

Chapter 2

Numerical and Analytic Predictions for Correlations in Gene Expression Noise

In this chapter we develop a mathematical model for stochastic gene regulation that is motivated by previous work on noise in λ cI repressor [19]. Simulated time-series data are used to predict features of cross correlation functions and to understand the effect of active and inactive forms of regulation. We extend the simulation-based analysis by developing analytic solutions to arbitrary cross correlation functions using a linearized approximation of the mathematical model. The analytic framework lends insight into the origins of the cross correlation function shape and simplifies analysis. Two applications are demonstrated: (1) Sensitivity analysis of cross correlation function features reveals which system parameters are most significant. (2) The cross correlation function is calculated for a cascade of arbitrary length to demonstrate generality to larger scale networks. Finally we discuss some limitations of correlation-based analysis methods.

2.1 Mathematical Models with Noise

We analyzed a simple three-gene circuit, shown in Fig. 2.1. The protein A is a transcription factor that represses production of B . Proteins A and C are constitutively expressed, meaning they are produced at a constant level and are not under the control of other transcription factors. A deterministic model for this system can be written using a Hill function to describe repression [34, 35]:

$$\dot{A} = \alpha_A - \beta A \tag{2.1}$$

$$\dot{B} = \frac{\alpha_B}{1 + (A/K)^n} - \beta B \tag{2.2}$$

$$\dot{C} = \alpha_C - \beta C. \tag{2.3}$$

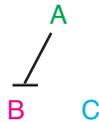


Figure 2.1: Three-gene circuit used in simulation. T-shaped arrow indicates repression of B by A .

In this model each protein is produced at a rate α_i and decays at a rate β ; the parameters K and n determine the properties of repression of B by A . The decay rate for all three proteins is assumed to be the same. This is a reasonable assumption if the proteins are stable (do not break down quickly). There are two ways that the level of protein in a cell can decrease: (1) As the cell grows in size the protein will dilute and (2) if the cell has an active mechanism for degrading proteins, such tagging them for recognition by proteases. Stable proteins are only governed by the first form of decay and thus their degradation is dependent upon the rate of cell division. By approximating dilution by a decaying exponential function, β can be calculated as

$$\beta = \frac{\log(2)}{T_{cc}}, \quad (2.4)$$

where T_{cc} is the length of the cell cycle.

The deterministic model has a single equilibrium point

$$A_{eq} = \frac{\alpha_A}{\beta} \quad (2.5)$$

$$B_{eq} = \frac{\alpha_B}{\beta(1 + (A_{eq}/K)^n)} \quad (2.6)$$

$$C_{eq} = \frac{\alpha_C}{\beta} \quad (2.7)$$

that is stable for all realistic biological parameters ($\alpha_i, \beta, K > 0$).

A more realistic model of gene expression accounts for noise in the expression of genes. Here we model the two classes of noise discussed in the Introduction. Extrinsic noise is assumed to affect all genes in the same way, while intrinsic noise is distinct for each individual gene. Thus, we add an extrinsic noise term, E , and intrinsic noise terms, I_i for $i = \{A, B, C\}$, to Eqns. (2.1)–(2.3):

$$\dot{A} = E + I_A + \alpha_A - \beta A \quad (2.8)$$

$$\dot{B} = E + I_B + \frac{\alpha_B}{1 + (A/K)^n} - \beta B \quad (2.9)$$

$$\dot{C} = E + I_C + \alpha_C - \beta C. \quad (2.10)$$

By setting the mean of these noise processes to zero we preserve the average equilibrium point of the system. We model noise as additive; other models, such as multiplicative noise, give similar qualitative results (not shown). Other properties of the noise sources are modeled explicitly using

biologically realistic parameters, described below.

Cellular noise sources have a finite correlation time that has been measured experimentally [19, 36]. We use Ornstein-Uhlenbeck processes to model noisy gene expression [37, 38]. These processes are described by the statistical values of mean, standard deviation, and correlation time, generating a noisy data trace that is continuous, allowing for numerical integration. In general, an Ornstein-Uhlenbeck process $X(t)$ can be written as

$$\frac{dX(t)}{dt} = -\frac{1}{\tau}X(t) + c^{1/2}\eta(t), \quad (2.11)$$

where τ is the correlation time of $X(t)$, c is a diffusion constant, and $\eta(t)$ is Gaussian white noise. As calculated in [38], as $t \rightarrow \infty$

$$\text{mean}\{X(t)\} = 0 \quad (2.12)$$

$$\text{var}\{X(t)\} = \frac{c\tau}{2}. \quad (2.13)$$

Rewriting Eqn. (2.11) in terms of the standard deviation of the noise, σ , and the correlation time, τ , we have

$$\dot{X} = -\frac{1}{\tau}X + \sqrt{\frac{2}{\tau}}\sigma\eta.$$

Extrinsic and intrinsic noise are modeled using Ornstein-Uhlenbeck processes by

$$\dot{E} = -\beta E + \theta\eta_E \quad (2.14)$$

$$\dot{I}_i = -\kappa I_i + \lambda_i\eta_i. \quad (2.15)$$

We assume the white noise terms η_E , η_A , η_B , and η_C are independent, identically distributed processes. The parameters β and κ define the time scale of the noise, while θ and λ set its standard deviation. These values were measured directly in [19]. Extrinsic noise was found to have a correlation time on the order of the cell cycle, while intrinsic noise had a much shorter characteristic time scale, $T_{int} = 5$ minutes. Thus β is described by Eqn. (2.4) (identical to the decay time of proteins) and $\kappa = \log(2)/T_{int}$. The standard deviation of extrinsic noise was measured directly for the λ cI system in [19] as $\sigma_{ext} = 0.35$. We assume the standard deviation of the intrinsic noise is related to the signal strength by setting $\sigma_{int,i} = \sqrt{\alpha_i}$, as in a Poisson process.

2.2 Simulation Results

We simulated the noisy expression of proteins A , B , and C numerically using the differential equations given in Eqns. (2.8)–(2.10) and (2.14)–(2.15) with the parameters listed in Table 2.1.

