This homework consists of 3 problems. This homework unfortunately tend to take much longer to complete, simply because analysis of noise tend to be more complicated, and takes more effort to gain interesting conclusions. But I have done considerable work to streamline the problems, so if you spend extra time, you will for sure be rewarded by the understanding you gained. If you are short on time, it is fine to come back to this homework later in your research career...

(Optional.) Gillespie

Write an implementation of the Gillespie algorithm to simulate stochastic trajectories. This helps you to gain intuitions about the stochastic models described in this homework. Here is an implementation of Gillespie if you want one that is ready to use: https://colab.research.google.com/drive/1qL7_Bk8_jVJWgPCUF5voUrIlnElhF8e8?usp=sharing.

1 Bursty gene expression

Noise becomes important for gene expression when the number of proteins expressed is low and the expression kinetics is bursty. We investigate how bursty-ness emerge in different mechanistic models of gene expression, and focus on the unregulated case, and leave regulatory mechanisms such as feedback to the next problem. The interpretations mainly follow this paper [1] giving a clear review of noise analysis for gene expression. By walking through this problem, we have covered the main takeaways about what people have learned about bursty gene expression.

1.1 Poisson from simple production and degradation (Optional)

Let us first establish the fact that the molecule count of a species regulated by simple production and degradation reactions follow a Poisson distribution. Let us consider the production and degradation of mRNA, for example.

$$G \xrightarrow{k_m} G + M, \qquad M \xrightarrow{\gamma_m} \emptyset. \tag{1}$$

1. Show that the chemical master equation is the following.

$$\frac{d}{dt}p(M,t) = [k_m G p(M-1) - k_m G p(M)] + [\gamma_m (M+1)p(M+1) - \gamma_m M p(M)]. \tag{2}$$

   Here M is the copy number of M, and G denotes the copy number of G. Note that G is constant.

2. Then show that at steady state, the probabilities follow

$$\begin{aligned}(M+1)p(M+1) =& (\lambda + M)p(M) - \lambda p(M-1), \qquad M = 1, 2, \dots \\ p(1) =& \lambda p(0),\end{aligned} \tag{3}$$

   where $\lambda = \frac{k_m G}{\gamma_m}$.

3. Use induction to show that the following formula holds for steady state $p(M)$:

$$p(M+1) = \frac{\lambda}{M+1}p(M), \qquad M = 0, 1, 2, \dots. \tag{4}$$

   (Hint: Show that it holds for $M = 0$, then show that for a given $M = 1, 2, \dots$, if $p(M+1) = \frac{\lambda}{M+1}p(M)$, then $p(M+2) = \frac{\lambda}{M+2}p(M+1)$. Then use induction.)

4. Use the above equation and the fact that $\sum_{M=0}^{\infty} p(M) = 1$ to show that $p(0) = e^{-\lambda}$, therefore

$$p(M) = \frac{\lambda^M}{M!} e^{-\lambda}. \tag{5}$$

This is the probability mass function (PMF) of a Poisson distribution with rate λ, which is also the mean. So we have established that the distribution of M follows a Poisson distribution.

1.2 Translational bursts

Let us first consider the vanilla process of transcription and translation in gene expression, and see how the signal-amplification due to the transcription and translation process itself can be viewed as bursty-ness and increase noise. In other words, this model considers gene expression bursty-ness to come from the fact that each mRNA transcript is expressed several times. This is the model used in [2] in 2002, for example, to explain the noise in gene expressions measured. This model is also used in [3] when the stochastic expression of the proteome and transcriptome of *E. coli* at the single molecule level was first measured in 2010, although this model has caveats which we will discuss at the end.

1. Consider the following reaction network for a simple transcription-translation process,

$$G \xrightarrow{k_m} G + M, \quad M \xrightarrow{k_p} M + P, \quad M \xrightarrow{\gamma_m} \emptyset, \quad P \xrightarrow{\gamma_p} \emptyset. \tag{6}$$

Since the gene molecule's number G does not change, only the numbers of mRNA and protein molecules change. Let (M, P) denote the number of mRNA M and protein P molecules respectively. Let $p(M, P, t)$ denote the probability that the number of mRNAs and proteins are M and P respectively at time t. Note that since the mRNA is regulated by a simple production and degradation process by itself, independent of P, we know from the previous subproblem that M follows a Poisson distribution with mean $\frac{k_m G}{\gamma_m}$.

Show that the chemical master equation of this reaction system is the following.

$$\begin{aligned}\frac{d}{dt}p(M,P,t) =& [k_m G p(M-1,P) - k_m G p(M,P)] + [k_p M p(M,P-1) - k_p M p(M,P)] \\ &+ [\gamma_m (M+1) p(M+1,P) - \gamma_m M p(M,P)] + [\gamma_p (P+1) p(M,P+1) - \gamma_p P p(M,P)].\end{aligned} \tag{7}$$

2. (Optional.) Let $\langle \cdot \rangle$ denote the average, e.g. $\langle M \rangle$ denotes the average of M. Let $\mathrm{Var}(\cdot)$ denote the variance, e.g. $\mathrm{Var}(M) = \langle M^2 \rangle - \langle M \rangle^2$, and let $\mathrm{Cov}(\cdot)$ denote the covariance, e.g. $\mathrm{Cov}(M, P) = \langle MP \rangle - \langle M \rangle \langle P \rangle$. Show that the moments of order two or less satisfy

$$\begin{aligned}\frac{d}{dt}\langle M \rangle =& k_m G - \gamma_m \langle M \rangle, \\ \frac{d}{dt}\langle P \rangle =& k_p \langle M \rangle - \gamma_p \langle P \rangle, \\ \frac{d}{dt}\left\langle M^2 \right\rangle =& 2k_m G \langle M \rangle + k_m G - 2\gamma_m \left\langle M^2 \right\rangle + \gamma_m \langle M \rangle, \\ \frac{d}{dt}\langle MP \rangle =& k_m G \langle P \rangle + k_p \left\langle M^2 \right\rangle - \gamma_m \langle MP \rangle - \gamma_p \langle MP \rangle, \\ \frac{d}{dt}\left\langle P^2 \right\rangle =& 2k_p \langle MP \rangle + k_p \langle M \rangle - 2\gamma_p \left\langle P^2 \right\rangle + \gamma_p \langle P \rangle.\end{aligned} \tag{8}$$

From the above, show that the mean and covariances satisfy

$$\begin{aligned}
\frac{d}{dt}\langle M\rangle &= k_m G - \gamma_m \langle M\rangle,\\
\frac{d}{dt}\langle P\rangle &= k_p \langle M\rangle - \gamma_p \langle P\rangle,\\
\frac{d}{dt}\operatorname{Var}(M) &= k_m G + \gamma_m \langle M\rangle - 2\gamma_m \operatorname{Var}(M),\\
\frac{d}{dt}\operatorname{Cov}(M,P) &= k_p \operatorname{Var}(M) - (\gamma_m+\gamma_p)\operatorname{Cov}(M,P),\\
\frac{d}{dt}\operatorname{Var}(P) &= k_p \langle M\rangle + 2k_p \operatorname{Cov}(M,P) + \gamma_p\langle P\rangle - 2\gamma_p \operatorname{Var}(P)
\end{aligned} \tag{9}$$

3. Solve the steady state moment equations to obtain the steady state moments, and show that they satisfy the following:

$$\begin{aligned}
\langle M\rangle &= \frac{k_m}{\gamma_m} G,\\
\langle P\rangle &= \frac{k_p}{\gamma_p}\frac{k_m}{\gamma_m} G = \frac{k_p}{\gamma_p}\langle M\rangle,\\
\operatorname{Var}(M) &= \frac{k_m}{\gamma_m} G = \langle M\rangle,\\
\operatorname{Cov}(M,P) &= \frac{k_p}{\gamma_m+\gamma_p}\operatorname{Var}(M) = \frac{\gamma_p}{\gamma_m+\gamma_p}\langle P\rangle,\\
\operatorname{Var}(P) &= \langle P\rangle + \frac{k_p}{\gamma_p}\operatorname{Cov}(M,P) = \langle P\rangle\left(1+\frac{k_p}{\gamma_m+\gamma_p}\right).
\end{aligned} \tag{10}$$

Here the expressions for $\operatorname{Cov}(M,P)$ and $\operatorname{Var}(P)$ are re-written with the equations for lower moments.

