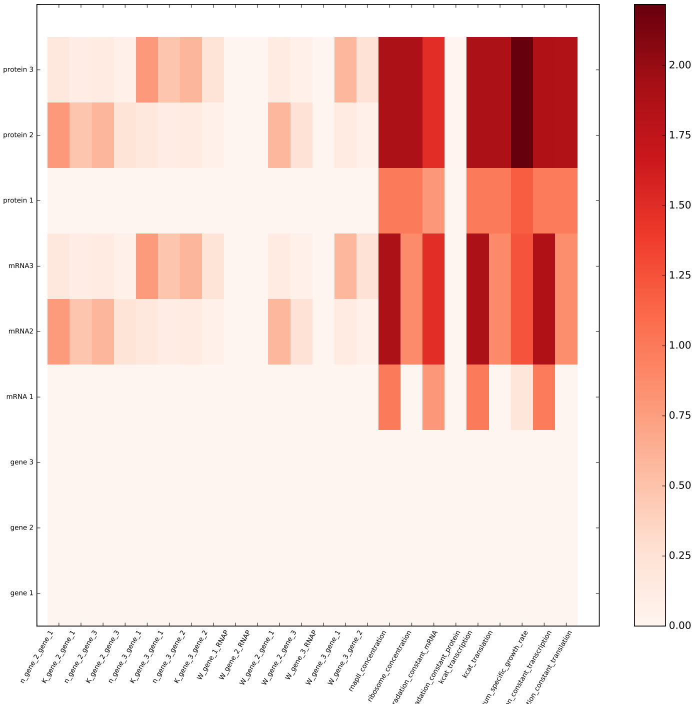

Figure 3: Heat map produced from singular value decomposition of three gene network during induction

Problem 3

Based on close examination of Figure 2 in the provided problem set, it appears that the levels of P3 and P1 and P2 are inversely related, so that either P1 or P2 represses the production of P3. Since it appears that the level of P1

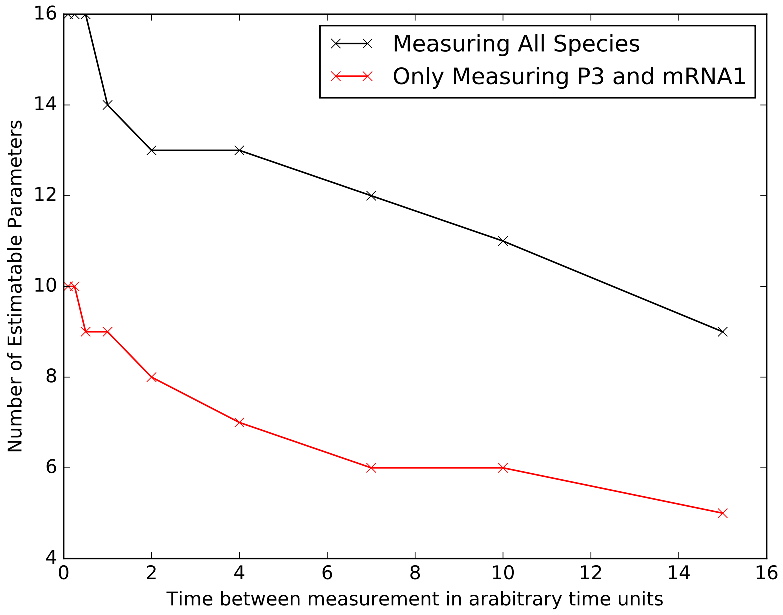

Figure 4: Number of estimable parameters at $\epsilon = .001$ as a function of sampling frequency

begins to increase before the level of P2 begins to rise, I hypothesized that the inducer interacts with P1. This leads to network 1:

$$\texttt{gene}_1\ \texttt{induces (gene}_2\texttt{)}$$
$$\texttt{gene}_2\ \texttt{activates (gene}_1\texttt{)}$$
$$\texttt{gene}_2\ \texttt{represses (gene}_3\texttt{)}$$

After optimization, this leads to a pretty decent fit of the model data, as shown in Figure 5. I used the Nelder-Mead algorithm as implemented by the NLopt package to perform the optimization, by minimizing the summed mean squared errors between the given protein data and the model prediction, with the bounds set to 100 times the hand fit parameters to 1/100th of the hand fit parameters.

Another possibility is that once gene 1 is induced, it activates gene 2 and inhibits gene 3, while protein 2 reduces the production of protein three, as described in network 2:

$$\texttt{gene}_1\ \texttt{induces gene}_2$$
$$\texttt{gene}_1\ \texttt{represses gene}_3$$
$$\texttt{gene}_2\ \texttt{represses gene}_3$$

This model, even after optimization, did not appear to do go as good of a job of capturing the experimental data, as it under-predicts all of the protein concentrations, as shown in Figure 6.

