

# Quantitative Biology: An Introductory Tour

Chris Wiggins

January 20, 2016



# Contents

<b>I Biological Physics</b>	<b>23</b>
<b>1 Physics at the Scale of the Cell</b>	<b>25</b>
1.1 The scale of the cell . . . . .	25
1.1.1 How big is a cell? . . . . .	25
1.1.2 What is inside a cell? . . . . .	27
1.2 Dynamics at the scale of the cell / The life aquatic . . . . .	29
1.2.1 History: Brownian motion . . . . .	29
1.2.2 Viscosity dependence . . . . .	29
1.2.3 Einstein's explanation . . . . .	29
1.2.4 Temperature: $k_B T = 4pNnm$ . . . . .	30
1.2.5 Drag: $F = bv$ . . . . .	30
1.2.6 Dimensional analysis: $\mu$ is not $b$ . . . . .	31
1.2.7 A dimensionful model . . . . .	32
1.2.8 A digression: physics not at the scale of the cell . . . . .	33
1.3 Diffusion . . . . .	33
1.3.1 Diffusion: $D = k_B T/b$ . . . . .	34
1.3.2 "Random walks in biology" . . . . .	34
1.4 Probability and probability distributions . . . . .	36
1.4.1 What just happened? . . . . .	36
1.4.2 The discrete case . . . . .	38
1.4.3 The continuum limit . . . . .	39
1.4.4 Transport: $\partial_t \rho = -\partial_x \rho \mathbf{v}$ . . . . .	41
1.4.5 Conservation laws: $\partial_t \rho = -J'$ . . . . .	41
1.4.6 Diffusion and drift $J = \rho v - D \rho'$ . . . . .	41
1.4.7 Application of diffusion with drift: the Chemotaxis story .	42
<b>2 Equilibrium Statistical Mechanics: <math>J = 0</math></b>	<b>45</b>
2.1 Biological applications . . . . .	46
2.1.1 Pollen grains ( $U \propto x$ ) . . . . .	46

2.1.2	Calibrating an optical trap ( $U \propto x^2$ ) . . . . .	46
2.1.3	The Ising model of DNA extension ( $U \propto \pm 1$ ): . . . . .	47
2.2	The partition function . . . . .	48
2.3	Linear response theory . . . . .	49
2.4	Electrostatic surprises in vivo . . . . .	50
2.4.1	DNA and electrostatic interactions ( $U \propto \ln x$ ) . . . . .	50
2.4.2	Electrostatic screening ( $U = qV \ll k_B T$ ) . . . . .	51
2.4.3	HW: linear response . . . . .	52
<b>3</b>	<b>Non-equilibria and transport: <math>J = constant</math></b>	<b>53</b>
3.1	Open equilibria . . . . .	53
3.2	escape rate . . . . .	54
3.2.1	Thermal ratchets: $U = -fx$ ; system size= $L$ . . . . .	54
3.2.2	Kramer's escape . . . . .	55
3.3	Chemical dynamics; "detailed balance" . . . . .	56
3.3.1	HW: Poisson equation . . . . .	56
3.3.2	Question . . . . .	57
3.3.3	Question, condensation . . . . .	58
3.3.4	What class should be able to do . . . . .	58
3.3.5	Thermal ratchet model . . . . .	60
3.3.6	90's physics . . . . .	61
3.3.7	HW . . . . .	65
3.3.8	Two state model . . . . .	65
3.4	Lecture 9 . . . . .	66
3.4.1	Check-up review . . . . .	66
<b>II</b>	<b>Systems Biology</b>	<b>71</b>
<b>4</b>	<b>Regulatory Dynamics</b>	<b>73</b>
4.1	Kinetic models: 1-d dynamical systems . . . . .	73
4.1.1	Chemical kinetics . . . . .	73
4.1.2	Michaelis-Menton kinetics . . . . .	74
4.1.3	Hill functions and regulation models . . . . .	75
4.2	Dynamical systems models for regulation . . . . .	75
4.2.1	Addendum to the parts list: informatic parts . . . . .	75
4.2.2	Mathematical models of transcriptional regulation . . . . .	76
4.3	Experiment: the toggle switch . . . . .	78
4.3.1	Lambda phage . . . . .	78
4.3.2	GFP . . . . .	78

4.3.3	SGNs . . . . .	78
4.4	Transduction . . . . .	78
4.4.1	Lambda phage . . . . .	78
4.4.2	Chemotaxis . . . . .	78
4.4.3	MapK . . . . .	78
4.4.4	HW . . . . .	82
4.4.5	Systems Biology . . . . .	83
4.4.6	Homework . . . . .	85
4.4.7	HW: Transcription upregulation . . . . .	87
4.4.8	HW given previously, inhibitor, down-regulation . . . . .	89
4.4.9	Up-regulation . . . . .	90
4.4.10	Student question . . . . .	93
4.5	Chemical kinetics review . . . . .	93
4.5.1	Up- and down-regulation . . . . .	93
4.5.2	Regulation revisited . . . . .	97
<b>5</b>	<b>Interactions: Dynamical systems</b>	<b>101</b>
5.1	Dynamical systems and linear algebra . . . . .	102
5.2	Kinetic models: 2-d dynamical systems . . . . .	103
5.3	Synthetic genetic networks: oscillation . . . . .	104
5.3.1	HW: Transcriptional regulation review . . . . .	104
5.3.2	Previous homework . . . . .	107
5.3.3	Simple two-dimensional dynamical systems . . . . .	109
5.3.4	A deeper look at dynamical systems . . . . .	110
5.3.5	HW: cooperative upregulation . . . . .	112
5.3.6	Genetic engineering and “synthetic biology” . . . . .	115
5.3.7	HW: Transcription, translation . . . . .	121
<b>6</b>	<b>Noise</b>	<b>125</b>
6.1	Motivating example: Van Kampen’s cat . . . . .	125
6.2	Review of simple chemical kinetics . . . . .	125
6.2.1	Decay: $\dot{x} = -rx$ . . . . .	125
6.2.2	Constitutive creation: $\dot{x} = \alpha_0$ . . . . .	125
6.2.3	Two-state systems: $N_a + N_b = \text{constant}$ . . . . .	125
6.2.4	Decay to a nonzero fixed point: $\dot{x} = \alpha_0 - rx$ . . . . .	125
6.2.5	Nonlinearity: $\dot{x} = -rx + x^2/(1 + x^2)$ . . . . .	125
6.2.6	Transcription-translation models . . . . .	125
6.2.7	Oscillations . . . . .	125
6.3	Stochastic kinetics . . . . .	126
6.3.1	Constitutive creation: $0 \rightarrow x$ . . . . .	126

6.3.2	Decay: $0 \rightarrow x$	126
6.3.3	Decay: $x \rightarrow 0$	126
6.3.4	Decay to a fixed point: $0 \leftrightarrow x$	126
6.3.5	Two-state systems: $N_a + N_b = constant$	127
6.4	The general case: $\dot{p}_n = -j(n) + j(n-1)$	127
6.4.1	Conservation	127
6.4.2	Moment dynamics	128
6.5	Return to the cat	129
6.6	Lecture 14	129
6.6.1	HW questions	129
6.6.2	Viruses and life and death	134
6.6.3	HW	140
6.7	Lecture 15	140
6.7.1	Transcriptional regulatory networks	140
6.8	Lecture 16	148
6.8.1	Review of stochastic description of chemical kinetics	148
6.9	Lecture 17	156
6.9.1	Chemical kinetics review	156
6.9.2	Dynamics and nondimensionalization review	159
6.9.3	Master equation kinetics	161
<b>7</b>	<b>Biological networks</b>	<b>165</b>
7.1	Questions and methods	166
7.1.1	Network statistics	166
7.1.2	Form vs function	166
7.2	What is to be measured?	166
7.2.1	Qualitative questions	166
7.2.2	Quantitative features	166
7.3	Modeling	167
7.3.1	Graph drawing: a caveat	167
7.3.2	Subgraph census	168
7.3.3	Price's model	168
<b>III</b>	<b>Learning models from data</b>	<b>169</b>
<b>8</b>	<b>Complexity Control</b>	<b>173</b>
8.1	Motivating example: inferring networks	173
8.1.1	Genetic networks	173
8.1.2	Scales of genomes	173

8.1.3	Densities of inferred genetic networks . . . . .	173
8.1.4	Overfitting data . . . . .	174
8.2	‘Model selection’ . . . . .	174
8.2.1	The Bayesian approach . . . . .	174
8.2.2	The machine learning approach . . . . .	174
8.3	Statistical significance . . . . .	175
8.3.1	Use and abuse of $p$ -values . . . . .	175
8.3.2	Bootstrap vs biology . . . . .	175
8.4	Data models . . . . .	175
8.4.1	Multinomial density estimation . . . . .	175
8.4.2	Mixture model density estimation . . . . .	175
8.4.3	Learning continuous distributions . . . . .	175
8.5	Applications . . . . .	175
8.5.1	Microarrays . . . . .	175
8.5.2	Binding sites as “motifs” . . . . .	175
8.5.3	Binding sites as binding sites . . . . .	175
<b>9</b>	<b>Complexity Control</b>	<b>177</b>
9.1	Learning regulatory interactions from data . . . . .	177
9.1.1	Models, data, and uncertainty . . . . .	178
9.1.2	Learning parameters from data . . . . .	179
9.2	Lecture 18 - Data . . . . .	179
9.2.1	Finding keywords in RNA sequence . . . . .	180
9.2.2	Minimize squared error . . . . .	182
9.2.3	Finding “important” words . . . . .	182
9.3	Lecture 19 . . . . .	185
9.3.1	Review . . . . .	185
9.4	Lecture 20 . . . . .	191
9.4.1	HW questions- Linear Regression . . . . .	191
9.5	Lecture 21 . . . . .	197
9.5.1	Modeling . . . . .	197
9.6	Lecture 22 - Final review . . . . .	205
9.6.1	Schrodinger’s cat . . . . .	205
9.6.2	Van Kampen’s (sp??) cat . . . . .	205
9.6.3	Maximum likelihood . . . . .	206
9.6.4	New material . . . . .	208
9.6.5	Outline of this part of class . . . . .	212

<b>10 Classification: modeling from data without data modeling</b>	<b>215</b>
10.1 A story . . . . .	215
10.2 Data without the model: prototype models . . . . .	215
10.3 Data modeling: classical classification . . . . .	215
10.4 Feature selection: classical approaches . . . . .	215
10.4.1 Discriminability . . . . .	215
10.4.2 Nonparametric statistics . . . . .	216
10.5 The machine learning approach . . . . .	216
10.6 Large margin classification . . . . .	216
10.6.1 SVMs . . . . .	216
10.6.2 Boosting . . . . .	216
10.7 Applications . . . . .	216
10.7.1 Clinical . . . . .	216
10.7.2 Network classification . . . . .	216
10.7.3 Network inference . . . . .	216
<b>11 Clustering</b>	<b>217</b>
11.1 Basic notions . . . . .	217
11.1.1 History . . . . .	217
11.1.2 Why would you do this? . . . . .	217
11.2 Orthodox clustering . . . . .	217
11.2.1 Generative . . . . .	217
11.2.2 Hierarchical . . . . .	217
11.2.3 How many clusters? . . . . .	218
11.3 Foreshadowing and context . . . . .	218
11.3.1 Unorthodox (supervised) clustering . . . . .	218
11.3.2 from side information to relevant information . . . . .	218
<b>12 Biological Information</b>	<b>219</b>
12.1 What is information? . . . . .	219
12.1.1 History: Shannon's postulates . . . . .	219
12.1.2 Information is average dependence . . . . .	219
12.1.3 Re-parameterization-invariance . . . . .	222
12.1.4 $D_{KL}$ : how many questions must I ask? . . . . .	223
12.2 Density estimation . . . . .	223
12.2.1 Maximum entropy distribution . . . . .	223
12.2.2 Complexity control . . . . .	224
12.2.3 Stat mech, revisited . . . . .	224
12.3 The information bottleneck . . . . .	225
12.4 What is modeling? . . . . .	226

<b>IV Coda</b>	<b>227</b>
<b>13 Summary</b>	<b>229</b>
13.1 Physics at the scale of cell . . . . .	230
13.2 Systems biology . . . . .	231
13.3 Modeling . . . . .	231
<b>14 Final thoughts: What is a model?</b>	<b>233</b>
<b>A Biology as told by a theoretical physicist</b>	<b>235</b>
A.1 What is biophysics? . . . . .	235
A.2 The “parts list:” a physicist cartoons biology . . . . .	236
A.2.1 Cells . . . . .	237
A.2.2 proteins . . . . .	237
A.2.3 small molecules . . . . .	237
A.2.4 polymers . . . . .	238
A.2.5 nucleic acids . . . . .	238
A.2.6 Final ideas . . . . .	238
A.3 “The machinery of life:” a biologist draws biology . . . . .	239
<b>B Dimensions of things</b>	<b>247</b>
B.1 Dimensional analysis . . . . .	247
B.2 Dimensional estimation . . . . .	249
<b>C A brief review of all of undergraduate mathematics</b>	<b>251</b>
C.1 Pre-calculus . . . . .	251
C.2 Differential calculus . . . . .	251
C.2.1 Differentiation and derivatives . . . . .	251
C.2.2 Taylor expansion, the most powerful and general tool in over 300 years of Calculus <sup>TM</sup> . . . . .	252
C.2.3 Application . . . . .	252
C.3 Integral calculus: the opposite of differential calculus . . . . .	253
C.3.1 Simple examples: . . . . .	254
C.3.2 The best trick in calculus: integration by differentiation . .	254
C.3.3 The Gaussian integral, the only integral you really need to know <sup>TM</sup> . . . . .	255
C.4 $\dot{x} = \alpha x$ : The only ODE you really need to know <sup>TM</sup> . . . . .	258
C.5 Solving the Diffusion Equation . . . . .	258
C.5.1 fourier space . . . . .	259
C.5.2 ODE . . . . .	259

C.5.3	solution . . . . .	259
C.5.4	come back . . . . .	260
C.6	Linear algebra . . . . .	260
C.6.1	Dynamical systems and linear algebra . . . . .	260
C.6.2	the simplest 2D dynamical system in the entire world . . . . .	261
C.6.3	Complex magic values? . . . . .	263
<b>D</b>	<b>Probability</b>	<b>265</b>
D.1	normalization . . . . .	266
D.2	expectation . . . . .	266
D.3	marginalization . . . . .	266
D.4	bayes rule . . . . .	266
D.5	probabilities and probability distributions . . . . .	266
<b>E</b>	<b>Elastica</b>	<b>267</b>
<b>F</b>	<b>Stirling's approximation</b>	<b>269</b>
F.1	Approximate approximation: $n! \approx N^N e^{-N}$ . . . . .	269
F.2	Next order: $n! \approx N^N e^{-N} \sqrt{2\pi N}$ . . . . .	269
F.3	Application to the binomial . . . . .	270
F.3.1	The binomial is nearly a Gaussian $q = 1/2$ , $j \approx N/2$ . . . . .	270
F.3.2	The binomial is nearly Poisson: $q \ll 1$ , $qN \sim 1$ $j \ll N$ . . . . .	272