4. Recall that the simple production-degradation process $\emptyset \rightleftharpoons X$ results in a Poisson distribution of X such that its variance is equal to the mean, $\operatorname{Var}(X) = \langle X\rangle$. Therefore, we can use the ratio of variance to mean to compare a distribution to the Poisson distribution. This is called the **Fano factor**, $F(X) = \frac{\operatorname{Var}(X)}{\langle X\rangle}$.

For the Fano factor of P, we see that

$$F(P) = \frac{\operatorname{Var}(P)}{\langle P\rangle} = 1 + \frac{k_p}{\gamma_m+\gamma_p}. \tag{11}$$

Typically, the degradation rate of proteins is much slower than the degradation rate of mRNAs, so $\gamma_p \ll \gamma_m$, we we obtain

$$F(P) = \frac{\operatorname{Var}(P)}{\langle P\rangle} \approx 1 + b, \tag{12}$$

where we define $b = \frac{k_p}{\gamma_m}$. Here b is the average number of proteins synthesized per transcript, which can be interpreted as the **burst size** of protein production caused by each transcription event.

We see that the noise is increased by a factor of b due to transcription-translation.

Some moderations of the interpretation above are due. This model's explanation for bursty gene expression is rather unsatisfactory for the following reason. It showed that, when mRNAs have short lifetimes $\gamma_p \ll \gamma_m$, the

noise of protein $\mathrm{Var}(P)$ is increased by a factor $b = \frac{k_p}{\gamma_m}$ which is the number of proteins synthesized per mRNA transcript. However, this increase in noise does not necessarily mean there are actually bursty translational events. Indeed, when there are hundreds of mRNAs in the cell all expressing proteins, the stochasticity of each mRNA's translational events is evened out, and the production of proteins is more like a continuous steady stream, rather than short bursts. Only when the number of mRNAs in the cell is close to zero or one would this translational burst interpretation be physically sound. However, most genes do not have such a low mRNA copy number, so the bursty features of fluorescence observed for gene expression CANNOT be mechanistically explained by the fact that each mRNA produces multiple proteins.

1.3 Random burst sizes give rise to Gamma or Negative Binomial distributions (Optional)

Instead of starting with a mechanistic explanation about how the bursts happen, we could also start with experimental observations and focus on the distributional shape. For example, it was observed that the burst size b of protein production is roughly exponentially distributed in [4], the famous paper that was the first to do single-cell single-molecule measurements of gene expression profiles. To be exact, since the burst size is about the number of protein molecules, b should take integer values. So the exact distribution could be that b follows a geometric distribution, as was considered in [5, 6] in 2000.

Let us investigate the resulting distribution on protein numbers given a geometric or exponentially distributed burst size b, while assuming the number of mRNA is roughly constant.

1. In [4], a simple argument was made to show that the distribution of protein number should follow a Gamma distribution.

   Consider again the model in Eqn (6), but with the translation replaced with bursty expression and the burst size is exponentially distributed with mean b. In other words, the translation reaction should be

   $$M \xrightarrow{k_p} M + BP,$$

   where B is a random variable with distribution $\mathrm{Exp}(\frac{1}{b})$ (mean b).

   We first simplify the scenario by assuming mRNA dynamics is fast, therefore its fluctuations can be ignored when considering the protein's dynamics. Indeed, since mRNA's degradation rate γ_m tends to be much larger than that of protein, γ_p, we can assume mRNA is always at a steady state distribution, and the production of proteins comes from a series of random independent burst events.

   Then, let us reason about the mRNAs, i.e. the number of transcription events, that contribute to the current value of protein number. Since the proteins are degraded at a rate of γ_p, the duration that contributes to the current value is $\frac{1}{\gamma_p}$. The number of mRNAs produced in this duration is therefore $k_m G \frac{1}{\gamma_p}$. So the effective number of transcripts that can cause bursty translations and contribute to the current number of proteins is $a = \frac{k_m G}{\gamma_p}$.

   Then, assuming the bursty translational events of these mRNA transcripts are independent of each other, the number of proteins produced per generation is the sum of a independent and identically distributed (iid) random variables of burst size B, where $B \sim \mathrm{Exp}(\frac{1}{b})$.

   Now, based on this argument, show that the number of proteins follow a Gamma distribution with $\mathrm{Gamma}(a, b)$, where a is the shape parameter, and b is the scale parameter.

   (For the exact derivation, see [7] and the book [8] referenced therein.)

2. Here are some interpretations of the Gamma distribution above, as is used in [4, 3]. Show that the coefficient of variation (CV) squared is $\frac{\mathrm{Var}(P)}{\langle P\rangle^2} = \frac{1}{a}$, and the Fano factor is $\frac{\mathrm{Var}(P)}{\langle P\rangle} = b$.

   Note that, from our analysis above, we have $a = \frac{k_m G}{\gamma_p}$, the average number of mRNA transcripts that contribute to the current number of proteins, and $b = \frac{k_p}{\gamma_m}$ is the average number of proteins produced per mRNA molecule. And they are related by $ab = \langle P\rangle$.

   Another observation from [3] is that the mRNA number and protein number for a gene seem un-correlated. Is this in agreement or disagreement with this model? (Hint: Look at $\mathrm{Cov}(M, P)$ analyzed in the previous subproblem.)

   (Note that although the same result is used in [3], which is a very famous paper, it had an error in the part describing the Gamma distribution for protein number. Here, a could be interpreted as the number of mRNAs per cell cycle as described in the paper, but its actual meaning that made the Gamma distribution argument work in the original derivation of [4, 7] was that a is the number of burst events that contribute to the current value of the protein. This is because it was assumed that mRNA degradation rate is much faster than that of proteins, therefore fluctuations of mRNA can be integrated out and proteins are produced in random independent burst events. See Section 3.2 of the Supplementary Information of [1] for more details. Also, the CV squared of protein count is a^{-1}, not a as described in the paper [3], while the Fano factor is indeed b.)

3. (Optional.) Considering the discrete nature of the number of proteins, we can also consider the translational burst size B to be a discrete random variable, therefore geometrically distributed $\mathrm{Geom(p)}$ with mean b, $p = \frac{1}{1+b}$, which has a probability mass function of

$$\mathbb{P}\{B = n\} = (1-p)^k p = (1 - \frac{1}{1+b})^n \frac{1}{1+b}.$$

   The interpretation is that the $\mathbb{P}\{B = n\}$ is the probability that the first successful trial happened after n failed Bernoulli trials, which are independent and have success probability p.

   Given that the sum of a iid geometric random variables with mean b yields a negative binomial random variable $\mathrm{NegBinom}(a, 1/b)$, argue that the distribution for the number of proteins should be $\mathrm{NegBinom}(a, 1/b)$, which has a probability mass function as follows,

$$\mathbb{P}\{P = n\} = (1-p)^n p^a \binom{a+n-1}{n} = \frac{b^n}{(1+b)^{a+n}} \binom{a+n-1}{n}.$$

   Note that this is consistent with the Gamma distribution above, since the negative binomial distribution is a discrete analog of the Gamma distribution. For a detailed explanation of this relations between the Gamma distribution and the negative binomial distribution, see the following excellent response from DeepSeek https://chat.deepseek.com/share/s13mmcengfd3aoio2z.

4. (Optional.) Our argument in this subproblem so far is more intuitive and crude than accurate and exact. In fact, we ignored the noise contribution to protein number from the fluctuations of mRNA copy number. This can be observed by considering the conditional probability of protein number for a fixed mRNA copy number, $p(P|M)$. Argue that, under $\mathrm{Geom}(\frac{1}{b})$ distributed translational bursts, $p(P|M)$ should follow a $\mathrm{NegBinom}(M, \frac{1}{b})$ distribution. This was shown in [5].

As a result of this, if $M = a$, then we recover our result above. However, since M fluctuates with a Poisson distribution, the noise of protein number P should be larger than predicted before. This means, the Fano factor should be larger than b. Indeed, as shown in the previous subproblem, we derived a Fano factor of $1 + b$.