Parameter	Value	Notes/Reference
α_A	1.39 molecules/cell/min	chosen so that $\alpha_A/\beta = K$
α_B	4.5 molecules/cell/min	arbitrary
α_C	1.39 molecules/cell/min	chosen to match α_A
β	0.0116 1/min	calculated from Eqn. (2.4) assuming 60 min cell cycle
K	120 nM	[19]
n	1.7	[19]
κ	0.139 1/min	[19]
θ	0.0532 (molecules/cell) ^{1/2} /min	from Eqn. (2.13) and [19] $\sqrt{\frac{2}{T_{cc}}}\sigma_{ext}$
λ_A	0.621 (molecules/cell) ^{1/2} /min	from Eqn. (2.13) $\sqrt{\frac{2\alpha_A}{T_{int}}}$
λ_B	1.12 (molecules/cell) ^{1/2} /min	from Eqn. (2.13) $\sqrt{\frac{2\alpha_B}{T_{int}}}$
λ_C	0.621 (molecules/cell) ^{1/2} /min	from Eqn. (2.13) $\sqrt{\frac{2\alpha_C}{T_{int}}}$
T_{cc}	60 mins	measured from experiments
T_{int}	5 mins	[19]

Table 2.1: Simulation Parameters

Sample simulation traces are shown in Fig. 2.2. The equilibrium point calculated from the deterministic model (Eqns. (2.5)–(2.7)) is an accurate description of the average behavior. The different time scales associated with extrinsic and intrinsic noise are apparent when considering limiting cases where only one source of noise exists, as seen in Fig. 2.3. With only extrinsic noise the three signals are positively correlated, though expression of B is also repressed by A . The cell cycle length governs the time scale of fluctuations. With only intrinsic noise, repression of B by A generates a delayed anti-correlation in the expression of these two genes. The intrinsic noise time scale is much faster than that due to extrinsic noise.

2.3 Cross Correlation Functions

The cross correlation between two signals $f(t)$ and $g(t)$ is $R_{f,g}$, defined as

$$\begin{aligned}
 S_{f,g}(\tau) &= \begin{cases} \frac{1}{N-|\tau|} \sum_{n=0}^{N-\tau-1} \tilde{f}(n+\tau)\tilde{g}(n) & \tau \geq 0 \\ S_{g,f}(-\tau) & \tau < 0 \end{cases} \\
 R_{f,g} &= \frac{S_{f,g}(\tau)}{\sqrt{S_{f,f}(0)S_{g,g}(0)}}, \tag{2.16}
 \end{aligned}$$

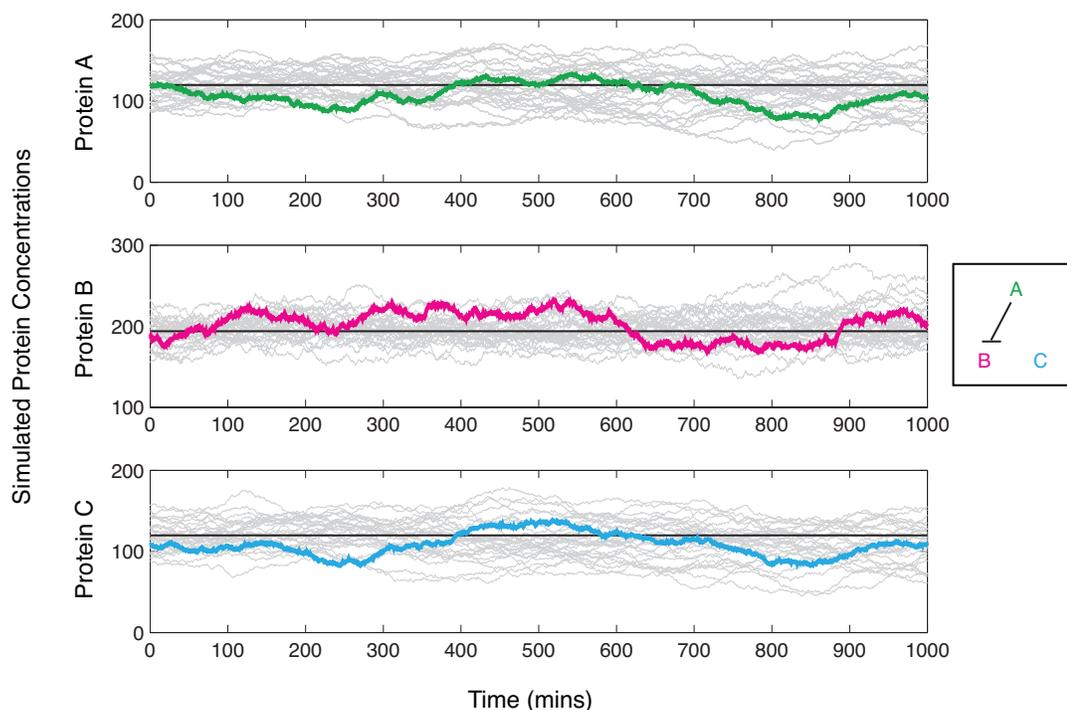


Figure 2.2: Simulated protein concentrations. Gray traces show 30 numerical examples of noisy gene expression. Black line indicates the steady-state equilibrium point. Individual traces from a single simulation are colored to show representative data.

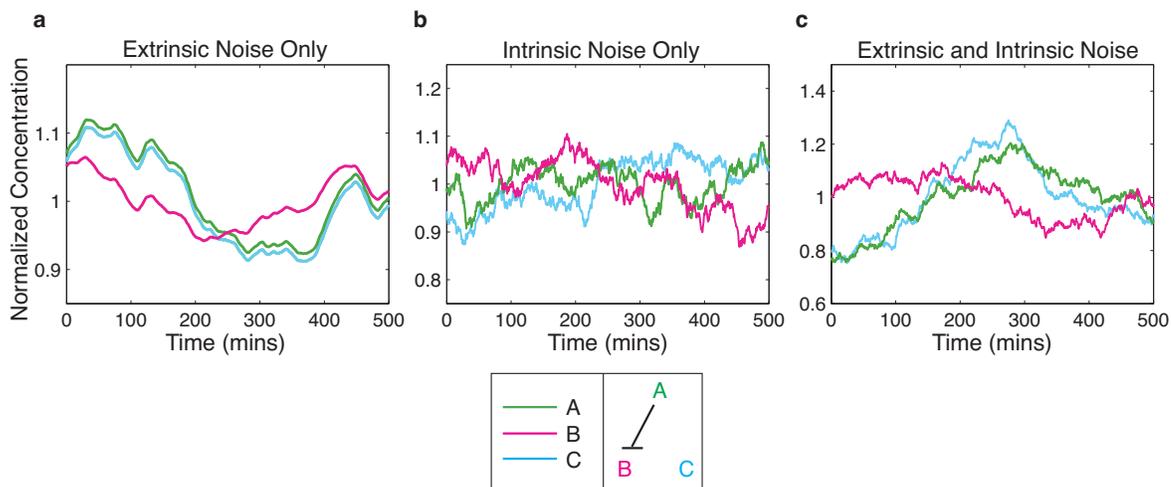


Figure 2.3: Sample simulation data under different types of noise. Protein A (green line) represses production of protein B (magenta line). Protein C (cyan line) is expressed constitutively. Data are normalized by mean concentration. Note that extrinsic noise positively correlates the time traces, while fluctuations in A (green curve) produce opposite fluctuations in B (magenta) at a delay (τ_{reg}). Simulated time traces are shown for three noise regimes, as indicated.

where N is the number of time points, τ is the time shift, and

$$\tilde{f} = f - \frac{1}{N} \sum_{n=0}^{N-1} f(n).$$

The mean-subtracted version of the cross correlation function, is sometimes referred to as the cross covariance.