# List of Figures

1.1	taken from [5] . . . . .	26
1.2	Leeuwenhoek's scope. . . . .	26
1.3	According to Goodsell . . . . .	28
1.4	two plates sliding past each other . . . . .	31
1.5	freely-jointed chain . . . . .	36
1.6	Pascal's triangle/diffusion dynamic . . . . .	38
1.7	Rate of change in $N$ is the rate at which stuff goes into the box minus the rate of stuff leaving the box. Since velocity is positive when things are moving to the right, we have Eqn. 1.64. . . . .	41
3.1	salt and large charged object . . . . .	57
3.2	DNA, $\Phi \propto \lambda \propto \frac{e}{b}$ . . . . .	57
3.3	DNA coiling, proteins wrap up DNA . . . . .	58
3.4	Charged object, salts . . . . .	59
3.5	motor protein, 8 nm lattice sites . . . . .	60
3.6	$x = 0 = x_- = x_0$ to $x_+ = L$ . . . . .	60
3.7	motor protein and optical trap, kinetics, probability . . . . .	61
3.8	Energy versus chemical reaction . . . . .	62
3.9	Doubly peaked probability . . . . .	62
3.10	Doubly peaked probability . . . . .	63
3.11	Doubly peaked probability, $U_a = U_b$ . . . . .	63
3.12	Doubly peaked probability . . . . .	64
3.13	Doubly peaked probability with length scales . . . . .	64
3.14	Doubly peaked probability with escape to left . . . . .	65
3.15	$P_l, P_R$ fluxes . . . . .	65
3.16	motor protein . . . . .	67
3.17	$U(x)$ . . . . .	68
3.18	. . . . .	68
3.19	Spring trap . . . . .	70

4.1	a gene, showing regulatory region with “motifs” (really binding sites)	75
4.2	a transcription factor, in action (here, repressing), in the major groove	76
4.3	A plot of $\gamma p$ and $\alpha(p)$ : decay and transcription/translation, indicating the existence of a “fixed point.”	77
4.4	reaction diagram	78
4.5	plot	78
4.6	ATP docks with protein binding site	79
4.7	$r(s)$	80
4.8	Cells, lipid bilayers, proteins, ATP, free phosphate, nucleic acid	81
4.9	Transcription: DNA copied to nucleic acid	81
4.10	$r/m_T$ versus $f$	82
4.11	DNA	82
4.12	Double well	83
4.13	enzyme	83
4.14	DNA	84
4.15	$r(f)$	85
4.16	Sketch tetramer	86
4.17	Transcription binding site	86
4.18	$r(f)$	87
4.19	transcription binding	87
4.20	tetramer sketch	88
4.21	Graphs of $\alpha_1^\pm$ and $\alpha_n^\pm$	90
4.22	Transcription, translation sketch	91
4.23	$\alpha$	92
4.24	Upregulating gene	92
4.25	double well	93
4.26	DNA transcription	93
4.27	$\alpha$	94
4.28	$\alpha$ for auto-upregulation	95
4.29	equilibrium points	96
4.30	$\alpha$ for auto-upregulation HW 1	96
4.31	$\alpha$ for cooperative upregulation	96
4.32	$\alpha$ for down-regulation	97
4.33	double well	98
4.34	vectors	99
5.1	mRNA	105
5.2	$\alpha$ plots	105
5.3	$f(y)$ fixed points	106
5.4	$\alpha$ plot	107

5.5	$\epsilon$	108
5.6	$y(\tau)$ plot	108
5.7	Fixed point at the origin	109
5.8	Stability in parameter space	109
5.9	$(a, b)$ -plane	110
5.10	Eigenvector coordinates	111
5.11	Stability in $(\tau, \Delta)$ -plane	112
5.12	DNA	113
5.13	Two fixed points	114
5.14	Plasmid	115
5.15	N-cycle	116
5.16	Stability diagram in $(\tau, \Delta)$ -plane	117
5.17	Stability cases	117
5.18	spiral trajectories	118
5.19	circle trajectories	118
5.20	stability	119
5.21	double well	119
5.22	phase plane for double well	119
5.23	eigendirections	120
5.24	eigenvector plot	121
5.25	$\tau, \Delta$	122
5.26	N-repressor cycle	123
6.1	repressor cycles	130
6.2	N-repressor cycle	130
6.3	HW cycle	131
6.4	Stable spiral	131
6.5	values	132
6.6	Cosine	133
6.7	N-cycle	133
6.8	N-cycle	134
6.9	Virus count	134
6.10	E. Coli. RNA	135
6.11	Pollen grain	135
6.12	Solution by characteristics	137
6.13	Shifted delta distribution	137
6.14	Poisson distribution	139
6.15	Transcriptional regulation	140
6.16	Solution of probability PDE	142
6.17	transcription	145

6.18	Stability of macroscopic system	149
6.19	$\lambda$ plot	149
6.20	Poisson distribution	150
6.21	$\alpha(x)$	150
6.22	Fixed point diagram	151
6.23	Distribution	151
6.24	Fixed point cases	152
6.25	Fixed point	153
6.26	Distribution with $\langle n \rangle = Vx_* = V\alpha/r$	153
6.27	$\alpha(x)$	153
6.28	Propagating average	154
6.29	MATLAB	157
6.30	MATLAB	158
6.31	Stability diagram	161
6.32	Decaying oscillation	161
6.33	Fixed points	164
6.34	Fixed points	164
9.1	GATACA	180
9.2	$\frac{x}{l} = N \tanh \left( \frac{Fl}{K_a T} \right)$	180
9.3	GATACA	181
9.4	mRNA microarray	181
9.5	data	181
9.6	microarray data	182
9.7	data plot	183
9.8	data	185
9.9	Heavisides	187
9.10	fit plot	188
9.11	data plot	190
9.12	polynomial fit plot	190
9.13	data	192
9.14	data	192
9.15	$S(K)$	196
9.16	data fits for $x(F)$	196
9.17	$\chi^2$	197
9.18	$\mathbf{y}(\mathbf{x}) \Rightarrow \hat{\mathbf{l}}$	197
9.19	$\mathcal{A}_g \Rightarrow \hat{\vartheta}_\mu, \mathcal{A}_g = \sum_\mu \mathcal{N}_g^\mu \vartheta_\mu$	198
9.20	DNA sketch	198
9.21	linear regression	199
9.22	DNA	199

9.23	BIC data . . . . .	200
9.24	Cross-validation error . . . . .	200
9.25	p-values plot . . . . .	200
9.26	robust regression p-value plot . . . . .	201
9.27	microarray data at different times . . . . .	201
9.28	CV matrix . . . . .	202
9.29	genes in buckets . . . . .	202
9.30	microarrays . . . . .	203
9.31	gender and height sketch . . . . .	203
9.32	Error versus $\vartheta$ . . . . .	203
9.33	Boy/girl data . . . . .	204
9.34	Linear regression data . . . . .	204
9.35	Support Vector Machines . . . . .	205
9.36	Schrodinger's cat . . . . .	206
9.37	E. Coli. . . . .	206
9.38	plus/minus data . . . . .	209
9.39	classification parameters . . . . .	209
9.40	Dividing data into two classes . . . . .	209
9.41	nice data fit . . . . .	209
9.42	Heaviside function . . . . .	210
9.43	probability plot . . . . .	210
9.44	boy and girl height data . . . . .	211
9.45	height plots without gender . . . . .	211
9.46	two peaked probability . . . . .	211
9.47	agglomerative hierarchical pairwise clustering . . . . .	212
9.48	classification . . . . .	212
9.49	Bayesian information criterion . . . . .	213
A.1	parts list, as drawn by a theoretical physicist . . . . .	236
A.2	scales of things . . . . .	239
A.3	ATP, the central energy storage molecule, being born. Note that it is not made the way protein is made (transcription and translation), nor the way “vesicles” are made (self-assembly via hydrophobic forces) . . . . .	240
A.4	A typical lipid bilayer, made via hydrophobic self-assembly, like oil and water. . . . .	241
A.5	scales of small things . . . . .	242
A.6	mechanical coupling . . . . .	243
A.7	mRNA being born . . . . .	243

A.8 A transcription factor binding to DNA. This interaction, between a protein and a nucleic acid, is the fundamental interaction of transcriptional regulation — genes turning “on” and “off.” There will be plenty more to say about this in Parts II and III. . . . .	244
A.9 Protein synthesis . . . . .	245
A.10 The ribosome, a machine made of nucleic acids and proteins for translating mRNA into protein. . . . .	246

## List of Tables



# Preface

## **Manifesto:**

There have been three revolutions in quantitative biology in the last decade. The first was biological physics: a wealth of new micromanipulation and microobservation techniques, along with a renewed open-mindedness by the physics community, led to the lowering of the boundary between cellular biology and (principally) soft condensed matter physics. Much of this research is in effect on the physics of biological materials. The second was systems biology – the study of collective phenomena in molecular biology – in some part a reaction to single-molecule biophysics. Some of this is informed by collective phenomena in physics (again, by soft condensed matter physicists), and some of it by engineering, and some, such as “network physics,” is still emerging as a research field. The third is the proliferation of ‘high-throughput’ biological technologies, coinciding with a WWW to facilitate the broad dissemination of the resulting copious data. This has then driven innovation in bioinformatics, machine learning, statistics, data mining, and computational biology. The overlap among the sub- and super-fields in the preceding sentence is a reflection of how dynamic and emerging this area of research is as well.

## **Logic:**

The logic of the book is as follows: In part one, we meet physics at the scale of the cell ( $k_B T$ , viscosity, and diffusion), then the mathematical modeling of diffusion with drift (and, as a special case, some biologically-motivated examples of equilibrium and non-equilibrium probability distributions), the final example of which is chemical reactions. This discussion leads naturally into part two, in which we describe the mathematical modeling of biochemical and transcriptional regulatory (a.k.a. “genetic”) networks, both from a deterministic dynamical systems perspective and a stochastic, master-equation perspective, and finally discuss statistical descriptions of biological networks. This discussion leads to part three, beginning

with the inference of biological networks, which in turn leads to the more general subject of data-driven approaches to biology, including complexity control, clustering, machine learning, and biological applications of information theory.

It is worth noting that this course does not cover. By limiting ourselves to phenomena at the scale of a cell and smaller, we are leaving out, among other things, biological pattern formation, mathematical physiology, ecological and macro-immunological modeling, and neuroscience. Neuroscience is by far a more ancient field, meaning physicists, mathematicians, and neuroscientists were talking to each other more than 20 years ago. With good reason, neuroscientists often remark that they have been doing biological physics, systems biology, and statistical approaches to biological data for a long time – and they are right. We will not discuss Hodgkin and Huxley, Hubel and Weisel, systems of integrate-and-fire neurons, or even  $V = IR$ . Some of the tools we will use are applied in, or even are the norm in, computational and theoretical neuroscience, but, focusing on fields emerging in the last decade, the material in this course stands apart from these subjects. Moreover, a number of excellent books already exist to introduce to this subject, including [13, 7], among others.

## This book:

### Need:

*Students* are very good eye. Problem is *boss*.

-Satoru Miyano, summarizing the driving forces in, and principal impediments to, interdisciplinary research

Universities are not designed in a way to promote interdisciplinary work. In order to ensure all students, faculty, square feet, and dollars are accounted for and put in easy-to-trace baskets, walls have been constructed which prevent the names and natures of these baskets from changing. An unfortunate byproduct of this is that it can encourage members of a given field constantly to ask “is this new research problem *really* physics (or chemistry, or applied mathematics, etc) — *i.e.*, does it look like things already recognized, familiar, and canonized.” Such taxonomic defense necessarily thwarts innovation, rather than fostering it.

Students, however, being the least inculcated into the establishment of accounting divisions, are the most immune to them. They are also often responsible for introducing faculty to interesting research being done by their colleagues in other departments, and are thereby, simply by being curious and taking advantage of the resources presented to them, often responsible for identifying and driving new opportunities for collaborations.

I hope that this book arms some students with the basic material they need to continue transgressing disciplinary barriers, despite the occasional objections of faculty and/or deanery, for the good of science and those of us who are lucky enough to engage in scientific research professionally.

## Prerequisites

What makes modeling hard is not that it is more of the same, *e.g.*, “quantitative biology is a really hard (physics/math/computer science) problem”; it is hard because you have to understand physics, mathematics and biology well enough to build models which are physically valid, mathematically meaningful, and biologically interesting. The key to doing this is not the ability to renormalize the most daunting Hamiltonian, nor to match the most baroque asymptotics, nor to have encyclopedic knowledge of every aspect and relevant player in a biological system. Modeling is the construction of a representation of a system which is complex enough to retain aspects deemed relevant, yet simple enough to be interpretable and to generalize to new systems. Modeling, as we shall discuss in more and more detail throughout the book, is hard.

As a consequence, the mathematics that we need will not be terribly technical. It’s not that we need a *lot* of mathematics, it’s that we’ll need certain key pieces of mathematics and will need to apply this small, selected amount of mathematics in ways to which students are probably not accustomed. A partial list of the mathematics we will use includes probability, partial differential equations, calculus, dimensional analysis, dynamical systems (and therefore some linear algebra), and statistics.

To encourage as broad a cross-section of students as possible, I have included appendices which cover a few mathematical topics in depth. The topics included are based strictly on questions I have received from students, and are not meant to be exhaustive. This is not a book about math.

## History and acknowledgments:

This book was not motivated by advice from my colleagues or senior faculty. In fact, it is generally accepted that writing a book as a junior person is a bad idea. Instead, it was motivated by requests from my students, who wanted to know more about topics which were clearly being discussed with great excitement by faculty and at seminars, but were unrepresented in any textbooks. I hope that this book will give clearer answers than I was able to give in my lectures, and I hope that it serves as some small thanks to my students for pushing me to communicate modern research clearly – warts and all.

The book grew out of a ‘special topics’ class for graduate students in Fall of 2001, followed by iterations of a course for advanced undergraduates and beginning graduate students, in spring of 2003 (12 students), 2004 (14 students), and 2005 (10 students), 2006 (14 students), and 2007, during which the content converged, slowly and noisily. I thank the students, drawn from a broad variety of disciplines, who showed the bravery to take the course and persistence to endure, for their many questions, comments, suggestions, and inspiration. Finally, the odious task of TeXing this book was begun by Jacob Hofman and Christina Tosti based on notes taken by themselves and by Eric Raleigh in Spring 2004. Abby Hofman considerably expanded on this over the summer of 2007. I owe them my thanks. I also want to acknowledge the many graduate teaching assistants 2003-2007 and one undergraduate teaching assistnat (Yoni Ben Tov) in the spring of 2008.

# **Part I**

# **Biological Physics**



## Chapter 1

# Physics at the Scale of the Cell

Modeling necessarily entails deciding what aspects will be described explicitly and how one will model deviations from that description. Often this means what will be described *deterministically* and which part will be modeled *probabilistically*. Since modeling is a central theme of this book, probability will necessarily be a central tool. This decision of separating the deterministic-described from the probabilistically-described might be called separation of “signal” from “noise,” or, in Sec. 1.4.5, “drift” vs “diffusion.”