1.4 Transcriptional bursts give rise to Poisson with a zero spike

Previously, the non-Poissonian behaviors all happen at the protein level, and mRNAs obeyed a Poisson distribution. However, it was soon observed that mRNAs have drastically non-Poisson behavior as well. Not only is the noise of mRNA number much larger than Poisson, it was also observed that quite often the distribution is bimodal, with a large spike at zero copy number. In 2014, [9] attributed this behavior in bacteria to the fact that the gene needs to be uncoiled before transcription, so the gene should be considered to have two states, on and off. See the reaction network below.

$$G_0 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} G_1, \qquad G_1 \xrightarrow{k} G_1 + M, \qquad M \xrightarrow{\gamma} \emptyset. \tag{13}$$

This is also sometimes called the telegraph model. We analyze the transcriptional noise in this case.

1. Let us first analyze the distribution of the gene, since its dynamics is independent of M. Let G_{tot} denote the total number of gene molecules. Show that the chemical master equation for $p(G_1, t)$, the probability that there are G_1 copies of the gene molecule in state G_1 at time t follows

$$\frac{d}{dt}p(G_1,t) = [\alpha(G_{\text{tot}} - G_1 + 1)p(G_1 - 1, t) - \alpha(G_{\text{tot}} - G_1)p(G_1, t)] + [\beta(G_1 + 1)p(G_1 + 1, t) - \beta G_1 p(G_1, t)], \tag{14}$$

for $G_1 = 1, 2, \ldots, G_{\text{tot}} - 1$, and the boundary cases follow

$$\begin{aligned} \frac{d}{dt}p(0,t) =& \beta p(1,t) - \alpha p(0,t), \\ \frac{d}{dt}p(G_{\text{tot}},t) =& \alpha p(G_{\text{tot}} - 1, t) - \beta p(G_{\text{tot}}, t). \end{aligned} \tag{15}$$

We can directly solve the above to get the solution for $p(G_1, t)$. Instead, we could also directly get the answer by intuitive arguments, and check that the answer is true.

Let $p_0(t) = \frac{G_0}{G_{\text{tot}}}$ and $p_1(t) = \frac{G_1}{G_{\text{tot}}}$ denote the fraction (probability) of gene molecules that a gene molecule is in G_0 or G_1 state at time t. Since each gene molecule transits between the G_0 and G_1 state independently, we have

$$\frac{d}{dt}p_1(t) = \alpha p_0(t) - \beta p_1(t) = \alpha(1 - p_1(t)) - \beta p_1(t).$$

This can be easily solved explicitly. And the probability $p(G_1, t)$ satisfies

$$p(G_1,t) = \binom{G_{\text{tot}}}{G_1} p_1(t)^{G_1} p_0(t)^{G_0},$$

a Binomial distribution with probability parameter $p_1(t)$.

Check that indeed this follows the CME for $p(G_1, t)$ above. So the steady state distribution is

$$p(G_1) = \binom{G_{\text{tot}}}{G_1} p_1^{G_1} p_0^{G_0},$$

where $p_1 = \frac{\alpha}{\alpha+\beta}$, and $p_0 + p_1 = 1$. This is the $\mathrm{Binom}(G_{\mathrm{tot}}, p_1)$ distribution for steady state G_1.

So the noise of the gene follows

$$\frac{\mathrm{Var}(G_1)}{\langle G_1\rangle^2} = \frac{p_0}{p_1 G_{\mathrm{tot}}} = \frac{p_0}{\langle G_1\rangle}.$$

Note that this noise is smaller than Poisson by a factor of p_0. So by having active and inactive states, and switching between them, we can have a low molecule number G_{tot} while having relatively small noise, as long as most molecules are active, $p_1 \approx 1$. However, this model did not account for the noise of G_{tot}, which could have noise contributions that make the noise of G_1 larger than Poisson.

2. Now we derive the distribution of M. Recall that, from the first subproblem, if $G_1 = G_{\mathrm{tot}}$ always, then steady state M follows $\mathrm{Poisson}(\lambda)$, where $\lambda = \frac{kG_{\mathrm{tot}}}{\gamma}$. Now that G_1 is $\mathrm{Binom}(G_{\mathrm{tot}}, p_1)$ distributed with $p_1 = \frac{\alpha}{\alpha+\beta}$, we see that M's steady state distribution should be close to a Poisson distribution with its rate following a binomial distribution. The actual distribution is slightly different from this since this argument assumes M reaches steady state distribution instantly when G_1 changes, while actually the dynamics of M and G_1 are correlated and changing simultaneously.

   Still, we can get a good enough answer to glean some intuitions based on this argument that M reaches steady state instantly when G_1 changes, which corresponds to the assumption that $\gamma \gg \alpha + \beta$, since the former is the rate for dynamics of M, and the latter is for the dynamics of G_1.

   Let us consider individual gene molecules, and consider the mRNAs produced by each gene molecule. In other words, let us first consider the case that $G_{\mathrm{tot}} = 1$. Let M_1 denote the number of mRNAs in this case. Show that, the steady state M_1 follows

$$M_1 \sim \begin{cases} \mathrm{Poisson}(\lambda_1), & \text{with probability } p_1 = \frac{\alpha}{\alpha+\beta}; \\ 0, & \text{with probability } p_0 = 1 - p_1; \end{cases} \tag{16}$$

   where $\lambda_1 = \frac{k}{\gamma}$.

   Then show that, the mean and variance of M_1 are

$$\begin{aligned} \langle M_1\rangle =& p_1\lambda_1 = \frac{\alpha}{\alpha+\beta}\frac{k}{\gamma}, \\ \mathrm{Var}(M_1) =& p_1\lambda_1(1 + p_0\lambda_1). \end{aligned} \tag{17}$$

   Conclude that, since M is the result of G_{tot} copies of the gene molecule, each independently producing mRNAs, we have M is the sum of G_{tot} coplies of M_1, therefore satisfies

$$\begin{aligned} \langle M\rangle =& p_1\lambda = \frac{\alpha}{\alpha+\beta}\frac{kG_{\mathrm{tot}}}{\gamma}, \\ \mathrm{Var}(M) =& p_1\lambda(1 + p_0\lambda_1). \end{aligned} \tag{18}$$

   So we see that the distribution of M is a "spike" at zero mixed with a Poisson distribution.

   From the above, we see that in the telegraph model, the Fano factor satisfies

$$F = \frac{\mathrm{Var}(M)}{\langle M\rangle} = 1 + p_0\lambda_1 = 1 + \frac{\beta}{\alpha+\beta}\frac{k}{\gamma}.$$

   So the noise of transcription is amplified by the factor $p_0\lambda_1$. This comes from the noise of G_1, since G_1 has a Fano factor of p_0, and this is amplified by the production rate λ_1 per G_1 molecule.

3. (Optional.) Now let us solve the mean and variance exactly. Write down the chemical master equation for $p(M, G_1, t)$, the probability that there are M copies of mRNA and G_1 coplies of the gene molecule in state G_1 at time t. Then calculate the steady state mean and variances, and show that

$$\frac{\mathrm{Var}(M)}{\langle M \rangle} = 1 + p_0 \lambda_1 \frac{\tau_G}{\tau_G + \tau_M}, \tag{19}$$

where $\tau_G = \frac{1}{\alpha+\beta}$ is the timescale of G, and $\tau_M = \frac{1}{\gamma}$ is the timescale of M.

So we see that compared to the result assuming fast M kinetics, the noise is slightly smaller, accounting for the timescale difference between M and G. When $\tau_G \gg \tau_M$, we recover the fast M limit. When $\tau_G \ll \tau_M$, we have fast G_1 dynamics, so G_1 practically fixed at the time scale that M varies, and there is no noise contribution to M from G_1.

1.5 Eukaryote transcriptional bursts give rise to plateaus and long-tails (Optional)

Later on, observations of transcript number distribution in mammalian cells seem to deviate even from the telegraph model, and call for a third state for the gene. For example, as shown in the data and argued by parameter inference in [10] from 2024, the number of mRNA transcripts tend to be long-tailed, bi-modal, or have a plateau near zero. This relate to further chromosomal and methylation mechanisms of transcriptional regulation in mammalian cells. Intuitively, due to the multitude of regulations, we can consider genes in mammalian cells to be "off by default". Then, upon removal of the super-coiling, methylations, etc, the gene enters a "weakly accessible" state, but is not actively transcribed. Lastly, when activating components such as transcription factors bind, the gene enters a highly active state.

