Cellular Systems Biology
and
Biological Network Analysis

Joel S. Bader
Department of Biomedical Engineering
Johns Hopkins University
(c) 2016

September 5, 2016

About the Author

Joel S. Bader, Ph.D., (joel.bader@jhu.edu, www.baderzone.org) is Professor of Biomedical Engineering at Johns Hopkins University, with a secondary appointment in Computer Science. He is Interim Director of the High Throughput Biology Center at the School of Medicine and a member of the Institute for Computational Medicine, the Institute for Genetic Medicine, and the Human Genetics Program. Prior to joining Johns Hopkins, Dr. Bader was employed by CuraGen Corporation (1995-2003) and is co-inventor of the Roche/454 Genome Sequencer. Dr. Bader has a Ph.D. in Theoretical Chemistry from U.C. Berkeley (1991), where he was an NSF Predoctoral Fellow, and performed post-doctoral research at Columbia University (1992-1995). Dr. Bader has a B.S. in Biochemistry from Lehigh University (1986, Phi Beta Kappa, Tau Beta Pi).

Research in the Bader lab focuses on the connection between genotype and phenotype, including human genetics, systems biology, and synthetic biology. The Bader lab has received funding from NIH, NSF CAREER, DOE, Microsoft, the Kleberg Foundation, and the Simons Foundation.

Preface

Cells are systems. Standard engineering and mathematics texts should provide an excellent introduction to understanding how cells behave, mapping inputs to outputs. Unfortunately, cells are not linear, time-independent systems. Saturation and cooperative response break linearity. Cellular states change with time. Cells are not even deterministic, violating the assumptions of non-linear systems analysis.

This book provides a self-contained introduction to cells as non-linear, time-dependent, stochastic, spatial systems. Each major section is motivated by a canonical biological pathway or phenomenon that requires the introduction of new concepts. All the required mathematical techniques are developed from the motivating examples.

The book is designed as a text for advanced undergraduate or graduate students. Prerequisites are univariate calculus, linear algebra, basic molecular biology, and rudimentary facility with a programming language for computational experiments. Linear systems and Laplace transforms are helpful, but are also reviewed in the initial chapters. Each chapter is designed to be covered in an hour lecture, and problems are provided in an Appendix.

This book is developed from course notes for "Systems Bioengineering III: Genes to Cells," taught by me since 2007 as a required course for our B.S. in Biomedical Engineering.

Joel S. Bader, Baltimore, MD

Contents

Part I

Cells as Linear Systems

Chapter 1

Cellular Signal Transduction

Chapter 2

Linear Systems Analysis

We left off last time with a model for a two-state biological signaling element,

$$(d/dt)x(t) = \beta(t) - \alpha x(t).$$

Here, $x(t)$ represents the concentration of the active form of a signaling molecule, usually meaning it is phosphorylated. The input is $\beta(t)$, and we consider it to be under our control. The rate that the activate form reverts to the inactive form is α.

Formally, we could write the solution as

$$[(d/dt) + \alpha]x(t) = \beta(t);$$

$$x(t) = [(d/dt) + \alpha]^{-1}\beta(t).$$

The problem is that we don't know what it means to take the inverse of an operator like the time derivative operator d/dt.

This is a lot like solving a matrix equation,

$$\mathbf{Ax} = \mathbf{b} - \alpha\mathbf{x}.$$

I use capital bold letters to indicate matrices and lower case bold to indicate column vectors. Elements of matrices and vectors are not bold, A_{ij} and x_i, We think about discretizing time so instead of $x(t)$ we have a vector $\mathbf{x}$ with elements $x_n = x(n\Delta t)$.

If we want this to be our actual problem, then $\mathbf{A}$ should be the time derivative operator in discrete form. Just to show you how we can do this, use the symmetric form

$$(d/dt)x_n = [x_{n+1} - x_{n-1}]/2\Delta t.$$

We also know that

$$(d/dt)x_n = \sum_{n'} A_{nn'}x_n = A_{n,n+1}x_{n+1} - A_{n,n-1}x_{n-1}.$$

$$A_{n,n'} = (1/2\Delta t)(\delta_{n',n+1} - \delta_{n,n-1}.$$

The discrete or Kronecker δ-function is 1 if its arguments are the same and 0 otherwise. One way to picture $\mathbf{A}$ is a tridiagonal matrix with 1's in the diagonal above the main diagonal, 0's in the main diagonal, and -1's in the diagonal below the main diagonal.

Formally, we could solve the algebraic equation as

$$\mathbf{x} = [\mathbf{A} + \alpha \mathbf{I}]^{-1}\mathbf{x}.$$

The matrix $\mathbf{I}$ is the identity matrix, with $I_{nn'} = \delta_{nn'}$ using our friend the δ-function. We wouldn't want to solve this by hand though because taking an inverse of a large matrix is difficult.