Here we’ll encounter the simplest and easiest-to-understand of all manners of biological noise: thermal, physical, “real” noise. It was one of the oldest applications of probability in biology, visualized in 1827, then quantified, and modeled mathematically in Einstein’s PhD Thesis. Moreover, the modeling tools we develop here will be used throughout the course.

## 1.1 The scale of the cell

### 1.1.1 How big is a cell?

Before describing physics at the scale of the cell, it’s wise to make sure we appreciate just what this scale is. Fig. 1.1 illustrates a progression of smaller and smaller objects, begining with one so small you have most likely never seen it, and which is three orders of magnitude bigger than anything we will model in this course.

One fact worth noting: the size of a cell,  $\sim 1\mu$ , is approximately the wavelength of visible light. This means that visualization of anything at this scale or smaller will be impossible with conventional (light) microscopy.

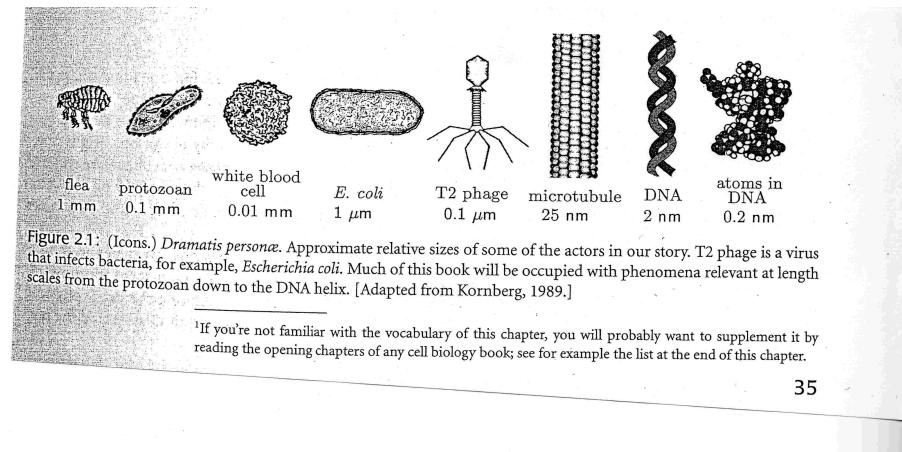


Figure 1.1: taken from [5]

**Question:**

On September 17, 1683, Leeuwenhoek wrote to the Royal Society about his observations of “many very little living animalcules, very prettily a-moving,” using a microscope such as the one pictured here (Fig. 1.2). These were among the first observations on living bacteria ever recorded.

Use this fact, along with the size of a bacterium, to determine a bound — and state whether it is a lower bound or an upper bound — on the wavelength of visible light.

**Answer:**

Leeuwenhoek’s observation sets the size of a bacterium as an upper bound on the wavelength of light. A bacterium is around a micron. To resolve something on the scale of a micron, light must be at the scale of (or smaller than) a micron.

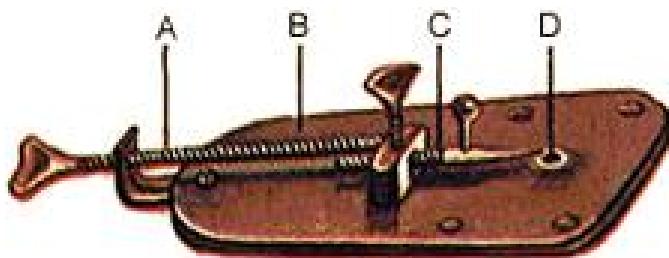


Figure 1.2: Leeuwenhoek’s scope.

### 1.1.2 What is inside a cell?

David Goodsell [3] is an expert at visualizing biology. One of countless examples is Fig. 1.3, taken from his website, which illustrates the inside of a single-celled organism, *Escherichia coli*. Recall that this is approximately a micron in diameter while looking at the substructures within the cell and reading his description:

This illustration shows a cross-section of a small portion of an *Escherichia coli* cell. The cell wall, with two concentric membranes studded with transmembrane proteins, is shown in green. A large flagellar motor crosses the entire wall, turning the flagellum that extends upwards from the surface. The cytoplasmic area is colored blue and purple. The large purple molecules are ribosomes and the small, L-shaped maroon molecules are tRNA, and the white strands are mRNA. Enzymes are shown in blue. The nucleoid region is shown in yellow and orange, with the long DNA circle shown in yellow, wrapped around HU protein (bacterial nucleosomes). In the center of the nucleoid region shown here, you might find a replication fork, with DNA polymerase (in red-orange) replicating new DNA.

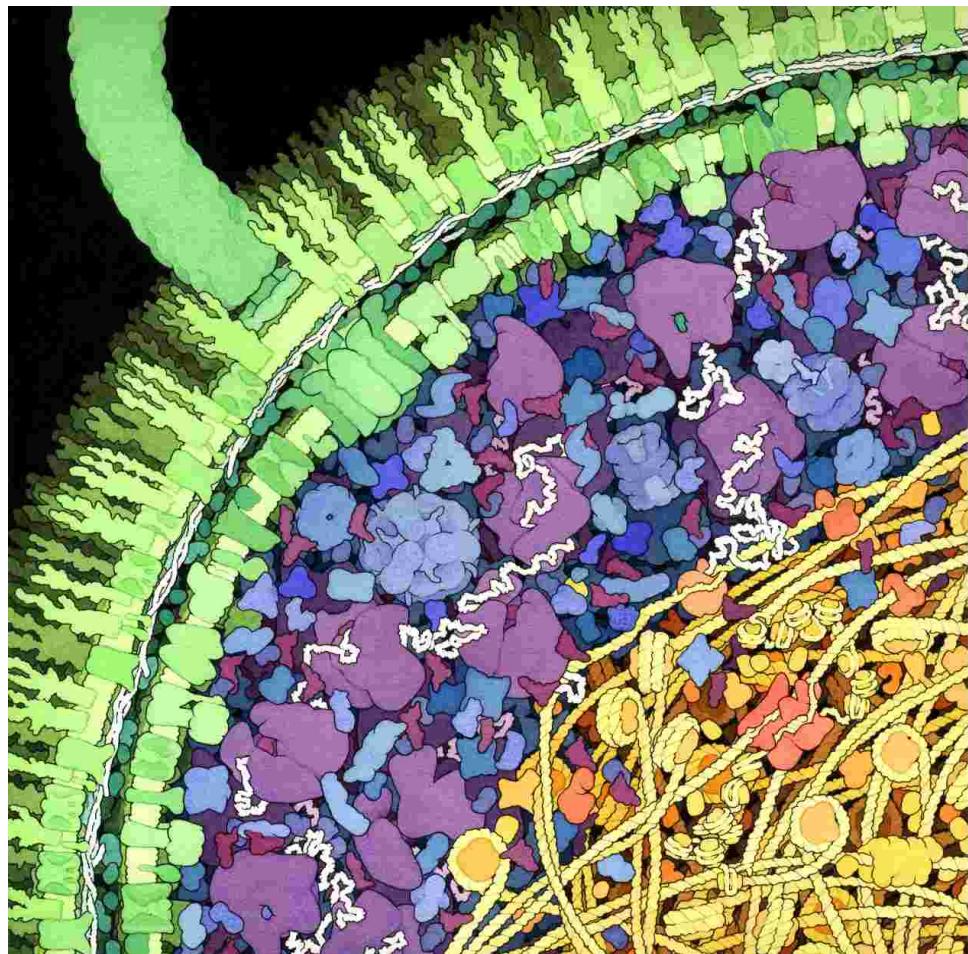


Figure 1.3: According to Goodsell

## 1.2 Dynamics at the scale of the cell / The life aquatic

### 1.2.1 History: Brownian motion

In 1828, the Botanist Robert Brown observed that pollen grains, about the same size as an *E. coli*, in water exhibit what Berg [6] might have called a “continuous riotous motion”, which PCN refers to as “a peculiar incessant dance.” [5] (page 109).

What surprised him in particular was the motion turned out to have nothing to do with the fact that Pollen is alive. To test this, he starved the pollen for a long time, simply leaving the Petri dish alone, and found that the motion had not changed.

### 1.2.2 Viscosity dependence

Today, repeating Brown’s observation is a common occurrence in undergraduate laboratories, and movies of Brownian motion can be found online. A great example is from the lab of Eric Weeks, which shows that the motion depends on the viscosity of the material in which it is embedded and the size of the particles: it decreases with increasing viscosity or increasing size of the particle. (*cf.* <http://www.physics.emory.edu/~weeks/squishy/BrownianMotionLab.html>)

Now let’s turn back to Fig. 1.3. It’s a beautiful image. Now picture Professor Weeks’ movies. Repeating this, picture the horrific movie of which Goodsell’s static image is a mere snapshot: DNA and proteins are created and translocated in the presence of a veritable mosh pit of thermal motion. Even for *E. coli* to get anywhere it must do so in the presence of strong and stochastic driving. It’s worth keeping this in mind as we begin to model the life aquatic.

### 1.2.3 Einstein’s explanation

Recall that Brown considered, among other possibilities, that Brownian motion will stop when the pollen grains “die.” This is radically different from the explanation put forth in Einstein’s “On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat” (May 1905; received 11 May 1905) *Annalen der Physik*, 17(1905), pp. 549-560.

A Java applet available online from AIP illustrates the phenomenon nicely. (*cf.* <http://www.aip.org/history/einstein/brownian.html>). The underlying physics, at the microscale, is simply  $F = ma$ . However, with scores of particles colliding, we see the emergence of two mesoscale “effective mechanisms”, of statistical origin: temperature and viscosity.

### 1.2.4 Temperature: $k_B T = 4pNm$

If you chill your vacuum cleaner, or even your television, to a degree above freezing, these appliances continue to work fine. But try this with a grasshopper, or even a bacterium, and you find that life processes practically stop. (After all, that's why you own a freezer in the first place.) Understanding the interplay of heat and work will prove to be crucial to the fundamental processes of life and will become a central obsession of this book" (PCN pp. 2-3)[5]

You may recall from other classes that temperature is proportional to the average kinetic energy of the particles.

What is temperature measured in? If you don't know, recall the dimensions of Boltzmann's constant  $k_B$ .

$$[k_B] = J/K \quad (1.1)$$

therefore,  $k_B T$  is energy; remember  $pV = nk_B T$ . Looking at this fact, we can see that it's a historical fluke that temperature is *not* measured in dimensions of energy — if we had known more thermodynamics, we would not have invented "Kelvins" (let alone degrees Celsius or Fahrenheit).

Room temperature, measured in Kelvins, is approximately

$$k_B T = 1.38 \cdot 10^{-23} J/K \cdot 300K \approx 4.14 \cdot 10^{-21} J. \quad (1.2)$$

Of course, we can express this energy in terms of a force and a distance, since work, force and distance are related via

$$W = F \cdot x \quad (1.3)$$

Note that if we are dealing with nanometers ( $10^{-9} m$ ) as a typical length scale,  $k_B T$  suggests a typical force scale of a picoNewton ( $10^{-12} N$ ), since

$$10^{-21} J = 10^{-21} Nm = 10^{-12} N \cdot 10^{-9} m = 1pNm \quad (1.4)$$

and we may rewrite, for a temperature around  $27C \approx 80F$ ,

$$k_B T \approx 4.14 \cdot pNm \quad (1.5)$$

expressing the characteristic energy scale for life at room temperature in terms of a characteristic force scale and a characteristic length scale.

### 1.2.5 Drag: $F = bv$

Consider two plates filled with fluid; you push, doing work. Where does that work go? We have done work, and therefore lost energy, which, via dissipation or "drag", has become heat. "That means that energy, which is conserved, is flowing from something useful (e.g., motion) to something useless (heat)." [5]

Figure 1.4: two plates sliding past each other

When we exert a force on the top plate, it moves at some velocity. The force necessary to reach a certain velocity  $v$  is increasing with  $v$ , increasing with the total size (or area  $A$ ) of the plates, but decreasing as we increase the distance  $d$  between the plates. For very small values of all of these, the result should be

$$F = \eta \frac{Av}{d} \quad (1.6)$$

where we have introduced a constant of proportionality  $\eta$ . We must do this in order to keep from comparing apples to bananas; that is, any mathematical model of our physical world must be *dimensionfully sound*. The bookkeeping to make sure this is true, and the implications of this bookkeeping, are sometimes called *dimensional analysis*.

### 1.2.6 Dimensional analysis: $\mu$ is not $b$

Dimensional analysis will not be entirely new. If someone tells you that distance = rate  $\times$  time, and distance is measured in miles, and time is measured in hours, you will expect that the rate will be quoted to you in *units* of miles per hour.

More generally, whether it is in miles per hour or parsecs per fortnight, it will be some unit of distance divided by some unit of time. Another way of saying this is that the rate will have *dimensions* of length per time. Feet, meters, and parsecs are all *units* measuring quantities which have *dimensions* of time.

A special notation used for this is brackets, for example

$[d] = \ell$ :	“distance has dimensions of <i>length</i> ”
$[r] = \ell/t$ :	“rate has dimensions of <i>length per time</i> ”
$[t] = t$ :	“time has dimensions of <i>time</i> ”

and

$$d = rt \Rightarrow [d] = [r][t]. \quad (1.7)$$

Now let's turn to the relationship we used to define viscosity:

$$F = A \frac{v}{l} \mu \Rightarrow [\mu] = \frac{Et}{\ell^3} \quad (1.8)$$

Similarly, it's useful to ponder the dimensions of our discussion of temperature:

$$[k_B] = E/T \rightarrow [k_B T] = E. \quad (1.9)$$

$k_B T$  measures an energy.

At some point when you were told  $F = ma$ , in your physics classes, you might have wondered how this is possible when, in our day-to-day lives, usually things stop moving when you push on them. In physics, this is often modeled by defining a restorative drag force  $F = -bv$ , proportional to the velocity at which something moves. (This fact was not lost on Aristotle, by the way, who proposed this algebraic relation.)

We can see, though, that this  $b$  can not be merely  $\mu$ , since

$$[b] = [F]/[v] = Et/\ell^2 \quad (1.10)$$

so some magic length must relate  $b$  (a property of the grain-in-water) to  $\mu$  (a property of the water). The length scale, of course, is  $a$  the radius of the bead:

$$b = c\mu a. \quad (1.11)$$

The fluid mechanician Stokes solved for the (dimensionless) constant  $c = 6\pi$ .

This is consistent with our intuition: a pea should be easier to drag through molasses than an olive or a tennis ball.

### 1.2.7 A dimensionful model

#### drag

in a viscous liquid, newton tells us that

$$m \frac{dv}{dt} = -bv, \quad (1.12)$$

which admits solutions of the form

$$v(t) = v_0 e^{-\frac{b}{m}t}. \quad (1.13)$$

$\tau = \frac{m}{b}$  gives a natural “deceleration time” for the system. Putting in numbers for something of density of water and of size approximately that of a cell,

$$\tau = \frac{\rho a^3}{6\pi\mu a} \propto \frac{\rho a^2}{\mu} \quad (1.14)$$

In cgs units,  $\rho = 1$  and  $\mu = .01$ , and 1 micron is  $a = 10^{-4}$ . From these,  $\tau \approx 10^{-6}s$ , meaning an E. coli should be able to coast, after turning off its flagellar motor, for about  $10^{-6}$  seconds.