However, I opted to carry the analysis further and used the method of Kremling et all to inform my decision about which experiment should be performed to distinguish between the two models. I once again used the Nelder Mead algorithm as implemented by NLopt to maximize u, essentially, the difference between the model performance under differing experimental conditions. I allowed the weights of the RNAP on the genes to vary, simulating inducing different genes, as well as the initial concentrations of the genes, simulating knock outs. Following this, it appeared that the most informative experiment to perform was to induce gene 1, which when provided with experimental results supported model structure 1. Even though model 2 has a lower AIC, it completely fails to capture the dynamics of protein 3.

I also decided to compare the case of the feed back loop to having the feed back loop absent.

$$\texttt{gene}_1\ \texttt{induces (gene}_2\texttt{)}$$
$$\texttt{gene}_2\ \texttt{represses (gene}_3\texttt{)}$$

The case with no feedback loop will be named case 5, because the intermediate cases did not prove illuminating. After parameter optimization, this proposed network layout also failed to precisely capture the protein 3 dynamics,

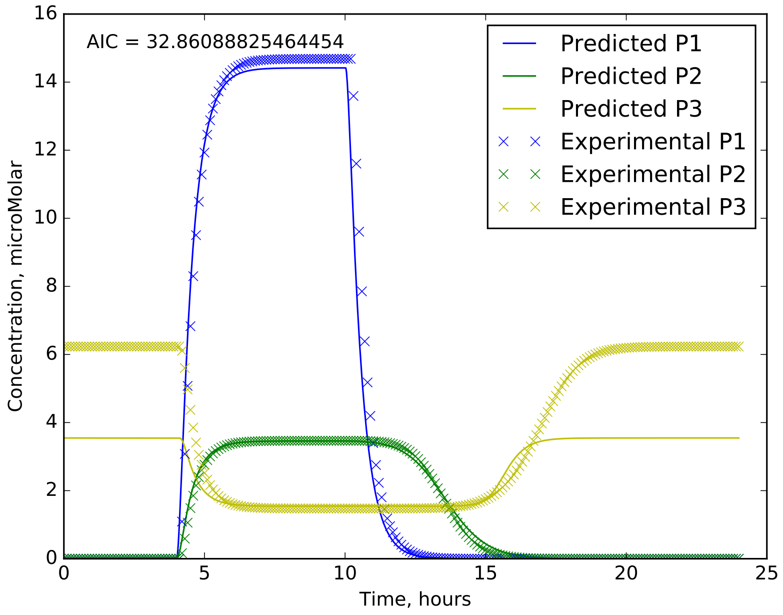

Figure 5: Model performance for network layout 1. Xs represent simulated data, lines represent model predictions.

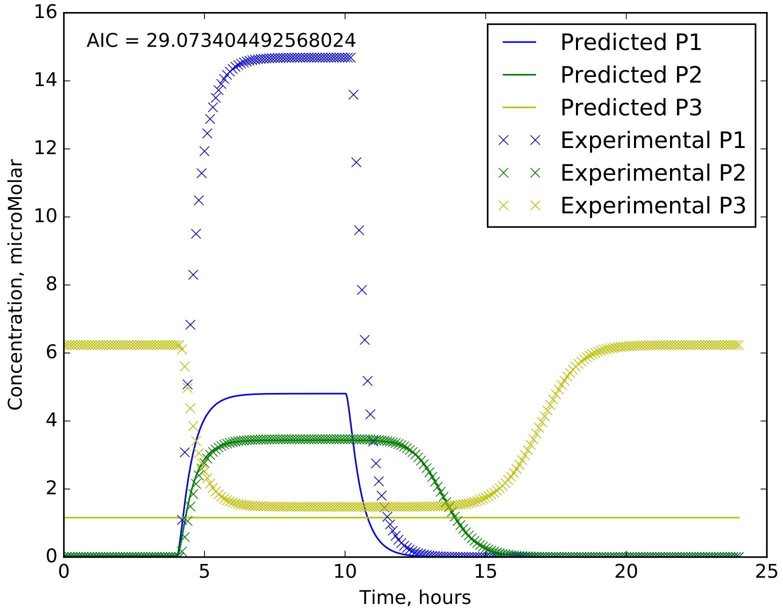

Figure 6: Model performance for network layout 2. Xs represent simulated data, lines represent model predictions.

but had a low AIC, partially because it contains fewer parameters than models 1 and 2. I figured that the easiest way to differentiate between models 1 and 5 would be to induce gene 2-as if the feedback loop is present, when gene 2 is induced, the concentration of protein 1 will rise, if not, it will remain low. Figure 7 shows the performance on this model on the experimental data originally given, and Figure 8 shows its performance when only gene 2 is induced.