To model this, we can consider the reaction network below.

$$G_0 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} G_1 \underset{k_3}{\overset{k_4}{\rightleftharpoons}} G_2, \qquad G_1 \xrightarrow{k_{p1}} G_1 + M, \;\; G_2 \xrightarrow{k_{p2}} G_2 + M, \qquad M \xrightarrow{\gamma} \emptyset. \tag{20}$$

Here G_0 is completely off, G_1 is weakly on, and G_2 is strongly on, so k_{p2} is significantly larger than k_{p_1}.

Explore the behavior of this model of bursty transcription through either analysis or simulations. Find scenarios where distributions from this model is unlikely to be explained by the telegraph model where the gene only has two states.

2 Summing up the noise in biomolecular systems

In the previous problem, we considered gene expression noise in an un-regulated setting. What happens if the gene is regulated? Could the noise be suppressed? Would there be a limit to noise suppression given a finite budget in terms of molecules expended? More generally, how should we understand the relations between means and covariances generated by the reaction mechanisms underneath?

To answer such questions, it would be cumbersome to solve the chemical master equations (CMEs) on a case by case basis. Also, when we consider highly nonlinear regulations, we could easily obtain CMEs that does not have moment closure, so we cannot solve the mean and variance exactly. Due to these reasons, we would like to derive general relations and formula governing the mean and variance for a given stochastic reaction network, and reveal the important structures in these relations. Also, we would like to employ approximations so that we can always solve and get some answers, since an approximate answer or any intuitive understanding is always better than an exact "no idea".

For this goal, let us consider general relations between the means and covariances, and employ linear noise approximation (LNA) to obtain formula that can always be used to solve for answers. This is largely based on

works from Johan Paulsson lab, from his early work [11] in 2004 deriving fluctuation-dissipation relations for LNA, to [12, 13] in 2016 deriving general relations for means and covariances. We are going to introduce these results in the reverse order of how they were discovered, since this fits the intuitive process of zooming in from a general picture.

General relation between means and covariances (Background)

The following is adapted from Section II.A of [14] (also available at https://chemaoxfz.github.io/assets/pdf/xiao,doyle-2019-coupled-reaction-networks-for-noise-suppression.pdf.) It was originally shown in [13, 12], and used in later works from the Hilfinger lab, such as

For a chemical reaction network with reactions

$$\boldsymbol{x} \xrightarrow{r_k(\boldsymbol{x})} \boldsymbol{x} + \boldsymbol{d}_k, \tag{21}$$

where $\boldsymbol{x} \in \mathbb{Z}^n_{\geq 0}$ is a vector of molecular species' counts, $r_k : \mathbf{R}^n \to \mathbf{R}_{\geq 0}$ is the reaction rate function for the kth reaction, and $\boldsymbol{d}_k \in \mathbb{Z}^n$ is the reaction stoichiometry vector for the kth reaction, $k = 1, \dots, m$, we can write the dynamics for the mean as

$$\frac{d}{dt}\langle x_i \rangle = \left\langle \sum_{k:d_{ik}>0} d_{ik} r_k(\boldsymbol{x}) \right\rangle - \left\langle \sum_{k:d_{ik}<0} |d_{ik}| r_k(\boldsymbol{x}) \right\rangle = \left\langle R_i^+(\boldsymbol{x}) \right\rangle - \left\langle R_i^-(\boldsymbol{x}) \right\rangle, \tag{22}$$

where $R_i^{\pm}(\boldsymbol{x})$ are the production or degradation rates of x_i.

We would like to write the quantities in terms of variables that are intuitive. So we introduce the **average lifetime** of x_i, denoted τ_i, and the **average step size** of x_i's change due to reactions that also change x_j, denoted $\left\langle s_{i|j} \right\rangle$. We define them in detail below.

At steady state, we have $\left\langle R_i^+(\boldsymbol{x}) \right\rangle = \left\langle R_i^-(\boldsymbol{x}) \right\rangle$. So we can define

$$\tau_i = \frac{\langle x_i \rangle}{\left\langle R_i^{\pm} \right\rangle} \tag{23}$$

as the average **lifetime** of x_i.

The average step size is defined as

$$\left\langle s_{i|j} \right\rangle = \sum_k \rho_{jk} |d_{ik}| \operatorname{sgn}\{d_{ik} d_{jk}\}, \qquad \rho_{jk} = \frac{|d_{jk}| \langle r_k \rangle}{\sum_{k'} |d_{jk'}| \langle r_{k'} \rangle}, \tag{24}$$

where ρ_{jk} is the probability that when x_j changes, this change comes from reaction k. The sign for the changes to x_i and x_j in reaction k is accounted for by $\operatorname{sgn}\{d_{ik} d_{jk}\}$, which is $+1$ if they are both increased or decreased, and -1 if the signs of change are opposite.

When $i = j$, we see $\left\langle s_{i|i} \right\rangle = \sum_k |d_{ik}| \rho_{ik}$, so it is an average of step sizes across all reactions that change x_i. For example, if x_1 has only one production reaction $x_1 \to x_1 + 1$ and one degradation reaction $x_1 \to x_1 - 10$, then $\left\langle s_{1|1} \right\rangle = \frac{1+10}{2}$, because production and degradation fluxes are always equal at steady state.

As for $i \neq j$, $\left\langle s_{i|j} \right\rangle$ is nonzero only if there are reactions that simultaneously change x_i and x_j. For an example, if the only reaction that have simultaneous changes to x_1 and x_2 is $(x_1, x_2) \to (x_1 - 1, x_2 + 1)$, i.e. one x_1 becomes

one x_2, and x_2 have no other production reactions so that this reaction accounts for all production fluxes of x_2, then $\left\langle s_{1|2} \right\rangle = -\frac{1}{2}$. It is negative because when x_1 decreases, x_2 increases. It is divided by 2 because this reaction accounts for all of x_2's production, therefore half of x_2's changes.

Note that the step sizes and the production and degradation rates are related to each other:

$$\left\langle s_{j|i} \right\rangle \left\langle R_i^{\pm} \right\rangle = \left\langle s_{i|j} \right\rangle \left\langle R_j^{\pm} \right\rangle. \tag{25}$$

Now we have the terminologies that can capture all the variations in a reaction system, so we can tackle the relation between mean and covariances. If we write the dynamics for the covariances, at steady state we have the following equation.

$$\mathrm{Cov}\left(x_i, R_j^+ - R_j^-\right) + \mathrm{Cov}\left(x_j, R_i^+ - R_i^-\right) + \sum_k d_{ik} d_{jk} \langle r_k \rangle. \tag{26}$$

One major achievement of the theoretical investigations in [13] is the re-writing of the above equation using the physically interpretable quantities of lifetimes and step sizes. The re-written equation is as follows:

$$\boldsymbol{U} + \boldsymbol{U}^\intercal + \boldsymbol{D} = 0, \tag{27}$$

where

$$U_{ij} = \frac{1}{\tau_j} \frac{\mathrm{Cov}\left(x_i, R_j^+ - R_j^-\right)}{\langle x_i \rangle \left\langle R_j^{\pm} \right\rangle}, \qquad D_{ij} = \frac{\sum_k d_{ik} d_{jk} \langle r_k \rangle}{\langle x_i \rangle \langle x_j \rangle} = \frac{1}{\tau_i} \frac{\left\langle s_{j|i} \right\rangle}{\langle x_j \rangle} + \frac{1}{\tau_j} \frac{\left\langle s_{i|j} \right\rangle}{\langle x_i \rangle}. \tag{28}$$

Here τ_i is the average life time of an x_i molecule, and $\left\langle s_{i|j} \right\rangle$ is the average step size of x_i's change from reactions that also change x_j.

$\boldsymbol{D}$ is the "diffusion" matrix, capturing the randomizing changes due to reactions, and $\boldsymbol{U}$ captures the correlations between concentrations and production degradation fluxes.