#### Forcing

Brown observed a staistical relationship of the excursion as a function of time  $x(t)$  of the pollen grains from the origin  $x = 0$ . You can reproduce this effect for yourself using the AIP’s java applet, or write your own simulation of it. If I start many pollen grains at time  $t = 0$  at the origin  $x = 0$ , they will spread in a cloud whose size  $\sigma^2 = Dt$  where  $[D] = \ell^2/t$ .

What Einstein showed (in his PhD thesis) was how to relate  $D$  to the infinitesimal thermal energy  $k_B T$  and the infinitesimal drag on a pollen grain  $b$ :

$$D = T/b = T/(6\pi\eta a) \quad (1.15)$$

a relationship now called the *Stokes-Einstein* relationship.

example problem:

$$mv_t = -bv \quad (1.16)$$

### Question:

what is  $t$  for a cell?

### Question:

what is  $\ell$  for a cell moving at 30 times its own length per second?

## 1.2.8 A digression: physics not at the scale of the cell

QM

$$t_{QM} \equiv h/k_B T = 1.6 \cdot 10^{-13} s \quad (1.17)$$

EM

CM

SM

## 1.3 Diffusion

One nearly-universal feature of diffusive motion is a simple relationship between the observed variance  $\sigma^2$  in the locations of each in an ensemble of pollen grains as a function of the time since they were all released (for simplicity's sake, let's assume they were all released from location  $x = 0$  at time  $t = 0$ ). Clearly  $\sigma(0) = 0$  — if we haven't waited any time, we know exactly where each pollen grain will be: wherever it is now. Brown observed that the uncertainty, measured as a squared distance  $\sigma^2$ , (or if you like an error bar) grows *linearly* in time. The constant of proportionality is denoted  $D$ , a parameter quantifying the "strength" of the diffusion, *i.e.*,

$$\sigma^2 = Dt, \quad (1.18)$$

and is called the *diffusion constant*.

We can discuss this in terms of statistics, defining  $\langle \dots \rangle$  as the average over pollen grains. If there is no difference between "right" and "left", then even at later times we must have  $\langle x \rangle = 0$ . Eqn. 1.18 can be expressed in terms of the mean squared displacement:  $\langle x^2 \rangle = Dt$ .

**Question:**

What are the dimensions of  $D$ ?

**1.3.1 Diffusion:  $D = k_B T / b$** 

Both  $k_B T$  and  $b$  are of the same origin: the coarse-grained representation of collisions with  $10^{23}$  or so water molecules, approximately one per picosecond.

**Dimensional Analysis, dimensional estimation, and Fermi problems**

“Usually only the simplest tools, like dimensional analysis, suffice to see what’s going on” (PCN p. xvi)

**Question:**

Calculate the dimensions of D from the diffusion equation

in class example: how high can you jump?

in class example: estimate D

**Homework:**

1. dimensional estimation: kinesin takes 1000 steps per second at room temperature. How important is quantum mechanics? (ask me if you need a hint)
2. dimensional analysis: what is the velocity of a protein at room temperature? (ask me if you need a hint)
3. dimensional analysis: if  $F = z \cdot v$ , and  $D$  has dimensions of  $\ell^2 / t$ , how can I combine  $z$  and  $D$  into something with units of energy? Einstein used this argument to find the size of atoms in 1905.

**1.3.2 “Random walks in biology”**

<sup>1</sup> The random walk model is one of the most simple and general ways to begin thinking concretely about probability. Let’s first consider a one-dimensional excursion of a pollen grain under Dr. Brown’s microscope. The grain begins at the origin  $x = 0$  at time  $t = 0$ . Dr. Brown looks under the microscope repeatedly, taking note of the location  $x_t$  the  $t^{\text{th}}$  time he looks (we are here considering time indexed by the dimensionless integer  $t$ ); each recording is separated by a small time  $dt$ . As a simple model, consider describing the displacements  $x_t - x_{t-1}$  statistically, that is, rather than specifying a particular set of values for these displacements, let us simply state some information about their expected values, from which we may compute the dynamics of the expected moments of  $x_t$ .

To begin, let us model the displacements as being one of two values,  $-\{\ell, \ell\}$ , or equivalently

---

<sup>1</sup>The subsection title is that of H. Berg’s excellent book [6]

1. let  $\eta_t \in \{-1, 1\}$  indicate which direction the grain moves after the  $t^{\text{th}}$  observation, i.e.  $x_{t+1} = x_t + \ell\eta_t$ ,
2.  $x_t$  and  $\eta_t$  are independent, thus  $\langle x_t \eta_t \rangle = \langle x_t \rangle \langle \eta_t \rangle$ ,
3.  $\eta_t$  satisfies  $\langle \eta_t \rangle = 0$ .

From this we may calculate the expected values of  $\langle x_t \rangle$  and  $\langle x_t^2 \rangle$ :

$$\langle x_t \rangle = \ell \sum_{t'=0}^t \langle \eta_{t'} \rangle = \ell \sum_{t'=0}^t 0 = 0 \quad (1.19)$$

$$\langle x_t^2 \rangle = \langle (x_t + \ell\eta_t)^2 \rangle = \langle x_t^2 \rangle + 2\ell\langle x_t \eta_t \rangle + \ell^2 \langle \eta_t^2 \rangle \quad (1.20)$$

$$= \langle x_t^2 \rangle + 2\ell\langle x_t \rangle \langle \eta_t \rangle + \ell^2 \langle 1 \rangle \quad (1.21)$$

$$= \langle x_t^2 \rangle + \ell^2 \quad (1.22)$$

$$\langle x_t^2 \rangle = \sum_{t'=0}^t \ell^2 = \ell^2 t. \quad (1.23)$$

We may write this in terms of the total time  $\tau = t \cdot dt$  of the experiment:

$$\langle x_t^2 \rangle = (\ell^2 / dt) \tau \equiv D\tau \quad (1.24)$$

where the diffusion constant  $D$  is a function of the time  $dt$  between observations and the typical magnitude of an excursion  $\ell$ , given this choice of  $dt$ .

### Biological application of diffusive modeling: DNA

This simple calculation is at the heart of innumerable stochastic models from fields as diverse as option pricing and polymer physics. As an illustration of the latter, consider a polymer composed of  $N$  small links, each of length  $\ell$ ; the  $t^{\text{th}}$  link is oriented according to the unit vector  $\hat{\mathbf{r}}_t$ , and  $\mathbb{R}_t$ , the location of the  $t^{\text{th}}$  monomer, is related to the previous monomer by

$$\mathbb{R}_{t+1} = \mathbb{R}_t + \ell\hat{\mathbf{r}}_t \quad (1.25)$$

from this the location of the final monomer  $\mathbb{R}_t$ , relative to the initial monomer  $\mathbb{R}_0$ , is given by

$$\mathbb{R}_t = \ell \sum_{t'=1}^N \hat{\mathbf{r}}_{t'} \quad (1.26)$$

To compute its typical size, we calculate its magnitude as

$$\langle R_t^2 \rangle \equiv \langle \mathbb{R}_t \cdot \mathbb{R}_t \rangle = \left\langle \sum_{t'=0}^N \ell \hat{\mathbf{r}}_{t'} \cdot \sum_{t''=0}^N \ell \hat{\mathbf{r}}_{t''} \right\rangle = \ell^2 \sum_{t'=0}^t \sum_{t''=0}^t \langle \hat{\mathbf{r}}_{t'} \cdot \hat{\mathbf{r}}_{t''} \rangle. \quad (1.27)$$

By constructing the calculation in terms of the sum, we must appeal to a slightly different statement of independence. For the pollen grain, we stated  $\langle \eta_t x_t \rangle = \langle \eta_t \rangle \langle x_t \rangle$ ; the analogous observation to make here is that distinct links are independent, such that

$$\langle \hat{\mathbf{r}}_{t'} \cdot \hat{\mathbf{r}}_{t''} \rangle = \delta_{t', t''} \equiv \begin{cases} \langle \hat{\mathbf{r}}_{t'} \rangle \cdot \langle \hat{\mathbf{r}}_t \rangle = 0, & t' \neq t'' \\ \langle \hat{\mathbf{r}}_{t'}^2 \rangle = 1, & t' = t'' \end{cases}. \quad (1.28)$$

Since each  $\hat{\mathbf{r}}_t$  is a random vector,  $\langle \hat{\mathbf{r}}_t \rangle = 0$ ; since each is a unit vector,  $\langle \hat{\mathbf{r}}_t^2 \rangle = \langle 1 \rangle = 1$ . From this,

$$\langle \mathbb{R}_t^2 \rangle = \ell^2 \sum_{t'=0}^t = t\ell^2. \quad (1.30)$$

Just as we wrote the pollen grain excursion in terms of the total time  $\tau$  of the experiment, here we may write the polymer diameter  $R_G \equiv \sqrt{\langle R_t^2 \rangle}$  in terms of the total backbone length  $L \equiv t\ell$ :

$$R_G = \sqrt{L\ell}, \quad (1.31)$$

meaning the diameter grows with the *square root* of the total length, just as the excursion of the pollen grain grows with the *square root* of the total time of the experiment.

Figure 1.5: freely-jointed chain

## 1.4 Probability and probability distributions

### 1.4.1 What just happened?

In Sec. 1.3.2, we calculated the dynamics for one particular statistic: the variance. But we can also work out an expression, beginning with the same model, of the entire probability distribution.

#### Question:

Write  $\langle \eta \rangle$  in terms of a probability

**A:**

$$\langle \eta \rangle = \sum_s p_s \eta_s = (1/2)(-1) + (1/2)(1) = 0$$

#### Question:

What properties must a function  $f(x, y, z)$  satisfy to be a probability  $p(x, y, z)$  in  $x, y, z$ ?

**A:**

Two properties:  $f(x, y, z) \geq 0$  and  $\int dx dy dz f(x, y, z) = 1$  (where the integral ranges over all possible values of  $(x, y, z)$ ).

**Question:**

Probability frequently uses the stunning abuse of notation  $p(x, y, \dots)$  with the understanding that  $p(x)$  is not necessarily the same function  $p(y)$ ; that is, the dependent variable is used to indicate which function is being discussed. Respecting this choice, how is  $p(x)$  related to  $p(x, y, z)$ ?

**A:**

$$p(x) = \int dy dz p(x, y, z).$$

**Question:**

How is  $p(y)$  related to  $p(x, y)$  and  $p(x|y)$ ?

**A:**

$$p(x, y) = p(x|y)p(y). \text{ Note that by symmetry, } p(x|y)p(y) = p(y|x)p(x).$$

**Question:**

Use the above to derive  $p_{t+1}^n$ , the probability that a pollen grain sits at the site  $n$  at time  $t + 1$ , in terms of  $p(n|n', \eta_t)$ , the probability the pollen grain sits at site  $n$  at a subsequent time given that it is sitting at site  $n'$  at the current time,  $p(\eta_t)$ , and  $p_t^n$ .

**A:**

Here is where we actually use the model — i.e., the dynamics  $x_{t+1} = x_t + \ell\eta_t$ . Specifically, the dynamics implies

$$p(n|n', \eta_t) = \delta_{n, n' + \ell\eta_t} \quad (1.32)$$

Then, turning the definitional crank,

$$p_{t+1}^n = \sum_{n'} \sum_{\eta=\pm 1} p(n|n', \eta) p_t^{n'} p(\eta) \quad (1.33)$$

$$= p_t^{n-1} p(\eta = 1) + p_t^{n+1} p(\eta = -1) \quad (1.34)$$

$$= \frac{1}{2}(p_t^{n-1} + p_t^{n+1}) \quad (1.35)$$

**Question:**

How is the statistic  $\langle x_t^2 \rangle$  related to a distribution  $p_t^n$ ?

**A:**

We identify  $x = \ell n$ ; then  $\langle x_t^2 \rangle \equiv \sum_n x^2 p_t^n = \ell^2 \sum_n n^2 p_t^n$

**Question:**

Use the above two answers to derive  $\langle x_{t+1}^2 \rangle$  in terms of  $\langle x_n^2 \rangle$ . Defining the dummy variables of summation  $u \equiv n + 1, v \equiv n - 1$ ,

$$\langle x_{t+1}^2 \rangle = \ell^2 \sum_n n^2 p_{t+1}^n \quad (1.36)$$

$$= \frac{1}{2}\ell^2 \sum_n n^2 p_t^{n-1} + \ell^2 \frac{1}{2} \sum_n n^2 p_t^{n+1} \quad (1.37)$$

$$= \ell^2 \frac{1}{2} \sum_v (v+1)^2 p_t^v + \frac{1}{2} \sum_u (u-1)^2 p_t^u \quad (1.38)$$

$$= \ell^2 \frac{1}{2} \sum_n ((n+1)^2 + (n-1)^2) p_t^n \quad (1.39)$$

$$= \langle x_t^2 \rangle + \ell^2 \quad (1.40)$$

### Question:

Independence of  $a$  and  $b$  has a precise meaning in probability:  $p(a,b) = p(a)p(b)$ . Use the independence of continuous  $a$  and discrete  $b \in \{-1, 1\}$  to show that the expectation of their product is the product of their expectations

**A:**

$$\langle ab \rangle = \int da \sum_{b=\pm 1} ab p(a, b) \quad (1.41)$$

$$= \int da p(a) \sum_{b=\pm 1} bp(b) \quad (1.42)$$

$$= \langle a \rangle \langle b \rangle \quad (1.43)$$

Figure 1.6: Pascal’s triangle/diffusion dynamic

### 1.4.2 The discrete case

Since we are considering only a discrete set of values of  $n$ , we have used a notation for a distribution defined over the integers (in the spatial variable  $n = x/\ell$ ) and the nonnegative integers  $t$  (in the time variable  $t = \tau/dt$ ):  $p_t^n$  = probability grain sits at location  $x = n\ell$  at time  $\tau = t \cdot dt$ .

Above, we derived

$$p_t^n = \frac{1}{2}(p_{t-1}^{n+1} + p_{t-1}^{n-1}). \quad (1.44)$$

Let’s consider the solution if the pollen grain starts at the origin. This specifies  $p_0^n = \delta_{n,0}$ .

Building Pascal’s triangle, we see that we may express this in terms of  $P_t^j$ , the probability that the pollen grain has moved to the right  $j$  times (and therefore moved to the left  $t-j$  times, such that its location  $n = j - (t-j) = 2j - t$ ):

$$P_t^j = \binom{t}{j} p^j q^{(t-j)} \quad (1.45)$$

where “ $q$ ” and “ $p$ ” indicate the probabilities of stepping left and right, respectively (which must sum to 1).

### Question:

Derive  $\langle x \rangle$  from  $P_t^j$ .