Temporal correlations were used to measure the propagation of noise through a network. Individual cross correlation curves were generated by calculating results for two time series of noisy gene expression; the mean values of many individual cross correlation functions describe the average behavior (Fig. 2.4). Although the expected value of the cross correlation function due to noise is a fairly simple curve, obtaining it requires generating statistics over many sets of time-series data.

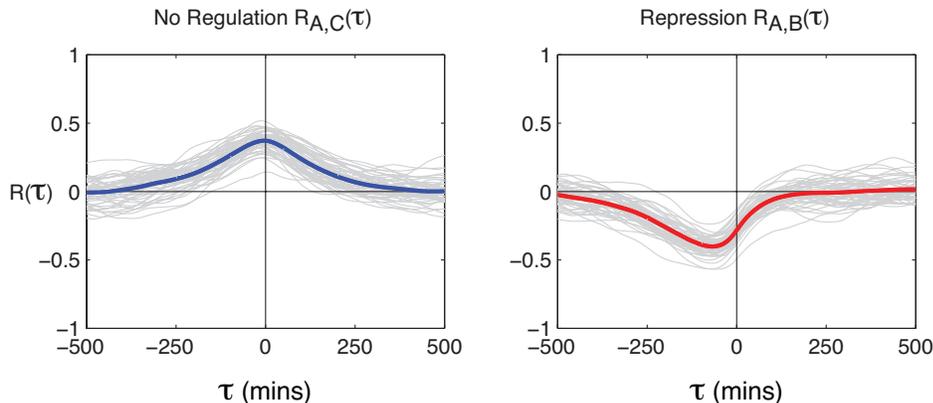


Figure 2.4: Cross correlation functions of simulated gene expression data. Gray curves show 50 cross correlation curves from individual time-series data. Mean values are shown in blue (no regulation) and red (repression).

Mean cross correlation functions for different noise regimes are shown in Fig. 2.5. Several features are apparent: (1) Repression appears as a dip at a delay time denoted by τ_{reg} , the effective regulation time. (2) The direction of regulation is given by the sign of τ_{reg} . Since A represses B, the dip occurs at $\tau_{reg} < 0$. (3) Extrinsic noise causes a positive peak in the cross correlation function close to $\tau = 0$, both with and without regulation. (4) The relative balance of intrinsic and extrinsic noise affects the magnitude of τ_{reg} . Together, these results indicate that cross correlation analysis, in combination with an understanding of physiological levels and types of noise, can be used to analyze the activity and direction of regulatory links.

Simulations were used to explore the effect of network activity on the shape of the cross correlation function. In Fig. 2.6 we varied the ratio of A_{eq} to K , which sets the position of the input on the sigmoidal Hill function curve. The dip is largest when A is in the region of the Hill function with the steepest slope (red dots). As A_{eq} moves to the saturating regions on the Hill function the magnitude

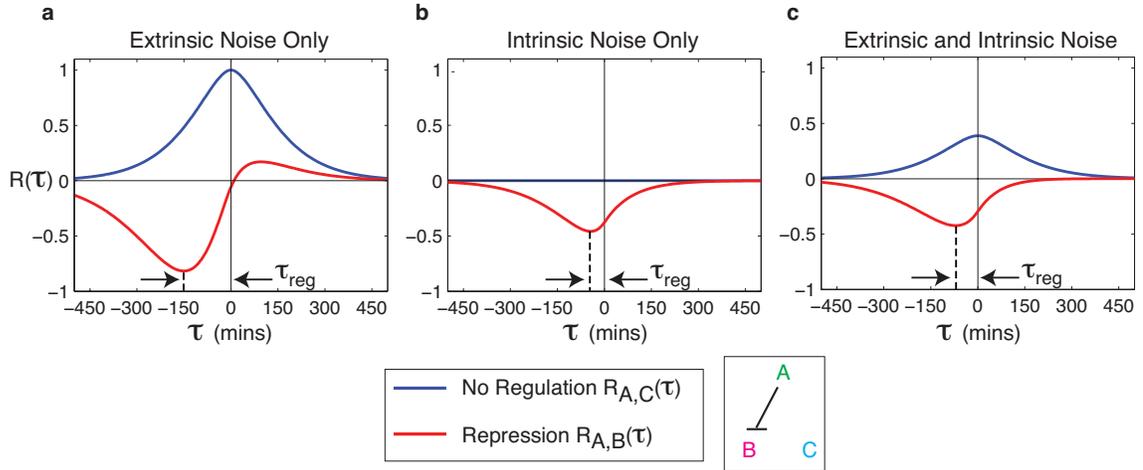


Figure 2.5: Mean cross correlation functions $R_{A,B}(\tau)$ and $R_{A,C}(\tau)$ are shown in red and blue, respectively. Note that active negative regulation causes a dip at τ_{reg} , while extrinsic noise results in positive correlation near $\tau = 0$.

of the dip in the cross correlation returns to zero (blue dots).

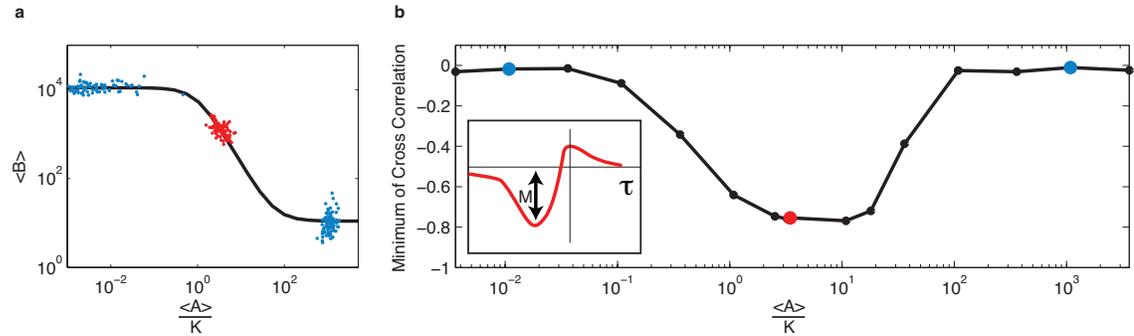


Figure 2.6: Cross correlation function shape depends on the activity of the network. (a) Ratio of the mean value of A (A_{eq}) to K versus the mean value of B . On average the shape of the curve is a Hill function. Several regions of this curve are explored in simulation with colored dots corresponding to data in (b). (b) Magnitude of the dip in the cross correlation function due to repression. As the mean value of A moves through regions on the Hill function with non-zero slope, the cross correlation function exhibits a characteristic dip due to repression. Inset shows schematic of dip magnitude.