Instead this is why we learned about eigenvectors and eigenvalues because they change the matrix inverse into a scalar inverse. We're going to do exactly the same thing here by thinking about eigenfunctions and eigenvalues.

An operator A operates on a function $f(t)$ to give a new function $Af(t) = g(t)$. We will limit ourselves to operators that we could express as matrices if we made time discrete. The main operator we will consider is the time derivative operator d/dt. We will simplify our problem is we can express everything in terms of eigenfunctions of d/dt, functions for which

$$(d/dt)f(t) \propto f(t).$$

The proportionality constant could be any scalar. Pure exponentials are eigenfunctions of d/dt,

$$(d/dt)e^{\lambda t} = \lambda e^{\lambda t}.$$

We use λ because everyone knows that λ is the name of a generic eigenvalue. Just the same way that a matrix can have many different eigenvectors, each with a different eigenvalue, an operator can have many eigenfunctions. Here we have an infinite number.

We could index each eigenfunction by its eigenvalue, $f_\lambda(t) = e^{\lambda t}$. If λ is pure real, then we have functions that grow or decay with time. We'll start instead with eigenvalues that are pure imaginary, $\lambda = i\omega$, because Fourier transforms seem more symmetric than Laplace transforms. Our convention is to think about basis functions $\phi_\omega(t) = e^{i\omega t}$.

Now really we could have any scalar in front of $\phi_\omega t$ and it would still have the same eigenvalue $i\omega$. This is the same as with eigenvectors where we fix the overall scale by insisting that eigenvectors are normalized to have a dot product of 1. Actually we want their dot products to be orthonormal. For functions, rather than the dot product, we use the inner product,

$$\langle f(t)|g(t)\rangle \equiv \int_{-\infty}^{\infty} dt\,[f(t)]^* g(t),$$

where $[f(t)]^*$ is the complex conjugate of $f(t)$. For eigenfunctions of d/dt we could abbreviate the inner product as $\langle \omega'|\omega\rangle$. If we are thinking about discrete time, then the ω values are also discrete, and we want $\langle \omega'|\omega\rangle = \delta_{\omega',\omega}$. We will do this as a homework problem to see that the correct scalar for $\phi_\omega(t)$ is $1/\sqrt{2\pi}$, so that

$$\phi_\omega(t) = (1/\sqrt{2\pi})e^{i\omega t}.$$

Notice that the inner product has two factors of $1/\sqrt{2\pi}$, and

$$\langle\omega'|\omega\rangle = (1/2\pi)\int_{-\infty}^{\infty} dt\, e^{-i\omega' t} e^{i\omega t}.$$

Math tends to split these factors symmetrically between $\langle\omega|$ and $|\omega\rangle$. Enygineering and physics usually puts the entire factor of $1/2\pi$ into $|\omega\rangle$ so that

$$x(t) = \int_{-\infty}^{\infty} d\omega \hat{x}(\omega)|\omega\rangle = \int_{-\infty}^{\infty} (d\omega/2\pi)\hat{x}(\omega)e^{i\omega t}$$

$$\hat{x}(\omega) = \langle\omega|x\rangle = \int_{-\infty}^{\infty} dt e^{-i\omega t} x(t).$$

While this would be the discrete Kronecker δ-function for a discrete time representation, in the limit that we have continuous time it becomes the Dirac δ-function, $\delta(\omega - \omega')$. For any finite value of $\Delta\omega = \omega - \omega'$, the integral goes to 0. Actually the convergence of the integral to 0 is tricky, but you can think about the indefinite integral being $e^{i\Delta\omega t}/i\Delta\omega$, which is evaluated at endpoints T and $-T$. These are so big that $e^{i\Delta\omega T}$ is oscillating so rapidly that it looks like 0.

When $\Delta\omega \to 0$, the function $\delta(\Delta\omega) \to \infty$, but in a very nice way: the area under the δ-function is 1. For any finite ε,

$$\int_{\omega-\varepsilon}^{\omega+\varepsilon} d\omega' \delta(\omega' - \omega) = 1.$$

This also makes integrals involving the δ-function easy,

$$\int_{-\infty}^{\infty} d\omega' f(\omega')\delta(\omega' - \omega) = f(\omega).$$

It just picks out the value of the rest of the integrand when its argument is 0.

If this doesn't make sense, don't worry. You'll prove all of this in homework.