(Caution about notation: $\left\langle s_{i|j} \right\rangle$ in our notation here is the same as s_{ji} in [13] and other related papers. Similarly, in the definition of U_{ij}, the term $R_j^+ - R_i^-$ is changed to $R_j^- - R_j^+$ in these other papers. We use production minus degradation to be consistent with notations for linearization of deterministic systems. See the next subsection.)

Eqn (27) is very powerful because of its generality. Note that this formula is applicable to all possible chemical reaction networks, and we have made no assumptions at all so it is exact! All we have done thus far is re-writing the quantities in the steady state equations for the first two moments of CMEs in terms of physically interpretable quantities. It may be questioned that since this is just a rewrite, why would it be useful at all? The power comes from exactly the fact that the quantities we express the relation in terms of are all physically relevant, and often experimentally determinable. Even in cases with little information, we have rough ranges on these numbers. Therefore, relations on these physical quantities are very powerful, since they can guide our intuition in asking questions, coming up with hypothesis, and designing experiments. Indeed, this is exactly where this result shines at: eliminates possible mechanisms and generates new hypothesis based on variations in data.

Let us note the physical and intuitive quantities we used: average abundances $\langle x_i \rangle$, steady state production degradation fluxes $\mathbf{R}_i^{\pm}$, the average lifetimes τ_i, the average step sizes $\left\langle s_{i|j} \right\rangle$.

2.1 General relation of noise in gene expression

To test your understanding of this formula (Eqn (27)) and get a sense of its power, let us consider its application for noise in gene expression. This is adapted from [12], which applied this formula to the dataset from [3].

Consider the following model that allows all possible dynamics for transcription-translation, with the only assumption that proteins are produced at a rate proportional to mRNA, and proteins have first-order degradation (or dilution.) This yields the following network:

$$x_2 \xrightarrow{\alpha x_1} x_2 + 1, \qquad x_2 \xrightarrow{\beta x_2} x_2 - 1. \tag{29}$$

Here we consider two species, x_1 is mRNA, and x_2 is protein. Note that we have only specified the dynamics of x_2, and have left the dynamics of x_1 and an arbitrary number of other species and their interactions un-specified.

Show that, using Eqn (27), we have

$$\mathrm{CC}_{12} = \frac{\mathrm{CV}_2}{\mathrm{CV}_1}\left(1 - \frac{1}{\langle x_2\rangle}\frac{1}{\mathrm{CV}_2^2}\right) \approx \frac{\mathrm{CV}_2}{\mathrm{CV}_1}, \tag{30}$$

where $\mathrm{CC}_{ij} = \frac{\mathrm{Cov}(x_i,x_j)}{\sqrt{\mathrm{Var}(x_i)\,\mathrm{Var}(x_j)}}$ is the Pearson correlation coefficient between x_i and x_j, and $\mathrm{CV}_i = \sqrt{\frac{\mathrm{Var}(x_i)}{\langle x_i\rangle^2}}$ is the coefficient of variation of x_i.

To simplify the notation relating correlation coefficients, coefficient of variations, and the terms in Eqn (27), let us define $\eta_{ij} = \frac{\mathrm{Cov}(x_i,x_j)}{\langle x_i\rangle\langle x_j\rangle}$, so $\eta_{ii} = \mathrm{CV}_i^2$, and $\mathrm{CC}_{ij} = \frac{\eta_{ij}}{\sqrt{\eta_{ii}\eta_{jj}}}$. The approximation at the end is because for most genes in the data of [3], $\langle x_2\rangle \mathrm{CV}_2^2 \gg 1$, so we can ignore this term.

(Hint: look at the equation for $2U_{22} + D_{22} = 0$, which should yield $\eta_{22} - \eta_{12} = \frac{1}{\langle x_2\rangle}$.)

Note that this gives a very powerful prediction, since it holds for *a class* of biomolecular systems! No matter what is the transcriptional regulation of this gene, just by knowing that the production of the protein by mRNA is first order, and the protein degrades by first order, then we know the correlation coefficient MUST be equal to the ratio of the coefficient of variations of the protein and the mRNA! Of course, without knowing the detailed mechanisms of transcriptional regulation, we cannot calculate CV_2 or CC_{12}. However, we can measure these quantities experimentally! Just like we observed $\langle x_2\rangle \mathrm{CV}_2^2 \gg 1$ to simply the relation based on data! So this can be used to rule out possible mechanisms. In [12], it was shown that the data from [3] does not match this result. Indeed, [3] observed that there is no correlation between mRNA number and protein number, while this result shows CC_{12} is always positive, and due to large protein noise it should be quite large! [13] also eliminated several other mechanisms based on such arguments, and hypothesized that the production rate of proteins must vary from one mRNA transcript to another. For example, maybe transcripts that have longer lifetimes also have decreased translational activity.

Fluctuation-dissipation for linear noise approximation (Background)

While the general relation in (27) is powerful, it tend to give equalities and inequalities that need experimental input to be useful. In other words, because U_{ij} is the covariance between x_i and the production and degradation fluxes of x_j, which could involve arbitrary high order moments, the system of equations in Eqn (27) is not closed, therefore not solvable to give concrete results. Therefore, we would like to perform approximations such that at least Eqn (27) becomes something that is solvable to give an answer.

To do so, we make the linear noise approximation (LNA), which assumes that the noise is not too large compared to the nonlinearities such that the production and degradation rates can be well-approximated by their

linearization,

$$R_i^{\pm}(\boldsymbol{x}) \approx R_i^{\pm}(\boldsymbol{x}^*) + \sum_{j=1}^{n} \frac{\partial R_i^{\pm}}{\partial x_j}(\boldsymbol{x}^*)\Delta x_j, \tag{31}$$

where $\boldsymbol{x}^*$ is the operating point around which the linearization is done. This is just like the linearization of a nonlinear control systems when studying the deterministic dynamics of a biomolecular system. Also, if we recall, a linearized deterministic control system can perfectly capture the behavior of the actual nonlinear system if the dynamics is controlled to happen close to the operating point. Similarly, the linearized stochastic system can also perfectly capture the behavior of the actual nonlinear stochastic system if the concentration is controlled to be close to the operating point AND the noise is controlled to be small enough so random variations does not go far from the operating point.

Let us assume there is a point $\boldsymbol{x}^*$ such that $R_i^{\pm}(\boldsymbol{x}^*) = \left\langle R_i^{\pm} \right\rangle$, and this point is the mean $\langle \boldsymbol{x} \rangle$, so that $R_i^{\pm}(\langle \boldsymbol{x} \rangle) \approx \left\langle R_i^{\pm} \right\rangle$, then we can perform linearization around $\langle \boldsymbol{x} \rangle$, and Eqn (27) becomes

$$\boldsymbol{M}\boldsymbol{\eta} + \boldsymbol{\eta}\boldsymbol{M}^{\intercal} + \boldsymbol{D}, \tag{32}$$

where

$$\eta_{ij} = \eta_{ji} = \frac{\mathrm{Cov}(x_i, x_j)}{\langle x_i \rangle \langle x_j \rangle}, \qquad M_{ij} = \frac{H_{ij}}{\tau_i}, \tag{33}$$

and

$$H_{ij} = \frac{\partial \log R_i^{+}}{\partial \log x_j}(\langle \boldsymbol{x} \rangle) - \frac{\partial \log R_i^{-}}{\partial \log x_j}(\langle \boldsymbol{x} \rangle) = \frac{\partial R_i^{+}(\langle \boldsymbol{x} \rangle)}{\partial x_j} \frac{\langle x_j \rangle}{\left\langle R_i^{-} \right\rangle} - \frac{\partial R_i^{-}(\langle \boldsymbol{x} \rangle)}{\partial x_j} \frac{\langle x_j \rangle}{\left\langle R_i^{+} \right\rangle}. \tag{34}$$

Here η_{ij} is the normalized covariance as defined before, with $\eta_{ii} = \mathrm{CV}_i^2$ as the squared coefficient of variation of x_i, and H_{ij} is the production-degradation order defined as the order of production minus the order of degradation of x_i with respect to x_j. We encountered the production-degradation order H_{ij} in our study of adaptation in homework 4, which encodes the structure of bioregulations. It is interesting to see it appears here again. In different contexts, H_{ij} is also referred to as reaction orders (e.g. for mass-action kinetics), logarithmic gains (e.g. from a controller perspective), sensitivities (e.g. when viewing concentrations as perturbations) or elasticities (e.g. in metabolic control analysis about how catalysis rates change with enzyme concentrations.)