**A:**

Since the number of times the grain moves right ( $n_r$ ) and the number of times the grain moves left ( $n_l$ ) must sum to  $t$ ,  $n_r + n_l = t$ ,  $\langle x \rangle = \ell \langle (n_r - n_l) \rangle = \ell \langle j - (t - j) \rangle = \ell (2\langle j \rangle - t)$ .

$$\langle j \rangle = \sum_j j P_t^j \text{sum}_j j \binom{t}{j} p^j q^{(t-j)} = p \partial_p (p+q)^t = pt(p+q)^{t-1} \quad (1.46)$$

and  $\langle x \rangle = t\ell(2p - 1) = 0$ .

### Question:

Derive  $\langle x^2 \rangle$  from  $P$ . Similarly,

$$\langle x^2 \rangle = \ell^2 \langle (n_r - n_l)^2 \rangle = \ell^2 \langle (2j - t)^2 \rangle = \ell^2 (4\langle j^2 \rangle - 4t\langle j \rangle + t^2) \quad (1.47)$$

$$\langle j^2 \rangle = (p \partial_p)^2 (p+q)^t = pt(p+q)^{t-1} + p^2 t(t-1)(p+q)^{t-2} \quad (1.48)$$

$$\langle x^2 \rangle = \ell^2 (4(t/2 + t(t-1)/4) - 4t^2/2 + t^2) = \ell^2 t. \quad (1.49)$$

### Question:

Express  $\langle j^m \rangle$  in terms of  $(p+q)^t$  for arbitrary integer  $m$

**A:**

$\langle j^m \rangle = (p \partial_p)^m (p+q)^t|_{q=1-p}$ . Because  $\langle j^m \rangle$  is called “the  $m^{\text{th}}$  moment of  $j$ ”, the function  $Z(p) \equiv (p+q)^t$  is called a “moment generating function”.

### 1.4.3 The continuum limit

As we showed above, the update rule for discrete time and discrete space,

$$p_{t+1}^n = \frac{1}{2} p_t^{n-1} + \frac{1}{2} p_t^{n+1} \quad (1.50)$$

can be used to derive a binomial distribution  $P_t^j$  for the number  $j$  of moves to the right. However, on our length scales, we see Brown's pollen grains as a continuous density  $\rho(x, \tau)$ , and regard the small length  $\ell$  and time  $dt$  as modeling and calculational tools only. To give further insight, let's consider the continuum limit with real-valued  $x$  and  $\tau$ .

The limit is derived by noticing that the update rule can be rewritten in terms of temporal and spatial differences:

$$p_{t+1}^n - p_t^n = \frac{1}{2}(p_t^{n-1} - p_t^n) - \frac{1}{2}(p_t^n - p_t^{n+1}) \quad (1.51)$$

which in turn suggests replacing the small differences with their continuum approximations, that is, derivatives, since:

$$dt\partial_\tau\rho(x, \tau) \approx (1/\ell)p_{t+1}^n - p_t^n \quad (1.52)$$

$$\ell\partial_x\rho(x, \tau) \approx (1/\ell)p_t^{n-1} - p_t^n \quad (1.53)$$

where we have identified  $\rho(x = n\ell, \tau = t \cdot dt) = p_t^n/\ell$ . The update rule is then approximated

$$dt\partial_\tau\rho(x, \tau) = \frac{1}{2}\ell\partial_x\ell\partial_x\rho \quad (1.54)$$

$$\partial_\tau\rho(x, \tau) = \ell^2/(2dt)\partial_x^2\rho \equiv D\partial_x^2\rho, \quad (1.55)$$

a *partial* differential equation known as the heat equation or diffusion equation.

A notational shorthand often used is to use dots to indicate temporal derivatives and primes to indicate spatial derivatives, that is:

$$\dot{\rho} \equiv \partial_\tau\rho \quad (1.56)$$

$$\rho' \equiv \partial_x\rho \quad (1.57)$$

in terms of which the diffusion equation reads

$$\dot{\rho} = D\rho'' \quad (1.58)$$

The initial condition  $p_t^n = \delta_n$  becomes  $\rho(x, \tau) = \delta(x)$ , whose solution is the normal distribution:

$$\rho(x, \tau) = \frac{e^{-\frac{1}{2}x^2/\sigma^2}}{\sqrt{2\pi\sigma^2}} \quad (1.59)$$

where  $\sigma^2 = 2D\tau$ . If this is not clear to you, you might want to refresh your PDE background with Appendix ??.

### Homework:

1. Use Stirling's approximation (*cf.* Appendix F) to approximate the binomial distribution

$$p_t^j = \frac{1}{2^t} \binom{t}{j} \quad (1.60)$$

when  $t \gg 1$  and  $x = t/2 + y$ ,  $y \ll t$ .

2. Show that the Gaussian

$$p(x, t) = \frac{e^{-x^2/(4Dt)}}{\sqrt{4\pi Dt}} \quad (1.61)$$

solves the diffusion equation

$$\partial_t p = D \partial_x^2 p \quad (1.62)$$

This solution may also be written as

$$p = \frac{e^{-\frac{1}{2}(x/\sigma)^2}}{\sqrt{2\pi\sigma^2}} \quad (1.63)$$

for  $\sigma^2 = 2Dt$ .

3. Estimate the stall time (deceleration time) for a bacterium. Use this to estimate the coasting distance if  $v \approx 1\mu/s$ .

#### 1.4.4 Transport: $\partial_t \rho = -\partial_x \rho v$

If you imagine a density of deterministic walkers, all moving with  $v = v(x)$ , then the rate of change of the number in some small window  $N \equiv \int_a^b dx \rho$  must obey

$$\partial_t N = \int_a^b dx \partial_t \rho = -\rho(b)v(b) + \rho(a)v(a) \quad (1.64)$$

or, expressed as a local condition,

$$\partial_t \rho = -\partial_x \rho v. \quad (1.65)$$

The term  $\rho v$  is sometimes called the “flux” and denoted  $J \equiv \rho v$ .

Note that, for this to be dimensionfully sound,  $\rho$  must have dimensions of  $[N]/[x] = 1/\ell$ .

Figure 1.7: Rate of change in  $N$  is the rate at which stuff goes into the box minus the rate of stuff leaving the box. Since velocity is positive when things are moving to the right, we have Eqn. 1.64.

#### 1.4.5 Conservation laws: $\partial_t \rho = -J'$

#### 1.4.6 Diffusion and drift $J = \rho v - D \partial_x \rho$

The general condition is a total flux from both deterministic drift and diffusion:

$$J = J_{\text{diffusion}} + J_{\text{drift}} = \rho v - D \partial_x \rho \quad (1.66)$$

or

$$\partial_t \rho + \partial_x \rho v = D \partial_x^2 \rho \quad (1.67)$$

from the conservation law

$$\partial_t \rho = -\partial_x J. \quad (1.68)$$

This is called a “conservation law” because it makes clear that the total amount of  $\rho$  is conserved if the flux  $J$  vanishes at boundaries or if  $J$  is a periodic function of  $x$ . That is, differentiating the normalization condition with respect to time,

$$1 = \int_a^b dx \rho_0 = \int_a^b dx \partial_\tau \rho = -J_b + J_a. \quad (1.69)$$

which vanishes if  $J_b = J_a$ , including as special cases  $J_a = J_b = 0$ , including as special cases  $a \rightarrow -\infty$   $b \rightarrow +\infty$  with  $\rho, \rho' \rightarrow 0$  as  $x \rightarrow \pm\infty$  (e.g., for diffusing pollen grains over the real line). Note that for a periodic domain, where  $a$  and  $b$  are separated by one period (the systems size), then  $J_a = J_b$  and conservation is enforced.

#### 1.4.7 Application of diffusion with drift: the Chemotaxis story

questions: how do bacteria get anywhere? how do they get food?

- some proteins regulate who goes in and out of the cell
- triggers network of proteins
- ejects protons, turning motor proteins around [page cut off]

bacteria samples food and takes a “time derivative” which regulates the length in time of the “tumble” (time during which motor proteins aren’t working together). shorter tumble = more time for run, get to food.

#### leftover homewework:

How many cells make up you?

Go to the RCSB protein data bank (<http://www.rcsb.org/>) and look up your favorite protein. If you don’t have one yet, try “kinesin”, or “Gene Regulating Protein”. How big is it? (in volume? in mass? in Daltons?) How many proteins could fit in a cell?

Convert  $k_B T$ , at room temperature, into picoNewtons and nanometers

what is the “molar concentration” of 1 object in a bacterium? (a concentration is just a number of objects per volume – irrespective of what the object is)

#### leftover homework:

Estimate how high you jump if you stay in the air for 1 second. (hint: use the natural acceleration given by gravity)

Estimate the stall force of kinesin given that it takes 8 nm steps (hint: use the natural energy given by room temperature)

Estimate stall time for a bacterium (I shouldn't call this a "stall time" exactly, it's more of a "deceleration time". Use the fact that it's decelerating with a force given by the drag force  $F = -bv$ , and this balances  $mass * (dv/dt)$ . Then estimate the size of the exponential decay time. Assume a spherical bacterium made of water, and a drag constant given by Stokes relation (as we discussed in class).)

Estimate coasting distance for a bacterium, assuming it is swimming initially at its own body length/second, and then comes to rest over one stall time



## Chapter 2

# Equilibrium Statistical Mechanics: $J = 0$

If we consider the statistical steady state  $\partial_t \rho = 0$ , then flux must be a constant:  $\partial_x J = 0$ . For a closed system in which flux vanishes at the boundaries

$$J = 0 \Rightarrow D\partial_x \rho - \rho v = 0 \quad (2.1)$$

(which is of the form  $\dot{x} = \alpha x$ ) thus

$$\frac{d\rho}{\rho} = \frac{vdx}{D} \quad (2.2)$$

$$\ln \rho/\rho_0 = \frac{1}{D} \int_0^x dx' v(x'). \quad (2.3)$$

In an overdamped, “Aristotelian” world, in which velocity is proportional to force, ( $v \propto F$ ) we define the drag constant as  $\zeta$ , and express the force in terms of a potential energy

$$\zeta v = -\frac{dU}{dx} \quad (2.4)$$

to find

$$\ln \frac{\rho}{\rho_0} = -\frac{1}{D} \frac{U}{\zeta} \quad (2.5)$$

thus  $\rho$  is the Boltzmann distribution :

$$\rho = \rho_0 e^{-\beta U} \quad (2.6)$$

with

$$\beta = \frac{1}{D\zeta}. \quad (2.7)$$

In our earlier discussion we argued for a characteristic deceleration time  $\Delta t = m/\zeta$ , and distance  $(\Delta x^2) = (\Delta t)^2 \langle v^2 \rangle = (T/m)(\Delta t)^2 = Tm/\zeta^2$  (from the characteristic thermal velocity), which would argue  $D \sim T/\zeta$ . However, we can also simply take this as the definition of temperature.

$$\beta = \frac{1}{D\zeta} \equiv \frac{1}{T}. \quad (2.8)$$

This relation is called “the Stokes Einstein relation” and is your first example of a “fluctuation-dissipation relation.”

## 2.1 Biological applications

### 2.1.1 Pollen grains ( $U \propto x$ )

example: pollen grains in closed vertical tube, gravitational potential  $U(y)$ , and drag constant  $b$ . Newton’s 2nd law gives

$$v(y) = -\frac{1}{b} \frac{dU}{dy} \quad (2.9)$$

since the tube is closed,  $J = 0$ , thus 2.1 gives

$$-D \frac{\partial \rho}{\partial x} = v\rho \quad (2.10)$$

which integrates to

$$\rho(y) = \rho_0 e^{-\frac{1}{bD} \int dy v(y)} \quad (2.11)$$

substituting 2.9 for  $v(y)$ , we arrive at

$$\rho(y) = \rho_0 e^{-\frac{1}{bD} U(y)}, \quad (2.12)$$

the Boltzmann distribution where  $bD = k_B T$ .

Note that this distribution suggests a moment-generating function, since, defining  $\alpha = mg/k_B T$ ,

$$\langle y^m \rangle = \rho_0 \int_0^\infty y^m e^{-mg y / k_B T} dy \quad (2.13)$$

$$= \rho_0 (-\partial_\alpha)^m \int_0^\infty dy e^{-\alpha y} \quad (2.14)$$

### 2.1.2 Calibrating an optical trap ( $U \propto x^2$ )

There are also many biophysical applications for which it is useful to study a quadratic energy, e.g., the energy of deviation from the center of an optical trap.

An optical trap is the focal point of a tightly focused laser beam. Just as a dielectric is “forced” into a capacitor, so is a glass bead drawn to the maximum of an electric field. (more discussion needed....)

defining  $\alpha \equiv k/k_B T$ ,

$$\rho = \rho_0 e^{-U/k_B T} = \rho_0 e^{-\frac{1}{2}kx^2/k_B T} \quad (2.15)$$

$$1 = \rho_0 \int_{-\infty}^{+\infty} dx e^{-\frac{1}{2}kx^2/k_B T} \quad (2.16)$$

$$\Rightarrow \rho = \frac{e^{-\frac{1}{2}kx^2/k_B T}}{\int_{-\infty}^{+\infty} dx e^{-\frac{1}{2}kx^2/k_B T}} \quad (2.17)$$

$$\langle x^2 \rangle = \frac{\int_{-\infty}^{+\infty} dx x^2 e^{-\frac{1}{2}\alpha x^2}}{\int_{-\infty}^{+\infty} dx e^{-\frac{1}{2}\alpha x^2}} \quad (2.18)$$

$$= -2\partial_\alpha \log \int_{-\infty}^{+\infty} dx e^{-\frac{1}{2}\alpha x^2} \quad (2.19)$$

$$\int_{-\infty}^{+\infty} dx e^{-\frac{1}{2}\alpha x^2} = \alpha^{-1/2} \int_{-\infty}^{+\infty} du e^{-\frac{1}{2}u^2} \quad (2.20)$$

$$\Rightarrow \langle x^2 \rangle = -2\partial_\alpha \ln \alpha^{-1/2} = \partial_\alpha \ln \alpha = \alpha^{-1} = k_B T/k \quad (2.21)$$

### 2.1.3 The Ising model of DNA extension ( $U \propto \pm 1$ ):

While the energy is easiest to state here, it's actually hardest to calculate.

#### History

single molecule biophysics

can construct one-dimensional model where DNA is made of links of length  $l$ , given by the persistence length of DNA, that can link to the right or left.

Now is a good time, if you haven't done it already, to calculate  $\lambda$ -phage's  $\sqrt{L\ell}$  and compare to the size of a virus.

#### modeling

the expected value for length of the chain is given by

$$\langle x \rangle = \frac{\sum_s x_s P_s}{\sum_s P_s} \quad (2.22)$$

where  $s = \pm$  indicates right or left and  $x_s = \pm l$ . using a Boltzmann distribution for  $P_s$  with an applied force  $F$ , we have

$$\langle x \rangle = \frac{\sum_s x_s e^{-\beta Fl}}{\sum_s e^{-\beta Fl}} = \frac{l e^{-\beta Fl} - l e^{\beta Fl}}{e^{-\beta Fl} + e^{\beta Fl}} = l \tanh(\beta Fl) \quad (2.23)$$

if one examines the partition function

$$Z(T) \equiv \cosh(\beta Fl) \quad (2.24)$$

it can be seen that

$$\frac{\partial Z}{\partial T} = -\frac{Fl}{k_B T^2} \sinh(\beta Fl), \quad (2.25)$$

arriving at the more general result:

$$\langle x \rangle = -\frac{k_B T^2}{F} \frac{\partial \ln Z}{\partial T}. \quad (2.26)$$

note: \* moment-generating functions in statistics \* moment-generating functions in physics

at low force, like a spring, no?

e.g., springs ( $U = x^2$ )  $\langle x \rangle$ ; the Gaussian integral the Gaussian, or two...