The shape of the cross correlation function is highly dependent upon the activity of the regulatory link. We explore the dependence of its features on system parameters further in Section 2.5. To summarize: We found that τ_{reg} is most sensitive to the cell cycle time, with longer cell cycles producing a longer τ_{reg} . The magnitude of the dip due to regulation, in contrast, is determined primarily by the slope of the regulation function—how switch-like the regulation is.

Simulations can be used to predict the shape of cross correlation functions for other network architectures. A transcriptional activator is shown in Fig. 2.7 under different noise environments. Regulation appears as a peak in the cross correlation at $\tau < 0$. With only intrinsic noise, this peak

is a mirror image of the dip in the cross correlation function due to repression. With extrinsic noise, this peak is combined with the positive, symmetric peak that is caused by global noise sources.

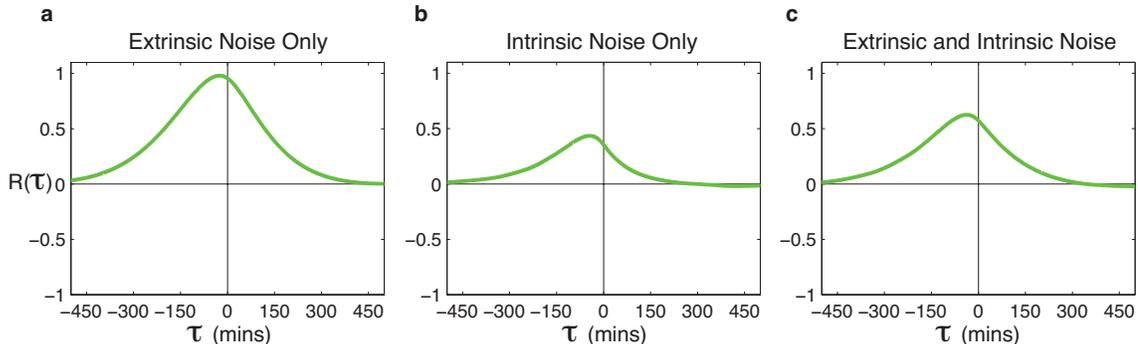


Figure 2.7: Cross correlation functions for an activator with (a) extrinsic noise only, (b) intrinsic noise only, and (c) both noise sources. Parameters are identical to those used in the prior repression simulations except $n = -1.7$.

2.4 Analytic Solutions for Cross Correlation Functions Due to Noise

In this section we develop an analytic method for calculating arbitrary cross correlation functions. This work was conducted in collaboration with Joe Levine.

We consider the stochastic differential equations in Eqns. (2.8)–(2.10) and (2.14)–(2.15). Assuming perturbations due to noise are small, we linearize the system about the equilibrium point given in Eqns. (2.5)–(2.7). Defining $a = A - A_{eq}$, $b = B - B_{eq}$, and $c = C - C_{eq}$ we obtain the following set of linear dynamics

$$\dot{a} = E + I_A - \beta a \quad (2.17)$$

$$\dot{b} = ga + E + I_B - \beta b \quad (2.18)$$

$$\dot{c} = E + I_C - \beta c \quad (2.19)$$

where g is the *local sensitivity*

$$g = -\frac{\alpha_B n \left(\frac{\alpha_A}{\beta K}\right)^{n-1}}{K \left(1 + \left(\frac{\alpha_A}{\beta K}\right)^n\right)^2}.$$

If we assume that the mean value of A is in the center of the Hill function so $A_{eq} = K$, as is the case with our simulated system, then this simplifies to $g = -\frac{\alpha_B n}{4K}$. This analysis is still possible regardless of the steady state value of A , but certain regimes (the middle of the Hill function, saturated edges of the nonlinearity) are better approximated by linear models. Note that the constant g is the only place that information about the nonlinearity enters the equations.

The cross correlation theorem states that cross correlation in the time domain is equal to multiplication in the frequency domain

$$R_{f,g}(\tau) = F^{-1}[\tilde{f}^* \tilde{g}],$$

where * denotes the complex conjugate and $f(t)$ and $g(t)$ are the two signals. However, we average over many cross correlation functions (as seen in Fig. 2.4), so we need to calculate the expected value of the cross correlation function over many realizations of the noise

$$E\{R_{f,g}(\tau)\} = E\{F^{-1}[\tilde{f}^* \tilde{g}]\}. \quad (2.20)$$

Taking the Fourier transform of Eqns. (2.17)–(2.19) we find

$$\tilde{a} = \frac{1}{\beta + i\omega}(\tilde{E} + \tilde{I}_A) \quad (2.21)$$

$$\tilde{b} = \frac{1}{\beta + i\omega}(\tilde{E} + \tilde{I}_B + g\tilde{a}) \quad (2.22)$$

$$\tilde{c} = \frac{1}{\beta + i\omega}(\tilde{E} + \tilde{I}_C) \quad (2.23)$$

$$\tilde{E} = \frac{\theta}{\beta + i\omega} \tilde{\eta}_E \quad (2.24)$$

$$\tilde{I}_i = \frac{\lambda_i}{\kappa + i\omega} \tilde{\eta}_i. \quad (2.25)$$

Below, we calculate cross correlation expressions for two cases: two independent genes (A and C) and a simple regulatory link with repression (A and B). The first, and simpler, case is worked through in detail, while the results of the second case are summarized.

2.4.1 Unregulated Case

We substitute Eqns. (2.21)–(2.25) into Eqn. (2.20), dropping tildes to simplify notation

$$\begin{aligned} E\{R_{a,c}(\tau)\} &= E\{F^{-1}\left[\frac{1}{\beta - i\omega}\left(\frac{\theta}{\beta - i\omega}\eta_E^* + \frac{\lambda_A}{\kappa - i\omega}\eta_a^*\right)\frac{1}{\beta + i\omega}\left(\frac{\theta}{\beta + i\omega}\eta_E + \frac{\lambda_C}{\kappa + i\omega}\eta_c\right)\right]\} \\ &= E\{F^{-1}\left[\frac{1}{\beta^2 + \omega^2}\left(\frac{\theta^2}{\beta^2 + \omega^2}\eta_E^*\eta_E + \frac{\theta\lambda_C}{(\beta - i\omega)(\kappa + i\omega)}\eta_E^*\eta_c\right.\right. \\ &\quad \left.\left. + \frac{\theta\lambda_A}{(\beta + i\omega)(\kappa - i\omega)}\eta_a^*\eta_E + \frac{\lambda_A\lambda_C}{\beta^2 + \omega^2}\eta_a^*\eta_c\right)\right]\}. \end{aligned}$$

Because the Fourier transform is a linear operation we can analyze each of the four terms individually.