(Caution: in our notation, H_{ij} is the order of production minus the order of degradation, while in most works from Paulsson and Hilfinger groups such as [13], this is defined as the order of degradation minus the order of production, so Eqn (32) becomes $\boldsymbol{M}\boldsymbol{\eta} + \boldsymbol{\eta}\boldsymbol{M}^{\intercal} = \boldsymbol{D}$.)

The Eqn (32) is also called the fluctuation-dissipation theorem, since it takes the same form as the fluctuation-dissipation theorem governing Brownian motion, diffusion processes, or continuous random walks. In fact, after linearization, we can see that all moments collapse to just moments of order two or less. Therefore, the relation between the means and covariances in Eqn (32) can always be implemented via a continuous stochastic process with Gaussian noise, which is also called the chemical Langevin equation (CLE.) However, it should be noted that Eqn (32) is just a moment equation, and there are many stochastic processes, including ones with discrete states, that can implement it. So making a linear noise approximation does not mean we have made Gaussian noise assumption. The noise can take arbitrary distribution, and Eqn (32) only specifies the mean-covariance relation that it has to satisfy.

Eqn (32) is also called the (continuous time) Lyapunov equation in matrix analysis and control theory. In fact, for a linear dynamical system $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$, it is stable, i.e. the real parts of $\boldsymbol{A}$'s eigenvalues are all negative, if and only if

there exists a positive definite matrix $\boldsymbol{P}$ such that $\boldsymbol{AP} + \boldsymbol{PA}^\intercal < 0$, i.e. $\boldsymbol{AP} + \boldsymbol{PA}^\intercal = \boldsymbol{Q}$ and $\boldsymbol{Q}$ is negative definite. In this case, the function $V(\boldsymbol{x}) = \boldsymbol{x}^\intercal \boldsymbol{P} \boldsymbol{x}$ is called the Lyapunov function. It certifies the stability of the system since it decreases over time and it is positive everywhere except at $\boldsymbol{x} = 0$. Compared to Eqn (32), we see that if we take $\boldsymbol{A} = \boldsymbol{M}$, then $\boldsymbol{Q} = -\boldsymbol{D}$. By definition of $\boldsymbol{D}$, we see that it is always positive definite. Therefore, if the normalized covariance matrix $\boldsymbol{\eta}$ exists, this is a certificate that the dynamical system $\dot{\boldsymbol{z}} = \boldsymbol{M}\boldsymbol{z}$ is stable. Recall from homework 4, $\dot{\boldsymbol{z}} = \boldsymbol{M}\boldsymbol{z}$ is the linearization of the deterministic dynamics of the reaction network in the fold-change variables, with $z_i = \frac{x_i - x_i^*}{x_i^*}$. So the fact that covariances exist and stability of the linearized system are tightly related. We will use this fact in the next problem.

2.2 Summing up the noise in gene expression

With the linear noise approximation, we can study noise of gene expression in relatively arbitrary settings, as was done in [11] in 2004. Let us consider a generic transcription-translation process.

$$x_1 \xrightarrow{R_1^+(x_1)} x_1 + 1, x_1 \quad \xrightarrow{R_1^-(x_1)} x_1 - 1, x_2 \xrightarrow{R_2^+(x_1,x_2)} x_2 + 1, x_2 \quad \xrightarrow{R_2^+(x_1,x_2)} x_2 + 1. \tag{35}$$

Here we can consider x_1 as the gene, and x_2 as the mRNA for a transcription process, or x_1 as mRNA and x_2 as proteins for a translation process. We assumed x_1's production and degradation only depends on itself, and x_2's production and degradation can depend on both x_1 and x_2. Note that all of the models considered in Problem 1 are examples of this network.

1. Show that Eqn (32) in this case becomes

$$\begin{bmatrix} H_{11}\eta_{11} + \frac{1}{\langle x_1 \rangle} & \frac{H_{21}}{\tau_2}\eta_{11} + \left(\frac{H_{11}}{\tau_1} + \frac{H_{22}}{\tau_2}\right)\eta_{12} \\ & H_{21}\eta_{12} + H_{22}\eta_{22} + \frac{1}{\langle x_2 \rangle} \end{bmatrix} = 0 \tag{36}$$

   (Hint: the structure of the network implies $H_{12} = 0$, and $D_{ii} = \frac{2}{\tau_i \langle x_i \rangle}$, $D_{12} = D_{21} = 0$. Also note the symmetry $\eta_{ij} = \eta_{ji}$.)

   Solve the above equation to obtain

$$\begin{aligned} \eta_{11} &= \frac{\mathrm{Var}(x_1)}{\langle x_1 \rangle^2} = \frac{1}{\langle x_1 \rangle}\frac{1}{-H_{11}}, \\ \eta_{22} &= \frac{\mathrm{Var}(x_2)}{\langle x_2 \rangle^2} = \frac{1}{\langle x_2 \rangle}\frac{1}{-H_{22}} + \eta_{11}\frac{H_{21}^2}{H_{22}^2}\frac{H_{22}/\tau_2}{H_{22}/\tau_2 + H_{11}/\tau_1}. \end{aligned} \tag{37}$$

   What are some of conclusions you can draw from this result? For example, what happens if $H_{22} \to 0$? This happens when x_2 auto-activates, for example. Also, we see that lifetimes matter. So what happens if x_1's dynamics is very fast, and $\tau_1 \to 0$? What happens if x_1's dynamics is very slow, and $\tau_1 \to \infty$? What happens if x_1 regulates x_2 in an ultra-sensitive fashion, so that $|H_{21}|$ is very large?

   What about auto-repression? If H_{11} or H_{22} are more negative, what happens? Does auto-repression also suppresses noise?

2. Let us compare this to the cases we have analyzed before in Problem 1.

   Let us first consider the simple transcription-translation model where x_1 is the mRNA, and x_2 is protein. In this case, $H_{11} = -1$, $H_{21} = 1$, and $H_{22} = -1$. Show that you have the same result as Eqn (10).

3. Then consider the case where x_1 is the active state of the gene G_1, and x_2 is mRNA. This case needs special treatment when converting to the production-degradation reactions, since the state-transition reactions of the gene $G_0 \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} G_1$ is not production-degradation by themselves. But since $G_{\text{tot}} = G_0 + G_1$ is conserved, we can denote $x_1^{\max} = G_{\text{tot}}$ to make the reaction rates depend on x_1 only, and obtain the following production-degradation reactions

$$x_1 \xrightarrow{R_1^+ = \alpha(x_1^{\max} - x_1)} x_1 + 1, \qquad x_1 \xrightarrow{R_1^- = \beta x_1} x_1 - 1, \tag{38}$$

Show that

$$H_{11} = -\frac{x_1^{\max}}{x_1^{\max} - x_1} = -\frac{1}{p_0}$$

in the notation of Problem 1, with $p_0 = \frac{\beta}{\alpha+\beta}$.

Then show that you obtain

$$\begin{aligned} \eta_{11} &= \frac{1}{\langle x_1 \rangle} p_0, \\ \eta_{22} &= \frac{\text{Var}(x_2)}{\langle x_2 \rangle^2} = \frac{1}{\langle x_2 \rangle} + \eta_{11} \frac{1/\tau_2}{1/\tau_2 + 1/(p_0 \tau_1)}. \end{aligned} \tag{39}$$

Here $\frac{1}{\tau_1} = \beta$, and $\frac{1}{\tau_2} = \gamma$. Compared to the notations in Eqn (19), the natural timescales are $\tau_G = \frac{1}{\alpha+\beta}$, and $\tau_M = \frac{1}{\gamma}$. Although $\tau_M = \tau_2$, we see that τ_G and τ_1 are different.

Show that, written in terms of τ_G and τ_M, we have

$$\eta_{22} = \frac{\text{Var}(x_2)}{\langle x_2 \rangle^2} = \frac{1}{\langle x_2 \rangle} + \eta_{11} \frac{\tau_G}{\tau_G + \tau_M}. \tag{40}$$

So this is the same result as Eqn (19).

Note that, in both of these cases, the linear noise approximation is exact, since the production and degradation rates are linear.