## 2.2 The partition function

The above several problems illustrated one particular type of moment-generating function, called a *partition function*: for any algebraic expression  $U(x)$ , we define  $\beta = 1/k_B T$  and introduce

$$Z(\beta) \equiv \int_{\Omega} dx e^{-\beta U(x)} \quad (2.27)$$

where  $\Omega$  is the set of allowed possible values of  $x$  (e.g.,  $0 < x < +\infty$  for the pollen grain;  $-\infty < 0 < \infty$  for the bead in an optical trap). The utility of  $Z(\beta)$  is that, since  $p = e^{-\beta U}/Z$ , we have

$$\langle U(x) \rangle = -\partial_{\beta} \ln Z(\beta). \quad (2.28)$$

Several of the above calculations can be related to this, since, for example, the height of the pollen grain  $\langle y \rangle$  can be derived as (defining  $u \equiv \beta mgy$ )

$$\langle y \rangle = \langle U \rangle / mg = \frac{-1}{mg} \partial_{\beta} \ln \int_0^{\infty} dy e^{-\beta mgy} \quad (2.29)$$

$$\langle U \rangle = -\partial_{\beta} \ln \frac{1}{\beta mg} \int_0^{\infty} du e^{-u} = -\partial_{\beta} (\ln \beta^{-1} + \ln C) = \beta^{-1} = k_B T \quad (2.30)$$

$$\langle y \rangle = k_B T / (mg). \quad (2.31)$$

Or, for the trapped bead, with  $U = \frac{1}{2} kx^2$ , (defining  $u \equiv \sqrt{\beta k}x$ ),

$$\langle x^2 \rangle = \langle U \rangle / (k/2) \quad (2.32)$$

$$\langle U \rangle = -\partial_{\beta} \ln \int dx e^{-\beta \frac{1}{2} kx^2} = -\partial_{\beta} \ln (\beta k)^{-1/2} \int du e^{-\frac{1}{2} u^2} = -\partial_{\beta} \ln (\beta^{-1/2} C) = \frac{1}{2} k_B T \quad (2.33)$$

$$\langle x^2 \rangle = k_B T / k \quad (2.34)$$

The partition function is very helpful in physical applications, when we are interested in calculating the expected value of an energy, or of functions related to the energy such as its derivative. For example, the specific heat

$$C \equiv \frac{d\langle U \rangle}{dT} = -\frac{1}{T^2} \partial_{\beta} \langle U \rangle = \beta^2 \partial_{\beta}^2 \ln Z. \quad (2.35)$$

## 2.3 Linear response theory

For the Ising model of DNA, note that, for 0 force, we should recover the statistics of the random walk, that is

$$\langle x \rangle = 0 \langle x^2 \rangle = \ell^2 \quad (2.36)$$

Since  $\tanh(x) \rightarrow x$  for  $x \ll 1$ , we have

$$\langle x \rangle_F \approx \beta F \ell^2 = (F/k_B T) \langle x^2 \rangle_0 \quad (2.37)$$

where the subscript emphasizes the expectation value as a function of force  $F$ , rather than for 0 force as in  $\langle x^2 \rangle_0 = \ell^2$ . Qualitatively, we have that in the small force, Hookian regime, the spring constant is given by  $k_B T / \sigma^2$  where  $\sigma^2$  is the size of fluctuations at zero force. This is a particularly useful result, for example, in biological problems where we can measure  $\langle x \rangle$  and  $\langle x^2 \rangle$  more easily than an applied  $F$ . Or, for example, in complicated models when we wish to estimate a small  $\langle x \rangle$  at fixed  $F$  (that is, when the model is too complicated to calculate  $\langle x \rangle(FF)$  for all  $F$ ).

This is a special case of a far more general result called linear response, namely, for arbitrary  $U(x)$ , if we impose a force  $F$  on the system, we find for small  $F$

$$\langle x \rangle_F \equiv \frac{\int dx x e^{-\beta(U(x)-Fx)}}{\int dx e^{-\beta(U(x)-Fx)}} \quad (2.38)$$

$$\approx \frac{\int dx x e^{-\beta U(x)} (1 + \beta Fx)}{\int dx e^{-\beta U(x)} (1 + \beta Fx)} \quad (2.39)$$

$$= \frac{\int dx x e^{-\beta U(x)} + \beta F \int dx x^2 e^{-\beta U(x)}}{\int dx e^{-\beta U(x)} + \beta F \int dx x e^{-\beta U(x)}} \quad (2.40)$$

$$= \frac{\langle x \rangle_0 + \beta F \langle x^2 \rangle_0}{1 + \beta F \langle x \rangle_0} \quad (2.41)$$

$$\approx (\langle x \rangle_0 + \beta F \langle x^2 \rangle_0)(1 - \beta F \langle x \rangle_0) \quad (2.42)$$

$$\approx \langle x \rangle_0 + \beta F(\langle x^2 \rangle_0 - \langle x \rangle_0^2) \quad (2.43)$$

$$= \langle x \rangle_0 + \beta F \sigma^2 \quad (2.44)$$

### Homework:

1.  $T$  and  $mg$  together define a characteristic length. Calculate that length for water molecules
2. Express  $\langle x^n \rangle$  in terms of  $Z$  for pollen grains
3. Calculate  $R_g$  for lambda phage DNA ( $L = 16.4\mu$  and for your DNA  $L \approx 2m/20$ ). You will need to know that  $\ell_{\text{DNA}} \approx 50nm$ .
4. Calculate  $Y$ , the Young's modulus, for DNA, actin, and microtubules. You will need to know their persistence lengths  $\ell$  (which are  $50nm$ ,  $7\mu$ , and  $3mm$ , respectively) and you will have to look up their radii (cf. Appendix E)

## 2.4 Electrostatic surprises in vivo

You might have heard by now that DNA is in fact charged: two charges per base pair, in fact. How, then, is it ever *not* simply a straight rod?

### 2.4.1 DNA and electrostatic interactions ( $U \propto \ln x$ )

background: polymer physics and DNA (examples of L/l: actin, microtubules, DNA) \* "DNA compaction" (viruses, nuclei, etc)

consider e&m interactions at  $k_B T$ . You'll recall from your physics classes the electric potential energy stored by two charges  $q_1, q_2$  at distance  $r$  (in MKS units) in free space:

$$U = \frac{q_1 q_2}{4\pi\epsilon_0 r} \quad (2.45)$$

where  $\epsilon_0$  is the "permittivity of free space"

$$k_B T = \frac{q^2}{4\pi\epsilon_0 r}, \quad (2.46)$$

from which a characteristic length known as the Bjerrum length is given by

$$l_B \sim \frac{q^2}{4\pi\epsilon_0 k_B T}. \quad (2.47)$$

if we model DNA as a line charge with linear charge density  $\lambda$  surrounded by particles with charge  $q$  we have a potential

$$V(r) \sim \lambda \ln\left(\frac{r}{r_0}\right) \quad (2.48)$$

which gives a potential energy

$$U(r) \sim q\lambda \ln\left(\frac{r}{r_0}\right). \quad (2.49)$$

dividing 2.49 by 2.46, we have

$$\frac{U}{k_B T} = \frac{q^2}{4\pi\epsilon_0 b k_B T} \ln\left(\frac{r}{r_0}\right) \quad (2.50)$$

where  $b$  is  $[]$ . now examine the expected value for  $r$  given a Boltzmann distribution,

$$\langle r \rangle = \frac{\int_a^R dr \int_0^{2\pi} d\theta r e^{-\beta U}}{\int_a^R dr \int_0^{2\pi} d\theta e^{-\beta U}}, \quad (2.51)$$

substituting 2.50 and simplifying gives (defining  $u = r/a$ ;  $\Lambda = R/a$ )

$$\langle r \rangle = \frac{\int_a^R dr \int_0^{2\pi} d\theta r^{2-\frac{l_B}{b}}}{\int_a^R dr \int_0^{2\pi} d\theta r^{1-\frac{l_B}{b}}} \quad (2.52)$$

$$\langle r \rangle / a = \frac{\int_1^\Lambda r^{2-z} dr}{\int_1^\Lambda r^{1-z} dr} \quad (2.53)$$

$$= \frac{2-z}{3-z} \frac{\Lambda^{3-z} - 1}{\Lambda^{2-z} - 1} \quad (2.54)$$

### Homework:

look at  $\lim_{a/R \rightarrow 0}$  of the above function, defining  $z \equiv \frac{l_B}{b}$ .

from last homework, should have learned that for  $z < 2$ , particles bunch up right against DNA, while for  $z > 2$ , particles are, on average, infinitely far away.

### 2.4.2 Electrostatic screening ( $U = qV \ll k_B T$ )

Screening is one of the ways electrostatics in an aqueous environment, including in vivo, is very different from electrostatics in vacuo (which is the kind of electrostatics you have probably been focusing on in Jackson, Griffiths, Haliday and Resnik, or any other textbook from which you learned electrostatics).

In the presence of salt, positive and negative ions dissociate. The free ions redistribute spatially, with each ion attracted to counterions. The energetic benefit of proximity to counterions is thwarted by entropy; the balance between the two sets a new length scale beyond which charged particles appear neutral – their charge has been “screened.”

Poisson says (ignore the  $4\pi$  if you choose different units):

$$\nabla \epsilon \nabla \phi = -4\pi \rho_Q, \quad (2.55)$$

while Boltzmann says

$$\rho_Q = (+q)\rho^+ + (-q)\rho^- = +q\rho_0 e^{-\beta q\phi} - q\rho_0 e^{+\beta q\phi} \sim \pm \rho_0 e^{\mp \beta q\phi} \quad (2.56)$$

meaning combining these in the presence of a single point charge described by  $Q\delta(\mathbf{r})$ , we have

$$\nabla^2 \phi = -\frac{4\pi q \rho_0}{\epsilon} (e^{-q\phi/k_B T} - e^{q\phi/k_B T}) + Q\delta(\mathbf{r}) \quad (2.57)$$

the first term on the right is often rewritten by noting

$$e^{-q\phi/k_B T} - e^{q\phi/k_B T} = -2 \sinh(\beta q\phi) \quad (2.58)$$

for  $T \gg q\phi$  we have

$$-\nabla^2 \phi = \frac{8\pi q^2 \rho_0}{\epsilon k_B T} \phi + (Q/\epsilon)\delta(\mathbf{r}) \quad (2.59)$$

which admits the solutions of the form<sup>1</sup>

$$\phi \sim \frac{Q}{4\pi \epsilon_0 r} e^{-\kappa r} \quad (2.60)$$

where  $\kappa^2 \equiv \frac{8\pi q^2 \rho_0}{\epsilon k_B T}$  and  $1/\kappa$  is known as the Debye length.<sup>2</sup> Note  $\kappa^2 = \rho_0 \ell_B$ .

---

<sup>1</sup>Remember that  $\nabla^2 = r^{-2} \partial_r r^2 \partial_r$  in 3D.

<sup>2</sup>a note on units: potential energy goes as  $\frac{q^2}{r}$ , so C<sup>2</sup> = J · m.

**Homework:**

(electrostatics in vivo)

1. Calculate  $\langle r \rangle$  when  $p \propto r^z$  and  $a < r < R$ . Express in the limit  $a/R \rightarrow 0$ .
2. Estimate the Bjerrum length  $\ell_B$  at room temperature.
3. Estimate the Debye screening length  $\kappa^{-1}$  at room temperature and at physiological salt concentration
4. Estimate the Manning parameter  $b/\ell_B$  for DNA.

**2.4.3 HW: linear response**

$$\langle x \rangle_F = \langle x \rangle_0 + \frac{F}{T} \sigma_0^2 \quad (2.61)$$

$$h = \{0, H\} \quad \text{height} \quad (2.62)$$

$$\langle h \rangle = \frac{H}{2} \quad (2.63)$$

$$\langle h^2 \rangle - \langle h \rangle^2 = \left[ \frac{1}{2} 0^2 + \frac{1}{2} H^2 \right] - \left( \frac{H}{2} \right)^2 = \frac{H^2}{4} = \sigma_0^2 \quad (2.64)$$

$$E = mgh \Rightarrow E_0 = mg \cdot 0 = 0 \quad (2.65)$$

$$E_H = mgH \quad (2.66)$$

ave height in presence of force

$$\langle h \rangle_F = \frac{He^{-mgH/T}}{1 + e^{-mgH/T}} \quad (2.67)$$

$$\approx \frac{H \left( 1 - \frac{mgH}{T} \right)}{1 + 1 - \frac{mgH}{T}} \quad (2.68)$$

$$= \frac{H \left( 1 - \frac{mgH}{T} \right)}{2 \left( 1 - \frac{mgH}{2T} \right)} \quad (2.69)$$

$$= \frac{H}{2} \left( 1 - \frac{mgH}{T} \right) \left( 1 + \frac{mgH}{T} \right) \quad (2.70)$$

$$\approx \frac{H}{2} \left( 1 + \frac{mgH}{T} - \frac{mgH}{2T} - \frac{(mgH)^2}{2T^2} \right) \quad (2.71)$$

$$= \frac{H}{2} \left( 1 - \frac{mgH}{2T} - \frac{(mgH)^2}{2T^2} \right) \quad (2.72)$$

$$\langle h \rangle = \frac{H}{2} - \frac{mgH^2}{4T} \quad (2.73)$$

$$= \langle h \rangle_0 - \frac{mg}{T} \sigma_0^2 \quad (2.74)$$

## Chapter 3

# Non-equilibria and transport: $J = \text{constant}$

Up until this point, ever since Chapter ??, we have been talking about the physics of dead matter. That is, by restricting ourselves to the  $J = 0$  case of diffusion with drift, we consider only systems in steady state which are ‘closed,’ with no flux going in or out.

Life, of course, is out of equilibrium. Living systems consume energy, dissipate heat, change, create, decay. Physical modeling of this begins with  $J \neq 0$ : nonequilibrium statistical physics.

Recall that we began with diffusion, incorporated drift, and found

$$\partial_t \rho = -\partial_x J. \quad (3.1)$$

Then, focusing on  $J = 0$  (appropriate to closed systems), we found equilibrium statistical mechanics. Before considering arbitrary  $J(x, t)$ , let’s first restrict ourselves to the statistical steady state:  $\partial_t \rho = 0$ , but  $J \neq 0$ .