We use two features of white noise to simplify analysis. First, white noise has a flat power spectral

density $\eta_i^*(\omega)\eta_i(\omega) = W_i$ and second $E\{\eta_i(t)\eta_j(t)\} = 0$ for $i \neq j$. Thus,

$$\begin{aligned} E\{\eta_i(t)\eta_j(t)\}_{i \neq j} &= E\{F^{-1}[F[\eta_i(t)\eta_j(t)]]\} \\ &= E\{F^{-1}[\eta_i^*(\omega)\eta_j(\omega)]\} \\ &= 0. \end{aligned}$$

Therefore if we have a deterministic function $G(\omega)$

$$\begin{aligned} E\{F^{-1}[G(\omega)\eta_i^*(\omega)\eta_j(\omega)]\}_{i \neq j} &= E\left\{\frac{1}{\sqrt{2\pi}}F^{-1}[G(\omega)] \star F^{-1}[\eta_i^*(\omega)\eta_j(\omega)]\right\} \\ &= \frac{1}{\sqrt{2\pi}}F^{-1}[G(\omega)] \star E\{F^{-1}[\eta_i^*(\omega)\eta_j(\omega)]\} \\ &= 0, \end{aligned}$$

where \star represents convolution. Due to these white noise properties the last three terms in the cross correlation expression become zero and the remaining term simplifies to

$$E\{R_{a,c}(\tau)\} = F^{-1}\left[\frac{1}{\beta^2 + \omega^2} \frac{\theta^2}{\beta^2 + \omega^2} W_E\right].$$

Applying the inverse Fourier transform we find

$$E\{R_{a,c}(\tau)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{\beta^2 + \omega^2} \frac{\theta^2}{\beta^2 + \omega^2} W_E e^{-i\omega\tau} d\omega.$$

This integral can be solved using Cauchy's Residue theorem. Specifically, we need to consider two cases: $\tau < 0$ and $\tau \geq 0$. In the first case we can apply Jordan's lemma if we use a contour that encircles the upper half plane (Fig. 2.8a). Using Cauchy's Residue theorem we find

$$\lim_{R \rightarrow \infty} \int_{C_R} f(z) dz + \int_{-R}^R f(z) dz = 2\pi i \sum \text{Res.}$$

By Jordan's lemma the contour at infinity, C_R , becomes zero and we are left with the integral we want to evaluate. In our integral we have a second-order pole at $z = i\beta$ and a second-order pole at $z = -i\beta$. Since we are closing the contour in the upper half plane we need to evaluate the residue at $z = i\beta$:

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{\beta^2 + \omega^2} \frac{\theta^2}{\beta^2 + \omega^2} W_E e^{-i\omega\tau} d\omega &= \frac{1}{2\pi} 2\pi i \text{Res}(i\beta) \\ &= \frac{\theta^2 W_E}{4\beta^3} e^{\beta\tau} (1 - \beta\tau). \end{aligned}$$

For $\tau \geq 0$ we can use Jordan's lemma if we choose a contour that encircles the lower half plane

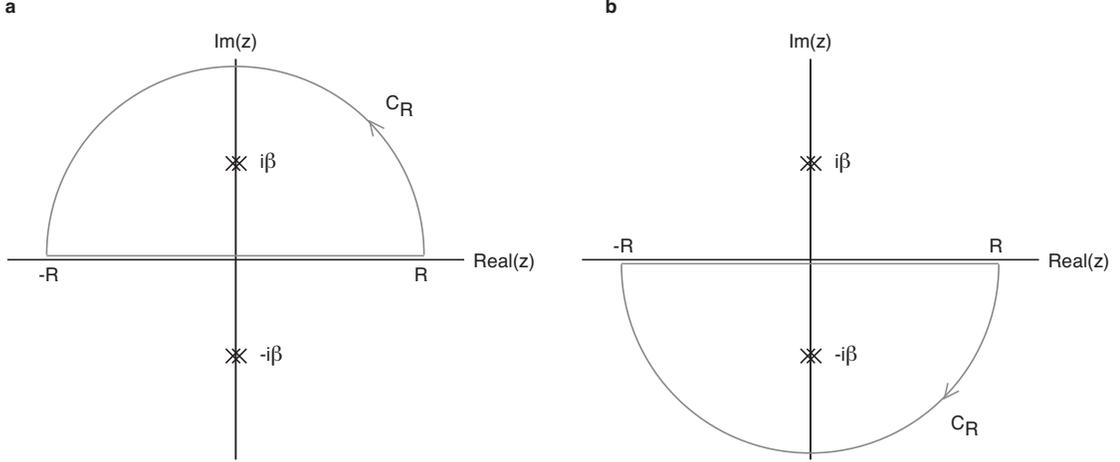


Figure 2.8: Contours used for evaluating Cauchy's Residue theorem. Pole locations shown are for the unregulated cross correlation expression.

(Fig. 2.8b). Since the direction of encirclement is now clockwise, the residue theorem has an additional negative sign:

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{\beta^2 + \omega^2} \frac{\theta^2}{\beta^2 + \omega^2} W_E e^{-i\omega\tau} d\omega &= \frac{1}{2\pi} (-2\pi i \text{Res}(-i\beta)) \\ &= \frac{\theta^2 W_E}{4\beta^3} e^{-\beta\tau} (1 + \beta\tau). \end{aligned}$$

Combining these two results we find

$$E\{R_{a,c}(\tau)\} = \frac{\theta^2 W_E}{4\beta^3} e^{-\beta|\tau|} (1 + \beta|\tau|).$$

Note that this expression for the cross correlation used a and c , the mean subtracted versions of A and C . These expressions are consistent with those calculated directly from the nonlinear simulations because we used the mean subtracted version of the cross correlation function (Eqn. (2.16)).