2.3 Limits on noise suppression versus coupling (Optional)

So far, beyond auto-regulations that directly change H_{ii}, we have not considered how feedback regulations can suppress noise. For example, if we make x_1's production rate R_1^+ dependent on x_2 and repressed by x_2, so that $H_{12} < 0$, would this suppress noise?

Intuitively, if x_2 is positively correlated with x_1, then the repression of x_1 should suppress noise, since larger x_1 cause larger x_2 and represses x_1, while smaller x_1 causes smaller x_2 and activates x_1. On the other hand, x_2 is noisy by itself, so this noise may increase the variation of x_1.

Let us investigate this noise suppression problem below. We will show that indeed noise suppression by such feedback is possible, but it has fundamental limits that is much worse than the typical case. However, if we get rid of the feedback requirement and instead let x_1 and x_2 be coupled with each other through reactions to change them simultaneously, then such limits on noise suppression can be broken again!

1. Consider the following network of net-change reactions,

$$\begin{aligned} x_1 &\xrightarrow{R_1^+(x_2)} x_1 + 1, \\ x_1 &\xrightarrow{x_1/\tau_1} x_1 - 1, \\ x_2 &\xrightarrow{\alpha x_1} x_2 + 1, \\ x_2 &\xrightarrow{x_2/\tau_2} x_2 - 1. \end{aligned} \tag{41}$$

Here, x_2's production is catalyzed by x_1, so x_2 acts like a sensor on x_1, and x_2 in turn is used to regulate the production of x_1, and both x_1 and x_2 have first-order degradations. We assumed there is no auto-regulation of x_1 or x_2 to simplify the problem, since we already know the roles of auto-regulation from the previous problems.

Show that, the fluctuation-dissipation equation Eqn (32) becomes the following in this case,

$$\begin{bmatrix} H_{12}\eta_{12} - \eta_{11} + \frac{1}{\langle x_1\rangle} & \frac{1}{\tau_1}H_{12}\eta_{22} + \frac{1}{\tau_2}\eta_{11} - (\frac{1}{\tau_1}+\frac{1}{\tau_2})\eta_{12} \\ & \eta_{12} - \eta_{22} + \frac{1}{\langle x_2\rangle} \end{bmatrix} = 0. \tag{42}$$

(Hint: $H_{11} = H_{22} = -1$, $H_{21} = 1$, $D_{12} = D_{21} = 0$, and $D_{ii} = \frac{2}{\tau_i\langle x_i\rangle}$.)

Note that the above already implies the following:

$$\begin{aligned} \eta_{11} &= \frac{1}{\langle x_1\rangle} + H_{12}\eta_{12} \\ \eta_{22} &= \frac{1}{\langle x_2\rangle} + \eta_{12}. \end{aligned} \tag{43}$$

So we see that, if x_1 and x_2 are positively correlated, therefore $\eta_{12} > 0$, and if $H_{12} < 0$, then η_{11} can be suppressed. Note that this comes at a cost of η_{22} increases.

2. Now we solve the noise $\eta_{11}, \eta_{12}, \eta_{22}$. Rearrange into a linear system of equations for the covariances, and show that the following holds.

$$\begin{bmatrix} 1 & -H_{12} & 0 \\ \frac{1}{\tau_2} & -(\frac{1}{\tau_1}+\frac{1}{\tau_2}) & \frac{1}{\tau_1}H_{12} \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \eta_{11} \\ \eta_{12} \\ \eta_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{\langle x_1\rangle} \\ 0 \\ \frac{1}{\langle x_2\rangle} \end{bmatrix} \tag{44}$$

Solve the above to obtain that

$$\eta_{12} = \frac{1}{1 - H_{12}}\left(\frac{\tau_1}{\tau_1+\tau_2}\frac{1}{\langle x_1\rangle} + H_{12}\frac{\tau_2}{\tau_1+\tau_2}\frac{1}{\langle x_2\rangle}\right). \tag{45}$$

Assume $H_{12} < 0$, and denote $\lambda_1 = \frac{\tau_1}{\tau_1+\tau_2}$, and $\lambda_2 = 1 - \lambda_1$, then

$$\begin{aligned} \eta_{12} &= \frac{1}{1+|H_{12}|}\left(\lambda_1\frac{1}{\langle x_1\rangle} - \lambda_2\frac{|H_{12}|}{\langle x_2\rangle}\right), \\ \eta_{11} &= \frac{1}{\langle x_1\rangle} - \frac{|H_{12}|}{1+|H_{12}|}\left(\lambda_1\frac{1}{\langle x_1\rangle} - \lambda_2\frac{|H_{12}|}{\langle x_2\rangle}\right), \end{aligned} \tag{46}$$

Interpret the above regarding noise suppression of x_1 via x_2's inhibition of x_1's production. When is noise suppressed? If we consider $H_{12} = -1$, what conclusions would you draw?

Then, consider the expression for η_{11}. Could we suppress noise indefinitely in this case? Is there any lower bound or limit on the noise suppression by this mechanism? If we assume we can arbitrarily tune the parameters $\lambda_1 \in [0, 1]$, and $|H_{12}| > 0$, what's the conclusion? What if we assume λ_1 is fixed, and we can only vary $|H_{12}|$?

3. Let us consider the cost of noise suppression. Let us define $N = \frac{N_2}{N_1}$, where $N_1 = \langle x_1 \rangle$ and $N_2 = \alpha \langle x_1 \rangle \tau_1$ are the effective signaling rates, defined as the number of production events of x_1 and x_2 respectively during the lifetime of x_1. Then $N = \alpha\tau_1$ is the number of x_2 production events used per x_1 production event, which is a per-x_1 molecule cost in signaling.

   Let us denote $|H_{12}| = h$. Also, we can use the fact that $\langle x_2 \rangle = \alpha\tau_2 \langle x_1 \rangle$, so $N = \frac{\langle x_2 \rangle}{\langle x_1 \rangle} \frac{\tau_1}{\tau_2}$. This can be used to write

$$\eta_{11} = \frac{1}{\langle x_1 \rangle} \left(1 - \lambda_1 \frac{h}{1+h} (1 - \frac{h}{N}) \right), \tag{47}$$

   where we used $\tau_2\lambda_1 = \tau_1\lambda_2$.

   Now, let us consider all but h is fixed, and minimize η_{11} by varying $h > 0$. Show that the minimizer is

$$h^* = -1 + \sqrt{1+N}. \tag{48}$$

   Then, plug this into the expression, rearrange, we have

$$\eta_{11}(h^*) = \frac{1}{\langle x_1 \rangle} \left(\lambda_2 + \lambda_1 \frac{2}{1 + \sqrt{1+N}} \right). \tag{49}$$

   Since $\lambda_1 + \lambda_2 = 1$, we see that η_{11} is in between two values, $\frac{1}{\langle x_1 \rangle}$ which is the Poisson noise, and $\frac{1}{\langle x_1 \rangle} \frac{2}{1+\sqrt{1+N}}$ which is the suppressed noise. Since for $N > 0$, the suppressed noise is always smaller, we have that the lower bound for η_{11} is achieved when $\lambda_1 = 1$ and the bound is

$$\eta_{11} \geq \frac{1}{\langle x_1 \rangle} \frac{2}{1 + \sqrt{1+N}}. \tag{50}$$

   Consider the large-signaling limit where $N \to +\infty$. Show that in this case,

$$\eta_{11} \approx \frac{1}{\langle x_1 \rangle} \frac{1}{\sqrt{N}}. \tag{51}$$

   Therefore, to suppress noise significantly, we need to start with a small enough λ_2, i.e. $\tau_2 \ll \tau_1$, and then use a large N to suppress noise.

   However, notice that the noise decreases very slowly with N. In fact, the decay of noise we are most familiar with comes from the law of large numbers, that as the number of signaling events N increase, the noise decreases as square root of N, i.e. $\mathrm{CV} \propto \frac{1}{\sqrt{N}}$, where CV is the coefficient of variation. However, since $\eta_{11} = \mathrm{CV}_1^2$, we see that in this case we obtain a quartic root noise decay, i.e. $\mathrm{CV}_1 \propto N^{1/4}$! This is significantly worse than a square root, and suggests that noise suppression would be very costly in biomolecular reactions.