### 3.1 Open equilibria

(PCN p420)

For a system open at either edge (e.g., a motor protein in transit, or a chemical reaction in which probability “leaks” to the stable state from the meta-stable state), we have constant  $J$ . The “absorbing” or “escaping” boundary condition for a continuous probability distribution function is  $\rho = 0$ . If we define  $x_0$  as the location where  $\rho = 0$ , we can integrate the relation (recalling  $\beta = 1/k_B T = 1/(Db)$ )

$$J = \rho v - D\rho' \quad (3.2)$$

$$J = -\rho \frac{1}{b} \frac{dU}{dx} - D\partial_x \rho \quad (3.3)$$

$$= -D(\beta \rho U' + \rho') \quad (3.4)$$

by observing (using primes for differentiation with respect to  $x$ )

$$\frac{J}{D} e^{\beta U} = -\frac{d}{dx}(\rho e^{\beta U}). \quad (3.5)$$

Integrating both sides from  $x_0$  we have

$$-(J/D) \int_{x_0}^x dx' e^{\beta U(x')} = \rho(x) e^{\beta U(x)} - \rho(x_0) e^{\beta U(x_0)} \quad (3.6)$$

$$= \rho(x) e^{\beta U(x)} \quad (3.7)$$

$$\rho(x) = -\frac{J}{D} e^{-\beta U(x)} \int_{x_0}^x e^{\beta U(x')} dx' \quad (3.8)$$

## 3.2 escape rate

Of course, this probability distribution must be *normalized* between  $x_- < x < x_+$ , thus

$$1 = \int_{x_-}^{x_+} dx \rho(x) = -\frac{J}{D} \int_{x_-}^{x_+} dx e^{-\beta U} \int_{x_0}^x dx' e^{\beta U(x')} \quad (3.9)$$

thus the signed escape time (remembering that  $J$  can be  $> 0$  or  $< 0$ ) is

$$\frac{1}{J} = \frac{-1}{D} \int_{x_-}^{x_+} dx e^{-\beta U(x)} \int_{x_0}^x e^{\beta U(x')} dx'. \quad (3.10)$$

Note that this is dimensionally sound:  $\tau \sim t \sim (dx)^2/D \sim l^2/(l^2/t) \sim t$ . It is also sign-sane. If the particle escapes at  $x_0 = x_+$  which is on the right boundary, then  $x < x_0$  in the inner integral and  $J > 0$ . Conversely, if the escape location  $x_0 = x_-$  is the left location, then  $x > x_0$  and  $J < 0$ .

### 3.2.1 Thermal ratchets: $U = -fx$ ; system size= $L$

One special case is constant force, exerted from  $x = 0$  to  $x = L$ . This is the simplest “rectified diffusion” model of a motor protein, in which the “ATP” is used to make the reaction completely irreversible (and hence  $\rho(0) = 0$ ). That is, the motor protein is trying to move to the left  $J < 0$  but the force pulls it to the right.

We then find (**HW: show this**)

$$\tau = -1/J = \frac{1}{D} \int_0^L dx e^{\beta f x} \int_0^x e^{-\beta f x'} dx' \quad (3.11)$$

$$= \frac{1}{D\beta f} \int_0^L dx e^{\beta f x} (e^{-\beta f x'} - 1) \quad (3.12)$$

$$= \frac{1}{D\beta f} \int_0^L dx (1 - e^{\beta f x'}) \quad (3.13)$$

$$= \frac{1}{D\beta f} \left( L - \frac{1}{\beta f} (e^{\beta f L} - 1) \right) \quad (3.14)$$

$$= \frac{L^2}{D} \frac{1}{\beta^2 f^2 L^2} (e^{\beta f L} - \beta f L - 1) \quad (3.15)$$

$$= \tau_{\text{diff}} g(\beta f L) \quad (3.16)$$

where the function  $g(x) = (\exp(x) - 1 - x)/x^2$  goes to  $\infty$  for large  $x$  and to  $1/2$  for small  $x$  (which you can show by Taylor expansion). This result shows that, as the force diverges, the time between steps goes to infinity (or the velocity goes to 0); in the limit of vanishing force we get the result we expect from rectified diffusion and dimensional analysis.

### 3.2.2 Kramer's escape

(aka Double-well potential, small  $T$ -limit)

A second special case is that of the Kramers' escape problem, in which we consider a double-well potential (with minima at  $x = x_-$  and at  $x = x_0 = 0$ ) and a maximum at  $x_+$ . The potentials are  $U_-$ ,  $U_+$ , and  $U_0$ .

As before, we have

$$\tau = \frac{1}{D} \int_{x_- - \epsilon}^{x_- + \epsilon} dx e^{-\beta U(x)} \int_0^x e^{\beta U(x')} dx' \quad (3.17)$$

We now consider the  $T \rightarrow 0, \beta \rightarrow \infty$  limit. In this limit the integrals are sharply dominated by the locations where the integrands are largest. For the inner integral, this is near  $x' \approx x_+$ , and we Taylor expand  $U(x') \approx U_+ + (1/2)(x' - x_+)^2 U''(x_+)$ .<sup>1</sup>

The characteristic length scales here are determined by the curvatures of the potential energies:

$$\kappa_{\{0,+,-\}}^2 \equiv \beta U''(x_{\{0,+,-\}}) \rightarrow +\infty \quad (3.19)$$

We then find, for any  $\epsilon \gg \kappa_-^{-1}$ ,

$$\tau \approx \frac{1}{D} \int_{x_- - \epsilon}^{x_- + \epsilon} dx e^{-\beta U(x)} \sqrt{\frac{2\pi}{\kappa_+^2}} e^{\beta U_+} \chi(x < x_+ < 0) \quad (3.20)$$

$$\approx \frac{1}{D} \sqrt{\frac{2\pi}{\kappa_-^2}} e^{-\beta U_-} \sqrt{\frac{2\pi}{\kappa_-^2}} e^{\beta U_+} \quad (3.21)$$

$$\approx \frac{2\pi}{D \kappa_- \kappa_+} e^{\beta(U_+ - U_-)} \quad (3.22)$$

**(HW: show this)** where the characteristic function  $\chi$  is defined here as

$$\chi(x < x_+ < 0) = \begin{cases} 1 & : x < x_+ < 0 \\ 0 & : \text{otherwise} \end{cases} \quad (3.23)$$

---

<sup>1</sup>In fact, this form (a Gaussian of vanishing width) is one of the definitions of the Dirac delta function:

$$\delta(x) \equiv \lim_{\sigma \rightarrow 0} \frac{e^{-\frac{1}{2}x^2/\sigma^2}}{\sqrt{2\pi\sigma^2}}; \quad (3.18)$$

In terms of the escape rate  $r = 1/\tau$ ,

$$r_{-0} \equiv \tau^{-1} \approx \frac{1}{2\pi} D \kappa_- \kappa_+ e^{-\beta(U_+ - U_-)}. \quad (3.24)$$

For the reverse reaction, in which  $\rho(x_-) = 0$ , we would have found

$$r_{0-} \approx \frac{1}{2\pi} D \kappa_0 \kappa_+ e^{-\beta(U_+ - U_0)}. \quad (3.25)$$

**HW:** Show that for the two-state dynamics

$$\dot{N}_- = -r_{-0} N_- + r_{0-} N_0 \quad (3.26)$$

$$\dot{N}_0 = -r_{0-} N_0 + r_{-0} N_- \quad (3.27)$$

the kinetic equilibrium (all time derivatives are 0) is the thermal equilibrium  $N_1/N_2 \propto \exp(-\beta(U_1 - U_2))$ .

old HW: 1 Limit 2 Limit

new HW 1 Calculate F vs X for the freely jointed chain 2 What is pushing against you? 3 Show Kramers escape (do the 2 Gaussian integrals) 3 show that the thermodynamic (Boltzmann) and dynamical equilibria are the same for Kramers rate, 2-state dynamics

### 3.3 Chemical dynamics; “detailed balance”

#### summary of part I (review for checkup)

- cells assemble via hydrophobicity (doesn’t require you to do work)
- things you make include: actin & microtubules, proteins (DNA, RNA)
- energy comes from: ADP  $\rightleftharpoons$  ATP + phosphate
- $k_B T = 4\text{pN} \cdot \text{nm}$  in the cell
- cell is over-damped, with Reynolds number  $10^{-6} \ll 1$
- stokes-Einstein relationship

$$D = \frac{k_B T}{6\pi\eta a} \quad (3.28)$$

#### 3.3.1 HW: Poisson equation

$$-\epsilon \nabla^2 \Phi = \rho_a + Q\delta(\mathbf{r}) \quad (3.29)$$

$$\rho_a = \rho_a^+ + \rho_a^- = q\rho_0 2 \sinh\left(\frac{q\Phi}{k_B T}\right) \quad (3.30)$$

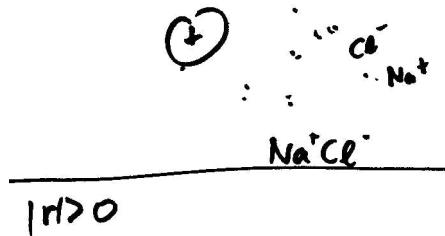


Figure 3.1: salt and large charged object

$$-\epsilon \nabla^2 \Phi \approx q \rho_0 2 \frac{q\Phi}{k_B T}, \quad |r| > 0 \quad (3.31)$$

$$-\epsilon \nabla \cdot \nabla \Phi = \frac{\rho_0 q^2 \Phi}{kT} + Q \delta(\mathbf{r}) \quad (3.32)$$

$$\int dV \nabla \cdot (\nabla \phi) = \int (\nabla \phi \cdot \hat{\mathbf{n}}) dA \quad \text{Divergence Theorem} \quad (3.33)$$

$$4\pi h^2 \nabla \Phi = Q \quad (3.34)$$

$$\Phi = c \frac{e^{-ar}}{r} \quad (3.35)$$

$c$  is determined by  $Q$  and  $a$  is determined by the length scale from the first term on the right hand side of Poisson's equation.

### 3.3.2 Question

Figure 3.2: DNA,  $\Phi \propto \lambda \propto \frac{e}{b}$ 

$$-\epsilon \nabla \cdot \nabla \Phi = \rho_a \quad (3.36)$$

$$\epsilon \int (\nabla \cdot \Phi) \cdot \hat{\mathbf{n}} = Q = \lambda l \quad (3.37)$$

$$\epsilon 2\pi r l \partial_r \Phi = \lambda l \quad (3.38)$$

$$\Phi = \Phi_0 \ln r \Rightarrow \frac{\lambda}{2\pi\epsilon} \ln r \quad (3.39)$$

$$\epsilon 2\pi \Phi_0 = \lambda \quad (3.40)$$

$$\mathcal{E} = q\Phi = q\Phi_0 \ln r \Rightarrow \frac{q\lambda}{2\pi\epsilon} \ln r, \quad \text{Energy} \quad (3.41)$$

$$\mathcal{E} = \frac{q^2}{2\pi\epsilon} \frac{1}{b} \ln \frac{r}{a} \quad (3.42)$$

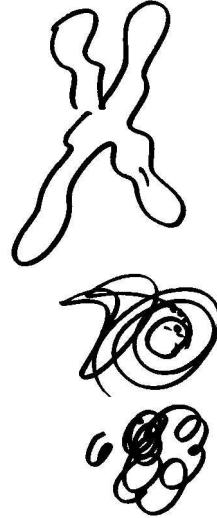


Figure 3.3: DNA coiling, proteins wrap up DNA

### 3.3.3 Question, condensation

$$Q = Q_0 e^{-xr} \quad (3.43)$$

$$\Phi = Q_0 \frac{e^{-xr}}{r} = \frac{Q(r)}{r} \quad (3.44)$$

### 3.3.4 What class should be able to do

Write equation for diffusion with drift in conservation law form. What enters into flux? Assume system is overdamped. Get equation with only  $b$  and  $U$ .

$$\dot{\rho} = -\partial_x T \quad (3.45)$$

$$J = \rho v - D\rho' \quad (3.46)$$

$$v = -\frac{1}{b} \frac{dU}{dx} \quad (3.47)$$



Figure 3.4: Charged object, salts

In closed system  $x \in [a, b]$ , say  $J_a = J_b$ . Prove total probability is conserved.

In addition, say  $J = 0$ . Derive  $\rho = \rho_0 e^{-\frac{U}{Db}} = \rho_0 e^{-\frac{U}{T}}$ .

Given the three equations on domain  $x \in [x_-, x_+]$  and an  $x_0$  such that  $\rho(x_0) = 0$  (absorbing boundary), derive:

$$-\frac{D}{J} = \int_{x_-}^{x_+} dx e^{-\beta U(x)} \int_{x_0}^x dx' e^{\beta U(x')} \quad (3.48)$$

Where do we use this model? Whenever we have system not in closed equilibria. Recall,

- “Closed equilibria”  $\Leftarrow \dot{\rho} = 0$  and  $J = 0$
- “Open equilibria”  $\Leftarrow \dot{\rho} = 0$  and  $J \neq 0$

Derivation:

$$J = \rho v - D\rho' \quad (3.49)$$

$$= -\frac{1}{b} \frac{dU}{dx} \rho - Dp' \quad (3.50)$$

$$= -D \left( \rho e^{\frac{U}{Db}} \right)' e^{-\frac{U}{Db}} \quad (3.51)$$

$$\frac{J}{D} e^{\frac{U}{Db}} = -\partial_x \rho e^{\frac{U}{Db}} \quad (3.52)$$

$$\beta = \frac{1}{k_B T} = \frac{1}{Db} \quad (3.53)$$

$$\frac{J}{D} \int_{x_0}^x dx' e^{\beta U(x')} = -\rho e^{\beta U(x)}, \quad \text{using } \rho(x_0) = 0 \quad (3.54)$$

$$-\frac{J}{D} e^{\beta U} \int_{x_0}^x dx' e^{\beta U(x')} = \rho \quad (3.55)$$

$$-\frac{J}{D} \int_{x_-}^+ dx e^{\beta U} \int_{x_0}^x dx' e^{\beta U(x')} = 1 \quad (3.56)$$