2.4.2 Regulated Case

Analysis of the cross correlation function due to repression is similar. After simplification using the white noise properties discussed in the previous section we find

$$\begin{aligned} E\{R_{a,b}(\tau)\} &= F^{-1}\left[\frac{\theta^2 W_E}{(\beta + i\omega)^2(\beta - i\omega)^2}\right] \\ &+ F^{-1}\left[\frac{g \theta^2 W_E}{(\beta + i\omega)^3(\beta - i\omega)^2}\right] \\ &+ F^{-1}\left[\frac{g \lambda_A^2 W_A}{(\beta + i\omega)^2(\beta - i\omega)(\kappa + i\omega)(\kappa - i\omega)}\right]. \end{aligned}$$

These three terms have a convenient interpretation. The first term is the artificial correlation due to extrinsic noise, the second term is extrinsic noise that has propagated through the link, the third term is intrinsic noise that has propagated through the link. Cauchy's Residue theorem and Jordan's lemma are applied to find

$$R_{a,b}(\tau) = \begin{cases} \frac{\theta^2 W_E}{4\beta^3} e^{\beta\tau} (1 - \beta\tau) \\ + \frac{g\theta^2 W_E}{16\beta^4} e^{\beta\tau} (3 - 4\beta\tau + 2\beta^2\tau^2) \\ + \lambda_A^2 g W_A \left(e^{\beta\tau} \frac{\kappa^2(1-2\beta\tau) - \beta^2(5-2\beta\tau)}{4\beta^2(\beta^2 - \kappa^2)^2} + e^{\kappa\tau} \frac{1}{2\kappa(\beta - \kappa)^2(\beta + \kappa)} \right) & \tau < 0 \\ \frac{\theta^2 W_E}{4\beta^3} e^{-\beta\tau} (1 + \beta\tau) \\ + \frac{g\theta^2 W_E}{16\beta^4} e^{-\beta\tau} (3 + 2\beta\tau) \\ + \lambda_A^2 g W_A \left(\frac{e^{-\beta\tau}}{4\beta^2(\kappa^2 - \beta^2)} + \frac{e^{-\kappa\tau}}{2\kappa(\beta + \kappa)^2(\beta - \kappa)} \right) & \tau \geq 0. \end{cases}$$

2.4.3 Summary

The cross correlation relations are summarized as

$$E\{R_{A,C}(\tau)\} = N_{A,C} \frac{\theta^2 W_E}{4\beta^3} e^{-\beta|\tau|} (1 + \beta|\tau|)$$

$$E\{R_{A,B}(\tau)\} = \begin{cases} N_{A,B} \left(\frac{\theta^2 W_E}{16\beta^4} e^{\beta\tau} (2g\beta^2\tau^2 - 4\beta(g + \beta)\tau + 3g + 4\beta) \right. \\ \left. + \lambda_A^2 g W_A \left(e^{\beta\tau} \frac{\kappa^2(1-2\beta\tau) - \beta^2(5-2\beta\tau)}{4\beta^2(\kappa^2 - \beta^2)^2} + e^{\kappa\tau} \frac{1}{2\kappa(\beta - \kappa)^2(\beta + \kappa)} \right) \right) & \tau < 0 \\ N_{A,B} \left(\frac{\theta^2 W_E}{16\beta^4} e^{-\beta\tau} (2\beta(g + 2\beta)\tau + 3g + 4\beta) \right. \\ \left. + \lambda_A^2 g W_A \left(e^{-\beta\tau} \frac{1}{4\beta^2(\kappa^2 - \beta^2)} + e^{-\kappa\tau} \frac{1}{2\kappa(\beta + \kappa)^2(\beta - \kappa)} \right) \right) & \tau \geq 0 \end{cases}$$

where the normalization factors are

$$N_{A,B} = \frac{1}{\sqrt{R_{A,A}(0)R_{B,B}(0)}}$$

$$N_{A,C} = \frac{1}{\sqrt{R_{A,A}(0)R_{C,C}(0)}}$$

$$R_{A,A}(0) = \frac{\theta^2 W_E}{4\beta^3} + \frac{\lambda_A^2 W_A}{2\beta\kappa(\kappa + \beta)}$$

$$R_{B,B}(0) = \frac{\theta^2(3g^2 + 6g\beta + 4\beta^2) W_E}{16\beta^5} + \frac{\lambda_A^2 g^2(\kappa + 2\beta) W_A}{4\kappa\beta^3(\kappa + \beta)^2} + \frac{\lambda_B^2 W_B}{2\kappa\beta(\kappa + \beta)}$$

$$R_{C,C}(0) = \frac{\theta^2 W_E}{4\beta^3} + \frac{\lambda_C^2 W_C}{2\beta\kappa(\kappa + \beta)}.$$

To compare these results to the nonlinear simulation we use the constants specified in Table 2.1, where $g = -\frac{\alpha_B n}{4K}$. W_E, W_i for $i = \{A, B, C\}$ are treated as binary variables and set to 0 or 1 to turn off and on extrinsic and intrinsic noise for comparison to Fig. 2.5. We assume $W_A = W_B = W_C$

for simplicity. Fig. 2.9 shows that the analytic solutions for cross correlations match the simulated nonlinear system extremely well.

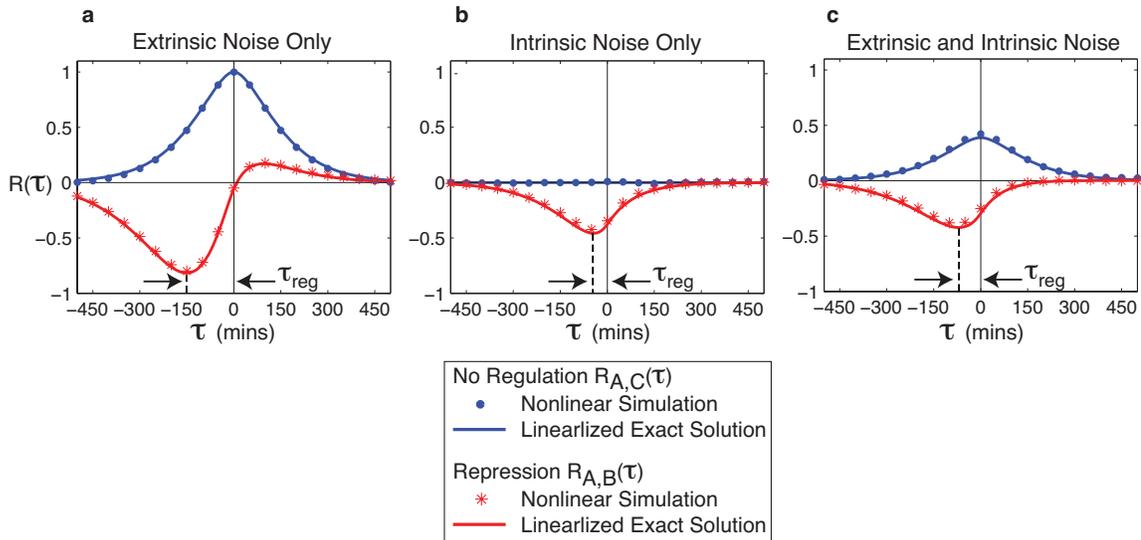


Figure 2.9: Comparison of analytically calculated cross correlation expressions and simulated data. Dots represent simulated data, while solid lines plot analytic solutions for the linearized model.