Returning to our problem, our plan is to write each of our time domain functions as a sum of eigenfunctions.

$$x(t) = \sum_{\omega} \hat{x}(\omega)|\omega\rangle.$$

$$\beta(t) = \sum_{\omega} \hat{\beta}(\omega)|\omega\rangle.$$

The terms $\hat{x}$ and $\hat{\beta}$ are just the expansion coefficients. Putting this expansion into the starting equation,

$$(d/dt)\sum_{\omega} \hat{x}(\omega)|\omega\rangle = \sum_{\omega} \hat{\beta}|\omega\rangle > -\alpha \sum_{\omega} \hat{x}(\omega)|\omega\rangle.$$

Now we can eliminate the time derivative in favor of the eigenvalue,

$$\sum_{\omega} (i\omega + \alpha)\hat{x}(\omega)|\omega\rangle = \sum_{\omega} \hat{\beta}(\omega)|\omega\rangle.$$

Remember that what we know is $\beta(t)$, which means that we should be able to figure out the expansion coefficients $\hat{\beta}(\omega)$. We want to solve for the output expansion coefficients $\hat{x}(\omega)$. We can do this for a particular value ω' by taking the inner product,

$$\sum_{\omega}(i\omega+\alpha)\hat{x}(\omega)\langle\omega'|\omega\rangle = \sum_{\omega}\hat{\beta}(\omega)\langle\omega'|\omega\rangle.$$

$$(i\omega'+\alpha)\hat{x}(\omega') = \hat{\beta}(\omega').$$

$$\hat{x}(\omega) = (i\omega+\alpha)^{-1}\hat{\beta}(\omega).$$

We can write down the formal solution,

$$x(t) = \sum_{\omega}\hat{x}(\omega)|\omega\rangle.$$

For continuous time, the sum becomes an integral, with details in the homework,

$$x(t) = (1/2\pi)\int_{-\infty}^{\infty} d\omega(i\omega+\alpha)^{-1}e^{i\omega t}\hat{\beta}(\omega).$$

Substituting the inner product that gives us the expansion coefficient $\hat{\beta}(\omega)$,

$$x(t) = (1/2\pi)\int_{-\infty}^{\infty} d\omega(i\omega+\alpha)^{-1}e^{i\omega t}\int_{-\infty}^{\infty} dt' e^{-i\omega t'}\beta(t')$$

We will next change the order of the integrals. We can usually do this for physical systems. We will always be able to do it in this class.

$$x(t) = \int_{-\infty}^{\infty} dt' \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \frac{\exp[i\omega(t-t')]}{i\omega+\alpha}\beta(t').$$

Let's think of this as a convolution or a filter,

$$x(t) = \int_{-\infty}^{\infty} dt' H(t-t')\beta(t'),$$

where the response function is

$$H(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \frac{\exp[i\omega(t-t')]}{i\omega+\alpha}.$$

Take a step back and breathe after the math blizzard. We have an output $x(t)$ that comes from an ODE model for a system that is driven by input $\beta(t)$. In a causal universe, $x(t)$ should only depend on the input at times before t,

$$x(t) = \int_{-\infty}^{t} dt' H(t-t')\beta(t').$$

Plot twist! Our integral doesn't stop at t. The integral goes to infinity. What are the possibilities?

1. We made a math mistake somewhere.
2. The universe (or our model for it) is not causal.
3. There is something special about the response function $H(t)$ for causal systems.

Spoiler alert: it's the last one. Response functions for classical causal systems are only non-zero for responses to inputs in the past. In other words, if the response function $H(t-t')$ is the response of the system at time t to an input at time t', then $H(t-t')$ must be 0 for $t < t'$. Next class we'll show this by doing the integral for our system's response function.

Chapter 3

The Laplace Transform and Complex Variables

Chapter 4

Signal Transduction Cascades and MAPK Signaling

Chapter 5

Generating Functions for Pharmacokinetics and Pharmacodynamics

Chapter 6

Positive Feedback and Caffeine Response

Part II

Cells as Non-linear Systems

Chapter 7

Saturation and Cooperative Response

Chapter 8

Joint Models of Transcription and Translation

Chapter 9

Positive and Negative Auto-Regulation

Chapter 10

Combinatorial Regulation

Chapter 11

Non-Linear Cascades and Logic Gates

Part III

Cells as Stochastic Systems

Chapter 12

Delta-Notch Signaling

Chapter 13

Stochastic Dynamics

Chapter 14

Noise in Gene and Protein Expression

Chapter 15

Stochastic Simulations and the Gillespie Algorithm

Chapter 16

Stability Analysis

Part IV

Cells as Spatial Systems

Chapter 17

Morphogen Gradient Patterning

Chapter 18

Diffusion

Chapter 19

Solving the Diffusion Equation

Chapter 20

Patterning and Noise

Part V

Cellular Networks

Chapter 21

Diffusion on a Network

Chapter 22

Network Topology, Motifs, and Clustering

Chapter 23

The Giant Component

Chapter 24

Network Partitioning and Spectral Clustering

Chapter 25

Metabolic Networks and Flux Balance Analysis

Appendix A

Problems