   Intuitively, this comes from the fact that to suppress the noise of x_1, we needed to use another species x_2, which is noisy in itself, therefore introducing additional noise when x_2 senses x_1 via x_1 catalyzing

x_2's production, and when x_2 actuates on x_1 by inhibiting x_1's production. The noise propagation and enhancement within the loop therefore make noise suppression much harder.

In fact, [15] in 2010 took this fact into consideration and used information theoretic methods to analyze arbitrary non-stationary memory-containing suppression of noise, by replacing $R_1^+(x_2)$ with $u[\mathcal{I}_t(x_2)]$, so the production rate of x_1 can be regulated by an arbitrary non-anticipatory functional allowing dependence on all of the information $\mathcal{I}_t(x_2)$ of x_2's past trajectory. They derived the following bound:

$$\eta_{11} \geq \frac{1}{\langle x_1 \rangle} \frac{2}{1+\sqrt{1+N}}. \tag{52}$$

We see that this is the same bound as the one we derived, with a quartic-root bound when N is large! Therefore the observation we had from our case study above is generally true.

(This bound is in fact more general than presented here, as [15] also allowed R_1^-, R_2^+ and R_2^- to take more general forms, as long as R_1^- depends on x_1 only, R_2^- depends on x_2 only, and R_2^+ can be arbitrary forms of signaling channel. The complication is that in these more general cases, the signaling rate N is not as well-defined, and channel capacity, the measure of information transmission, needs to be used in the bound. So the only assumption is the system architecture: x_2 first senses x_1's concentration by x_1 influencing the production of x_2, and then x_2 actuating on x_1 by influencing x_1's production rate.)

4. Based on the above results, it may seem that noise suppression is indeed hopeless in biomolecular systems! The fundamental limit on noise suppression is really severe, and it is so general that all biological scenarios cannot escape it.

   However, although it feels natural and without loss of generality to consider the feedback architecture, or the sensing-and-then-actuate architecture, as in our above example and in [15], biomolecular reactions do not necessarily obey such architecture. For example, reactions of the form $X_1 \to X_2$, or $(x_1, x_2) \to (x_1 - 1, x_2 + 1)$ in net change notation, happens very often in biology. Here x_1 is transformed into x_2, such as via catalysis reactions. Also, the simultaneous production or degradation of x_1 and x_2 is also very common, such as $(x_1, x_2) \to (x_1 + 1, x_2 + 1)$. This happens when x_1 and x_2 are two genes in the same operon, for example, or they are the protein products from a cleavage reaction.

   Biomolecular systems containing reactions of the above form, where x_1 and x_2's production and degradation are coupled in one reaction, is termed a **coupled** reaction system, and used in [14] to show that the quartic-root noise suppression bound from [15] can be easily broken when coupling is utilized, as is done in nature.

   Consider the following coupled variant of the reaction network in Eqn (41)

$$\begin{aligned} x_1 &\xrightarrow{R_1^+(x_2)} x_1 + 1, \\ (x_1, x_2) &\xrightarrow{x_1/\tau_1} (x_1 - 1, x_2 + n), \\ x_2 &\xrightarrow{x_2/\tau_2} x_2 - 1. \end{aligned} \tag{53}$$

   Here the degradation of x_1 and the production of x_2 are coupled into one reaction transforming x_1 into n molecules of x_2. Note that $\alpha = \frac{n}{\tau_1}$ in this case, so $N = \alpha\tau_1 = n$.

   Use linear noise approximation to show that the fluctuation dissipation equation Eqn (32) is the following in this case:

$$\begin{bmatrix} H_{12}\eta_{12} - \eta_{11} + \frac{1}{\langle x_1 \rangle} & \frac{1}{\tau_1}H_{12}\eta_{22} + \frac{1}{\tau_2}\eta_{11} - (\frac{1}{\tau_1} + \frac{1}{\tau_2})\eta_{12} - \frac{1}{\tau_1}\frac{1}{2\langle x_2 \rangle} - \frac{1}{\tau_2}\frac{n}{2\langle x_1 \rangle} \\ & \eta_{12} - \eta_{22} + \frac{n+1}{2\langle x_2 \rangle} \end{bmatrix} = 0. \tag{54}$$

(Hint: $D_{22} = \frac{n+1}{\tau_2\langle x_2\rangle}$, $\left\langle s_{1|2}\right\rangle = -\frac{1}{2}$, $\left\langle s_{2|1}\right\rangle = -\frac{n}{2}$, so $D_{12} = -\frac{1}{\tau_1}\frac{1}{2\langle x_2\rangle} - \frac{1}{\tau_2}\frac{n}{2\langle x_1\rangle}$.)

Solve the above to obtain

$$\eta_{11} = \frac{1}{\langle x_1\rangle}\left(1 - \lambda_1\frac{h}{1+h}\left(1 - \frac{N}{2} - \frac{h}{2}\frac{N+1}{N}\right)\right), \tag{55}$$

where $h = |H_{12}|$, assuming $H_{12} < 0$.

Now optimize over h, and compare with the fundamental limit on noise suppression. Show that the noise can indeed get lower than the bound!

References

[1] J. Paulsson, "Models of stochastic gene expression," *Physics of Life Reviews*, vol. 2, no. 2, pp. 157–175, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1571064505000138

[2] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, "Regulation of noise in the expression of a single gene," *Nature Genetics*, vol. 31, no. 1, pp. 69–73, May 2002. [Online]. Available: https://doi.org/10.1038/ng869

[3] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, "Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells," *Science*, vol. 329, no. 5991, pp. 533–538, 2010. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1188308

[4] L. Cai, N. Friedman, and X. S. Xie, "Stochastic protein expression in individual cells at the single molecule level," *Nature*, vol. 440, no. 7082, pp. 358–362, Mar 2006. [Online]. Available: https://doi.org/10.1038/nature04599

[5] J. Paulsson, O. G. Berg, and M. Ehrenberg, "Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation," *Proceedings of the National Academy of Sciences*, vol. 97, no. 13, pp. 7148–7153, 2000. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.110057697

[6] J. Paulsson and M. Ehrenberg, "Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks," *Phys. Rev. Lett.*, vol. 84, pp. 5447–5450, Jun 2000. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.84.5447

[7] N. Friedman, L. Cai, and X. S. Xie, "Linking stochastic dynamics to population distribution: An analytical framework of gene expression," *Phys. Rev. Lett.*, vol. 97, p. 168302, Oct 2006. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.97.168302

[8] N. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. Amsterdam: Elsevier, 2007.

[9] S. Chong, C. Chen, H. Ge, and X. S. Xie, "Mechanism of transcriptional bursting in bacteria," *Cell*, vol. 158, no. 2, pp. 314–326, Jul 2014. [Online]. Available: https://doi.org/10.1016/j.cell.2014.05.038

[10] Z. Fang, A. Gupta, S. Kumar, and M. Khammash, "Advanced methods for gene network identification and noise decomposition from single-cell data," *Nature Communications*, vol. 15, no. 1, p. 4911, Jun 2024. [Online]. Available: https://doi.org/10.1038/s41467-024-49177-1

[11] J. Paulsson, "Summing up the noise in gene networks," *Nature*, vol. 427, no. 6973, pp. 415–418, Jan 2004. [Online]. Available: https://doi.org/10.1038/nature02257

[12] A. Hilfinger, T. M. Norman, and J. Paulsson, "Exploiting natural fluctuations to identify kinetic mechanisms in sparsely characterized systems," *Cell Systems*, vol. 4, no. 2, 2016.

[13] A. Hilfinger, T. M. Norman, G. Vinnicombe, and J. Paulsson, "Constraints on fluctuations in sparsely characterized biological systems," *Phys. Rev. Lett.*, vol. 116, p. 058101, Feb 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.116.058101

[14] F. Xiao, M. Fang, J. Yan, and J. C. Doyle, "Coupled reaction networks for noise suppression," in *2019 American Control Conference (ACC)*, 2019, pp. 1547–1554.

[15] I. Lestas, G. Vinnicombe, and J. Paulsson, "Fundamental limits on the suppression of molecular fluctuations," *Nature*, vol. 467, no. 7312, p. 174, 2010.