We have calculated the cross correlation functions for two types of regulation. This method can be applied more generally to larger networks, provided perturbations due to noise are small enough that linearization is a valid approximation.

2.5 Sensitivity Analysis

Two prominent features of the cross correlation curve for repression are the location of the dip, τ_{reg} , and the magnitude of the dip, M (shown schematically in Fig. 2.10). We calculate how sensitive these features are to variations in the system parameters. These results indicate which parameters play a primary role in setting the features of the cross correlation function.

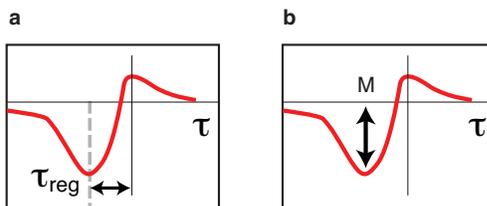


Figure 2.10: Schematic of cross correlation function features (a) τ_{reg} and (b) M .

To find τ_{reg} for repression we take

$$\left. \frac{dR_{a,b}(\tau)}{d\tau} \right|_{\tau=\tau_{reg}} = 0.$$

In general it is not possible to find a closed form solution for τ_{reg} , however numerical root finding methods can be applied. The nominal parameter values for sensitivity analysis are the same as those used in the simulation. For a feature of the cross correlation function, y , we find the normalized sensitivity

$$S_i = \frac{y(p_i + \Delta p_i) - y(p_i - \Delta p_i)}{2 \Delta p_i y(p_i)}.$$

For $\Delta p_i = 0.05 p_i$ the parameteric sensitivities are shown in Table 2.2. Large values indicate that the feature y is very sensitive to that parameter. The sign of the sensitivity indicates whether y will get larger ($S_i > 0$) or smaller ($S_i < 0$) as the parameter is increased.

Parameter	τ_{reg}	M
β	-108.5	2.39
g	18.9	-85.7
θ	11.3	5.5
κ	4.0	-12.4
λ_A	-0.9	2.5
λ_B	0.0	0.0
λ_C	0.0	0.0

Table 2.2: Normalized Sensitivities

τ_{reg} is most sensitive to the parameter β , which sets the time scale of both protein decay and extrinsic noise. As the cell cycle ($\log(2)/\beta$) gets longer, the location of the dip moves further away from zero. M is most sensitive to the local sensitivity g , which is negative for repression. As g becomes less negative, the repressor has less of an effect on its target and the dip gets smaller. In the extreme case when $g = 0$, which indicates an inactive or non-existent regulatory connection, the dip disappears.

2.6 Transcriptional Cascades

It is interesting to ask what the limits are to using these cross correlation functions. For two genetic components that are very far away, will the correlation eventually average to zero or will measurement be the limiting factor (it may be prohibitive to measure noisy signals in single cells for long periods of time)? To explore this question we derived an expression for the cross correlation function of a cascade of arbitrary length. The signals between the first and last element in the cascade are compared. This type of question was asked in [39] where they built a synthetic cascade to measure noise, and in [33] with a numerical study.



Figure 2.11: Cascade of length n . Interactions alternate between repressors (T-arrows) and activators (normal arrows) to preserve the net repression effect.

The cascade shown in Fig. 2.11 is described by a system of linear equations:

$$\begin{aligned} \dot{x}_1 &= E + I_1 - \beta x_1 \\ \dot{x}_2 &= E + I_2 + g_1 x_1 - \beta x_2 \\ \dot{x}_3 &= E + I_3 + g_2 x_2 - \beta x_3 \\ &\dots \\ \dot{x}_n &= E + I_n + g_{n-1} x_{n-1} - \beta x_n. \end{aligned}$$

The Fourier transforms used in the cross correlation calculation are

$$\begin{aligned} \tilde{x}_1 &= \frac{1}{\beta + i\omega} (\tilde{E} + \tilde{I}_1) \\ \tilde{x}_n &= \sum_{k=1}^n \frac{1}{(\beta + i\omega)^{n-k+1}} \left(\prod_{j=k}^n g_j \right) (\tilde{E} + \tilde{I}_k) \end{aligned}$$

where $g_n = 1$. The full cross correlation expressions were calculated using an automated Mathematica script, given in Appendix A. In general, the `Residue` command in Mathematica is an efficient way to calculate cross correlation functions.

Cross correlations from cascades with between two and eight elements are compared in Fig. 2.12. Even for long chains of transcription factors we still see non-negligible correlation values. Thus, it is likely that experimental limitations—such as the length of a movie—will restrict ability to see distant temporal events. In addition, real genetic networks will have other inputs to intermediate elements that may confound analysis.

2.7 mRNA dynamics

To this point we have only considered protein dynamics and have not considered regulation at the mRNA level. Using analytic methods we can show that mRNA dynamics, and other fast time-scale processes, can be neglected to a first approximation.

The systems we have considered so far are of the form

$$\dot{p}_B = -\beta p_B + g_p p_A + n,$$

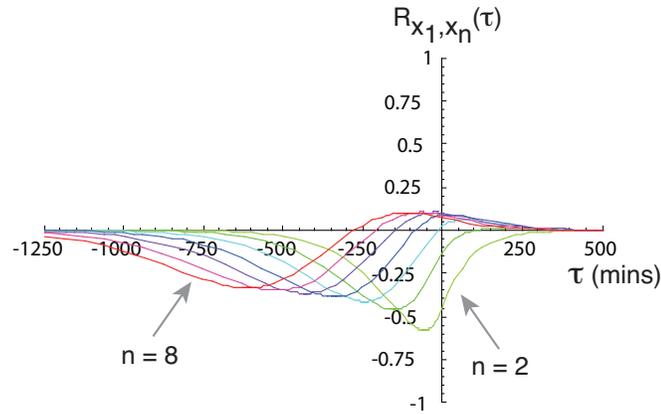


Figure 2.12: Cross correlation functions for different length cascades. Cross correlations are between the signals x_1 and x_n for $n = 2, 3, \dots, 8$. $g = 0.02$ where $g_n = 1$ and $g_i = -g$ for odd i and $g_i = g$ for even i .

where p_A and p_B are the input and output protein concentrations, respectively, g_p is the local sensitivity, and n is a grouped noise term. The Fourier transform of p_B is

$$\tilde{p}_B = \frac{1}{\beta + i\omega} (g_p \tilde{p}_A + \tilde{n}).$$

Extending this model to include mRNA dynamics (as in [23], using a linearized form of the equations), we have

$$\begin{aligned} \dot{m}_B &= -\beta_m m_B + g_m p_A + n_m \\ \dot{p}_B &= -\beta(p_B - m_B) + n_p, \end{aligned}$$

where m_B is the mRNA concentration and n_m and n_p are noise in the mRNA and protein production processes. Taking the Fourier transforms we find

$$\tilde{p}_B = \frac{1}{\beta + i\omega} \left(\frac{\beta g_m}{\beta_m + i\omega} \tilde{p}_A + \tilde{n} \right),$$

where we have grouped all the noise terms into n .