Model 5 performs very well in this case, as quantified by the low AIC for both the case of gene 1 and gene 2 being induced, leading me to hypothesize that it is more likely than model 1, as it has a low AIC for the case of inducing gene 1. In conclusion, I believe model 5 is the most probable, followed by model 1, and then 2.

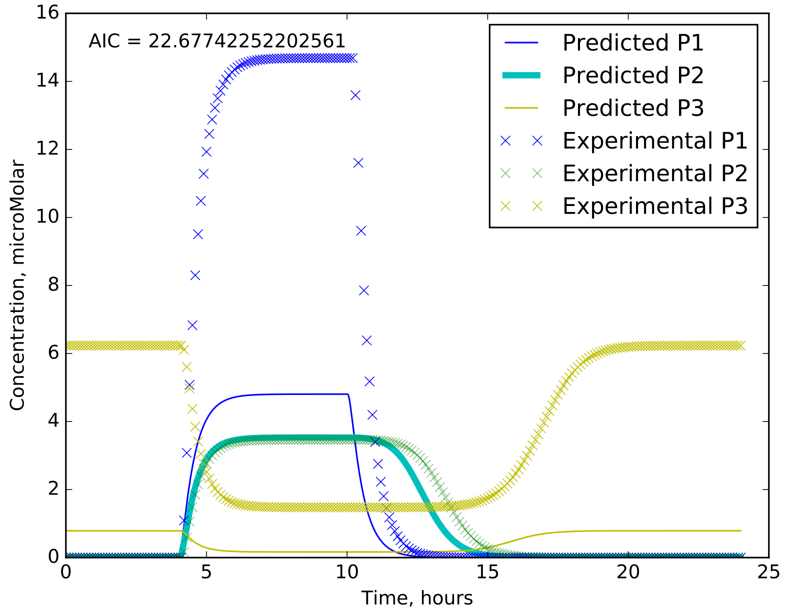

Figure 7: Model performance for network layout 5. Xs represent simulated data, lines represent model predictions. Using parameters generated by fitting on inducing gene 2. While the fit is far from perfect, these parameters do capture the shape of the protein interactions.

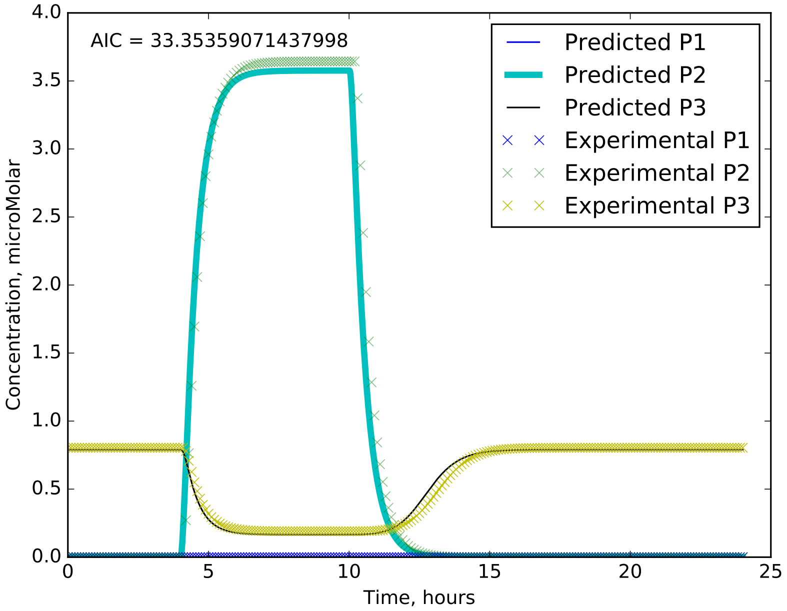

Figure 8: Model performance for network layout 5 when gene 2 is induced. Xs represent simulated data, lines represent model predictions.