Comparing these two equations, as long as

$$g_p \approx \frac{\beta g_m}{\beta_m + i\omega}$$

we can ignore the mRNA dynamics. If both models have the same steady-state behavior ($i\omega = 0$) then $g_p = \beta g_m / \beta_m$. The protein lifetime is $\beta = 0.0116$ 1/min, based on a 60 minute cell cycle time. mRNA lifetimes, in contrast, are on the order of 2 minutes, or $\beta_m = 0.3466$ 1/min [40]. Thus, the protein dynamics are 30 times slower than the mRNA dynamics and are thus expected to be

dominant in determining the system dynamics. As long as the mRNA dynamics are significantly faster than the protein dynamics it is reasonable to approximate the system using only the protein model.

2.8 Degenerate Cross Correlation Functions

There is not a unique relationship between the shape of a cross correlation function and the network architecture that generated it. Here, we explore two classes of networks that result in degenerate cross correlation functions. In these situations, two cross correlations look similar or identical even though they are the result of different network architectures.

2.8.1 Redundant Network Elements

Consider the network diagram shown in Fig. 2.13, assuming that network connection and expression properties are identical for the links A - B and A - C . Proteins B and C are redundant network elements because they are controlled by the same input, A , and extrinsic noise affects them in the same way. Thus, in the extreme case where only extrinsic noise is present, $B(t)$ and $C(t)$ will be identical. Consequently, the cross correlation function $R_{B,D}(\tau) = R_{C,D}(\tau)$.

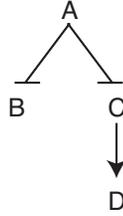


Figure 2.13: Network with redundant components

Intrinsic noise helps to discriminate between $R_{B,D}(\tau)$ and $R_{C,D}(\tau)$ because intrinsic noise in C propagates to D causing additional time-lagged correlation. Intrinsic noise in B , because it is uncorrelated with noise in D , will not have the same effect.

Analytic solutions for both cross correlation functions are summarized by

$$\begin{aligned}
 R_{B,D}(\tau) &= E\{F^{-1}[f_{Be}^* f_{De} + f_{BA}^* f_{DA}]\} \\
 R_{C,D}(\tau) &= E\{F^{-1}[f_{Ce}^* f_{De} + f_{CA}^* f_{DA} + f_{CC}^* f_{DC}]\},
 \end{aligned}$$

where $f_{i,j}$ is the Fourier transform of the differential equations describing the dynamics of protein j in response to noise from i , and e is extrinsic noise. Assuming the network connections are identical

for A - B and A - C we find the difference between the two cross correlation functions

$$R_{C,D}(\tau) - R_{B,D}(\tau) = E\{F^{-1}[f_{CC}^* f_{DC}]\}.$$

This term represents how noise in C propagates to D , but is only non-zero if there is intrinsic noise (or other inputs) affecting C .

Thus, mathematically $R_{C,D}(\tau) \neq R_{B,D}(\tau)$ unless there is no intrinsic noise in C . In practice it may be difficult to distinguish between the two cross correlation functions even if they are not identical.

This simple example illustrates how redundant network elements can confound analysis. More complicated networks will have similar problems any time there are two or more network elements that are controlled by the same inputs. Intrinsic noise or other signals that affect individual genes can help to distinguish correlations due to regulation from correlations due to redundant network elements.

2.8.2 Parametric Degeneracies

A second class of degeneracies comes from uncertainty in parameters. For example, Fig. 2.14 shows two cascades. If we measure $R_{A,C}(\tau)$ is it possible to determine that there is a middle element in the network or will the cross correlation function look like $R_{X,Y}(\tau)$? In this example extrinsic noise

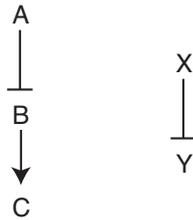


Figure 2.14: Example of a network showing parametric degeneracy

is helpful in discriminating between the two cascades. In the two-step cascade, extrinsic noise affects A , B , and C , and thus enters into the cross correlation function in three ways, while in the one-step cascade it only enters twice.

If we consider the limiting case where only intrinsic noise is present there are still ways adjust parameters to make the two cross correlation functions look nearly identical. For example, if B degrades quickly, but the net strength of the network is the same as in X - Y it is possible to choose parameters that result in very similar cross correlation functions (Fig. 2.15).

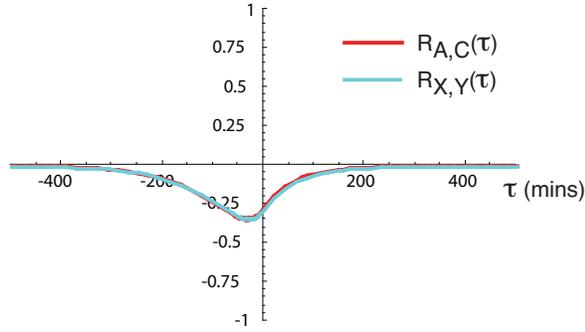


Figure 2.15: Cross correlation functions for two cascades with degenerate parameters. Simulation parameters are $\beta_x = \beta_y = \beta_a = \beta_c = 1/60 \text{ mins}^{-1}$, $\beta_b = 1/5 \text{ mins}^{-1}$, $g_{xy} = g_{ab}g_{bc} = -0.01$, $W_i = 1$.

In this example the intrinsic noise-only dynamics are described by

$$\begin{aligned}\dot{a} &= -\beta_a a + \eta_a \\ \dot{b} &= -\beta_b b + \eta_b + g_{ab} a \\ \dot{c} &= -\beta_c c + \eta_c + g_{bc} b \\ \\ \dot{x} &= -\beta_x x + \eta_x \\ \dot{y} &= -\beta_y y + \eta_y + g_{xy} x\end{aligned}$$

where intrinsic noise has been approximated by white noise and parameters are given in the figure caption.

Since biochemical parameters are often unknown or uncertain there are many possible situations where the shape of two cross correlation functions may look very similar.

2.8.3 Implications for Network Identification

Even though the cross correlation function does not uniquely determine the network architecture and parameters that generate it, it is still a useful tool. In particular, the cross correlation function can be used as a sensitive measure of network activity or it can suggest possible links in networks that are only partially mapped. Indeed, much of our knowledge about biochemical regulation comes from correlation-based reasoning and including temporal measurements extends these tools.