Supplemental Material

Details of Estimable Parameters

```
0.1,
n_gene_2_gene_1,K_gene_2_gene_1,n_gene_3_gene_1,K_gene_3_gene_1,W_gene_1_RNAP,W_gene_2_gene_1,W_gene_2_gene_3,W_gene_3_gene_1,W_gene_3_gene_2,rnapII_concentration,ribosome_concentration,degradation_constant_mRNA,kcat_transcription,kcat_translation,maximum_specific_growth_rate,sa
n_gene_3_gene_1,K_gene_3_gene_1,W_gene_1_RNAP,W_gene_3_gene_1,W_gene_3_gene_2,rnapII_concentration,degradation_constant_mRNA,kcat_transcription,kcat_translation,maximum_specific_growth_rate
0.25,
n_gene_2_gene_1,K_gene_2_gene_1,n_gene_3_gene_1,K_gene_3_gene_1,W_gene_1_RNAP,W_gene_2_gene_1,W_gene_2_gene_3,W_gene_3_gene_1,W_gene_3_gene_2,rnapII_concentration,ribosome_concentration,degradation_constant_mRNA,kcat_transcription,kcat_translation,maximum_specific_growth_rate,sa
n_gene_3_gene_1,K_gene_3_gene_1,W_gene_1_RNAP,W_gene_3_gene_1,W_gene_3_gene_2,rnapII_concentration,degradation_constant_mRNA,kcat_transcription,kcat_translation,maximum_specific_growth_rate
0.5,
n_gene_2_gene_1,K_gene_2_gene_1,n_gene_3_gene_1,K_gene_3_gene_1,W_gene_1_RNAP,W_gene_2_gene_1,W_gene_2_gene_3,W_gene_3_gene_1,W_gene_3_gene_2,rnapII_concentration,ribosome_concentration,degradation_constant_mRNA,kcat_transcription,kcat_translation,maximum_specific_growth_rate,sa
n_gene_3_gene_1,K_gene_3_gene_1,W_gene_1_RNAP,W_gene_3_gene_2,rnapII_concentration,degradation_constant_mRNA,kcat_transcription,kcat_translation,maximum_specific_growth_rate
1.0,
n_gene_2_gene_1,n_gene_3_gene_1,W_gene_1_RNAP,W_gene_2_gene_1,W_gene_2_gene_3,W_gene_3_gene_1,W_gene_3_gene_2,rnapII_concentration,ribosome_concentration,degradation_constant_mRNA,kcat_transcription,kcat_translation,maximum_specific_growth_rate,saturation_constant_translation
n_gene_3_gene_1,K_gene_3_gene_1,W_gene_1_RNAP,W_gene_3_gene_2,rnapII_concentration,degradation_constant_mRNA,kcat_transcription,kcat_translation,maximum_specific_growth_rate
2.0,
n_gene_2_gene_1,n_gene_3_gene_1,W_gene_1_RNAP,W_gene_2_gene_1,W_gene_2_gene_3,W_gene_3_gene_1,W_gene_3_gene_2,rnapII_concentration,ribosome_concentration,degradation_constant_mRNA,kcat_transcription,kcat_translation,maximum_specific_growth_rate
n_gene_3_gene_1,K_gene_3_gene_1,W_gene_1_RNAP,W_gene_3_gene_2,rnapII_concentration,degradation_constant_mRNA,kcat_translation,maximum_specific_growth_rate
4.0,
n_gene_2_gene_1,n_gene_3_gene_1,W_gene_1_RNAP,W_gene_2_gene_1,W_gene_2_gene_3,W_gene_3_gene_1,W_gene_3_gene_2,rnapII_concentration,ribosome_concentration,degradation_constant_mRNA,kcat_transcription,kcat_translation,maximum_specific_growth_rate
n_gene_3_gene_1,W_gene_1_RNAP,W_gene_3_gene_2,rnapII_concentration,degradation_constant_mRNA,kcat_translation,maximum_specific_growth_rate
7.0,
n_gene_2_gene_1,n_gene_3_gene_1,W_gene_1_RNAP,W_gene_2_gene_1,W_gene_2_gene_3,W_gene_3_gene_1,W_gene_3_gene_2,rnapII_concentration,ribosome_concentration,degradation_constant_mRNA,kcat_translation,maximum_specific_growth_rate
n_gene_3_gene_1,W_gene_1_RNAP,W_gene_3_gene_2,rnapII_concentration,degradation_constant_mRNA,kcat_translation
10.0,
n_gene_2_gene_1,n_gene_3_gene_1,W_gene_1_RNAP,W_gene_2_gene_1,W_gene_2_gene_3,W_gene_3_gene_1,W_gene_3_gene_2,rnapII_concentration,degradation_constant_mRNA,kcat_translation,maximum_specific_growth_rate
n_gene_3_gene_1,W_gene_1_RNAP,W_gene_3_gene_2,rnapII_concentration,degradation_constant_mRNA,kcat_translation
15.0,
W_gene_2_gene_1,W_gene_2_gene_3,W_gene_3_gene_1,W_gene_3_gene_2,rnapII_concentration,degradation_constant_mRNA,kcat_translation,maximum_specific_growth_rate,saturation_constant_translation
W_gene_3_gene_2,rnapII_concentration,degradation_constant_mRNA,kcat_translation,maximum_specific_growth_rate
```