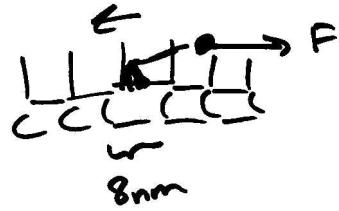


Figure 3.5: motor protein, 8 nm lattice sites

### 3.3.5 Thermal ratchet model

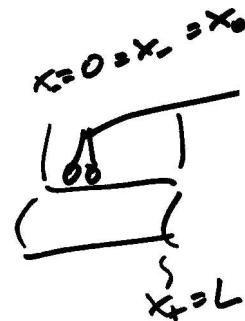
A reasonable model for a motor protein not going backward:

$$-\frac{D}{J} = \int_{x_-}^{x_+} dx e^{-\beta U(x)} \int_{x_0}^x dx' e^{\beta U(x')} \quad (3.57)$$

$$J_{\text{drift}} = \rho v \quad (3.58)$$

$$\tau = -\frac{1}{J} = \frac{1}{D} \int_{x_-}^{x_+} dx e^{\beta U} \quad (3.59)$$

$$U = -Fx \quad (3.60)$$

Figure 3.6:  $x = 0 = x_- = x_0$  to  $x_+ = L$ 

Introduce dummy variables

$$v = -\beta U = \beta F x \quad (3.61)$$

$$dv = (\beta F) dx \quad (3.62)$$

$$\alpha = \beta F L \quad (3.63)$$

Then we integrate:

$$\tau = \frac{1}{D} \int_0^L dx e^{-\beta Fx} \int_0^x dx' e^{-\beta Fx'} \quad (3.64)$$

$$= \frac{1}{D} \left( \frac{1}{\beta F} \right)^2 \int_0^\alpha dv e^{-v} \int_0^v dv' e^{-v'} \quad (3.65)$$

$$= \frac{L^2}{D} \frac{1}{\alpha^2} \left[ \int_0^\alpha dv e^v (1 - e^{-v}) \right] \quad (3.66)$$

$$= \frac{L^2}{D} \frac{1}{\alpha^2} \left[ \int_0^\alpha dv (e^v - 1) \right] \quad (3.67)$$

$$= \frac{L^2}{D} \frac{1}{\alpha^2} [e^\alpha - 1 - \alpha] \quad (3.68)$$

Time-scale here is set by diffusion

$$\tau_D = \frac{L^2}{D} \quad (3.69)$$

$$\tau = \tau_D \frac{1}{\alpha^2} \left( 1 + \alpha + \frac{1}{2} \alpha^2 - 1 - \alpha \right) \quad (3.70)$$

$$= \frac{1}{2} \tau_D \quad (3.71)$$

### 3.3.6 90's physics

Optical trapping of motor proteins

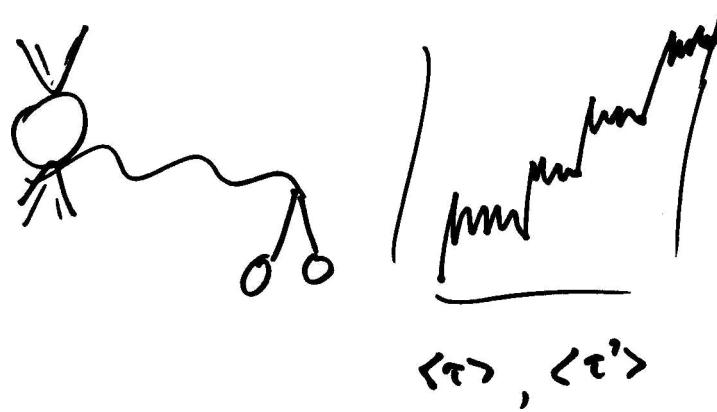


Figure 3.7: motor protein and optical trap, kinetics, probability

Kramer's escape over double potential well with local and global minimum.  
Consider this two open systems that interchange with each other

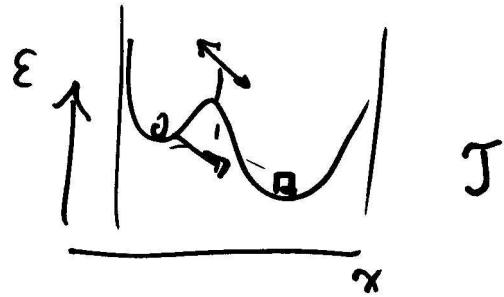


Figure 3.8: Energy versus chemical reaction

$$P(\text{doublewell}) = ? \quad (3.72)$$

$$P \propto \rho_0 e^{-\beta U} \quad (3.73)$$

so

$$P(\text{doublewell}) = P_l + P_r \quad (3.74)$$

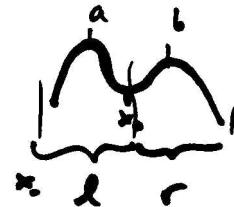


Figure 3.9: Doubly peaked probability

$$P_l = \int_{x_-}^{x_0} dx e^{-\beta U} \quad (3.75)$$

$$P_r = \int_{x_0}^{x_+} dx e^{-\beta U} \quad (3.76)$$

Integrands are close to Gaussians (sharply peaked). So if peak is at  $x = a$  we can estimate

$$\int_{x_-}^{x_0} dx e^{-g(x)} \approx e^{-g(a)} \sqrt{\frac{2\pi}{q''(x)}} \quad (3.77)$$

which leads to

$$P_l = p_0 e^{-\beta U_a} \sqrt{\frac{2\pi}{\beta U''_a}} \quad (3.78)$$

$$P_r = p_0 e^{-\beta U_b} \sqrt{\frac{2\pi}{\beta U''_b}} \quad (3.79)$$



Figure 3.10: Doubly peaked probability

$$\frac{P_l}{P_r} = e^{\beta(E_l - E_r)} \sqrt{\frac{U''_b}{U''_a}} \quad (3.80)$$

Defining

$$l_a = \sqrt{\frac{1}{\beta U''_a}} \quad (3.81)$$

$$l_b = \sqrt{\frac{1}{\beta U''_b}} \quad (3.82)$$

this becomes

$$\frac{P_l}{P_r} = \frac{l_a}{l_b} e^{\beta(U_a - U_b)} \quad (3.83)$$

Example of double well with two equal minima but with different widths.



Figure 3.11: Doubly peaked probability,  $U_a = U_b$

Now we will define rate of escaping from the left

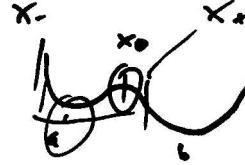


Figure 3.12: Doubly peaked probability

We assume

$$x_0 > x_* \quad (3.84)$$

$$x_1 < a \quad (3.85)$$

$$x_+ > x_* . \quad (3.86)$$

$$\frac{D}{J} = - \int_{x_-}^{x_+} dx e^{-\beta U(x)} \underbrace{\int_{x_0}^x dx' e^{\beta U(x')}}_{-e^{\beta U(x_*)} \sqrt{\frac{2\pi}{\beta U''(x_*)}} H(x_* - x)} \quad (3.87)$$

$$= \left[ \int_{x_-}^{x_+} dx e^{-\beta U(x)} \right] e^{\beta U_*} \sqrt{\frac{2\pi}{\beta U''_*}} \quad (3.88)$$

$$\approx e^{-\beta U_a} \sqrt{\frac{2\pi}{\beta U''_a}} e^{\beta U_*} \sqrt{\frac{2\pi}{\beta U''_*}} \quad (3.89)$$

$$\tau_+ = \frac{1}{J} = \frac{2\pi}{D} T \sqrt{\frac{1}{U''_a U''_b}} e^{\beta(U_* - U_a)} \quad (3.90)$$

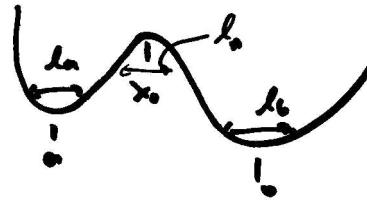


Figure 3.13: Doubly peaked probability with length scales

Length and time scales:

$$\begin{aligned} l_a &= \frac{1}{\sqrt{\beta U''_a}}, & l_b &= \frac{1}{\sqrt{\beta U''_b}}, & l_* &= \frac{1}{\sqrt{\beta U''_*}} \\ \tau_+ &= \frac{2\pi}{D} l_* l_a e^{-\beta(U_* U_a)} \end{aligned} \quad (3.91)$$

### 3.3.7 HW

1. Escaping to the right. Hint:

$$\tau_- = -\frac{1}{J_-}, \quad (3.92)$$

$$x_- < a, \quad (3.93)$$

$$x_+ > b, \quad (3.94)$$

$$x_0 \leq x_* \quad (3.95)$$



Figure 3.14: Doubly peaked probability with escape to left

2. Show steady state dynamics ratio is the same as the probability ratio from before using

$$\dot{P}_l = -\frac{1}{\tau_+} P_l + \frac{1}{\tau_-} P_r \quad (3.96)$$

$$\dot{P}_r = -\frac{1}{\tau_-} P_r + \frac{1}{\tau_+} P_l \quad (3.97)$$

### 3.3.8 Two state model

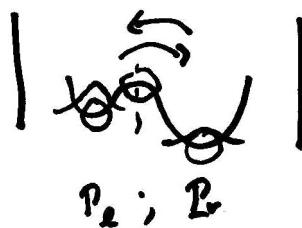


Figure 3.15:  $P_l, P_r$  fluxes

Flux dynamical system

$$\tau_+ = \frac{1}{J_+} \quad (3.98)$$

$$\tau_- = \frac{-1}{J_-} \quad (3.99)$$

$$\frac{d}{dt}P_l = -\frac{1}{\tau_+}P_l + \frac{1}{\tau_-}P_r = \dot{P}_l \quad (3.100)$$

$$\frac{d}{dt}P_r = -\frac{1}{\tau_-}P_r + \frac{1}{\tau_+}P_l = \dot{P}_r \quad (3.101)$$

$$\dot{P}_+ + \dot{P}_- = 0 \quad (3.102)$$

For some of you this may be your first example of a dynamical system. I can set the LHS to 0 to show how probabilities are related to each other and, reassuringly, the kinetic equilibrium of the dynamical system is the thermodynamic equilibrium (as we could have found by integrating the Boltzmann distribution). Note that this does not mean that the ratio of probabilities is set by the energies, since there may be many more ways of being on the left than being on the right – this is manifest in the “volume factor” which appears under Laplace approximation (*cf.* ??).

## 3.4 Lecture 9

### 3.4.1 Check-up review

- Problem 14

$$\epsilon \nabla^2 v = n_0 \left( q e^{\frac{qv}{k_B T}} - q e^{-\frac{qv}{k_B T}} \right) \quad (3.103)$$

If  $qv \ll k_B T$  we can use the Taylor series for the exponential  $e^x = 1 + x$ , yielding

$$\epsilon \nabla^2 v = n_0 \left[ q \left( 1 + \frac{qv}{k_B T} \right) - q \left( 1 - \frac{qv}{k_B T} \right) \right] \quad (3.104)$$

$$= \frac{2n_0 k^2 v}{k_B T} \quad (3.105)$$

So we have

$$\nabla^2 v = \left( \frac{2n_0 k^2}{\epsilon k_B T} \right) v. \quad (3.106)$$

Now we consider the dimensions of the last equation. We see that the left hand side has dimensions of  $v/l^2$  so the fraction in parentheses must have dimensions of  $1/l^2$ . This length is called the Debye length

$$\lambda_D = \sqrt{\frac{\epsilon k_B T}{2n_0 q^2}}. \quad (3.107)$$

The harder problem was show this is a solution to the above equation

$$v = \frac{ce^{-\kappa r}}{r} \quad (3.108)$$

where  $\kappa = \frac{1}{\lambda_D}$

- Problem 15

Find the other length scale using  $k_B T, q, \epsilon$ . Do this using Coulomb's law

$$F = \frac{q_1 q_2}{\epsilon r^2} \quad (3.109)$$

which implies the dimensions

$$\left[ \frac{q^2}{\epsilon} \right] = \frac{E}{l} l^2 = E \cdot l \quad (3.110)$$

and gives the length

$$\left[ \frac{q^2}{\epsilon} \frac{1}{k_B T} \right] = \frac{E \cdot l}{E} = \lambda_B. \quad (3.111)$$

- Problem 16

Show

$$\lambda_D = \sqrt{\frac{1}{n_0 \lambda_B}}. \quad (3.112)$$

- Problem 17

Motor protein tries to walk to the left with opposing force to the right.

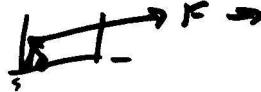


Figure 3.16: motor protein

$$\tau = -\frac{1}{J} = \frac{1}{D} \int_{x_-}^{x_+} dx e^{-\beta U} \int_{x_0}^x dx e^{\beta U} \quad (3.113)$$

$$J = \rho v - D \rho' = 1 \frac{1}{b} \rho U' - D \rho \quad (3.114)$$

$$= -\frac{1}{b} (\rho U' + k_B T \rho') \quad (3.115)$$

- Problem ?

$$\tau = \frac{1}{D} \int_{-L/2}^{L/2} dx e^{-\beta U(x)} \int_{-L/2}^{L/2} dx' e^{\beta U(x')} \quad (3.116)$$

Define  $\alpha = e^{\beta \epsilon}$ .

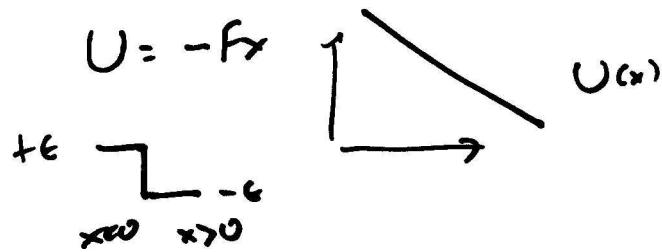
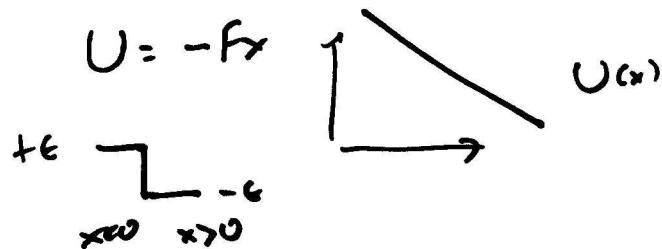
Figure 3.17:  $U(x)$ 

Figure 3.18:

$$e^{-\beta U} = e^{-\beta \epsilon} = \alpha^{-1}, \quad x < 0 \quad (3.117)$$

$$\Rightarrow e^{\beta U} = \alpha, \quad x < 0 \quad (3.118)$$

$$e^{-\beta U} = e^{\beta \epsilon} = \alpha, \quad x > 0 \quad (3.119)$$

$$\Rightarrow e^{\beta U} = \alpha^{-1}, \quad x > 0. \quad (3.120)$$

$$\begin{aligned} \tau &= \frac{1}{D} \left[ \int_{-L/2}^0 dx \alpha^{-1} \int_{-L/2}^x dx' \alpha \right. \\ &\quad \left. + \int_0^{L/2} dx \alpha \left( \int_{-L/2}^0 dx \alpha + \int_0^x dx' \alpha_1 \right) \right] \end{aligned} \quad (3.121)$$

$$\begin{aligned} &= \frac{1}{D} \left[ \int_{-L/2}^0 dx \int_{-L/2}^x dx' \right. \\ &\quad \left. + \alpha^2 \int_0^{L/2} dx \int_{-L/2}^0 dx + \int_0^{L/2} dx \int_0^x dx' \right] \end{aligned} \quad (3.122)$$

$$= \frac{1}{D} \left[ \int_{-L/2}^0 dx \left( x + \frac{L}{2} \right) + \alpha^2 \int_0^{L/2} dx \frac{L}{2} + \int_0^{L/2} dx x \right] \quad (3.123)$$

$$= \frac{1}{D} \left[ \frac{1}{2} x^2 \Big|_{-L/2}^0 + \left( \frac{L}{2} \right)^2 + \alpha^2 \left( \frac{L}{2} \right)^2 + \frac{1}{2} x^2 \Big|_0^{L/2} \right] \quad (3.124)$$

$$= \frac{1}{D} \left( \frac{L}{2} \right)^2 \left[ -\frac{1}{2} + 1 + \alpha^2 + \frac{1}{2} \right] \quad (3.125)$$

(3.126)

Recall the result for the thermal ratchet

$$\tau = \frac{1}{D} \frac{L^2}{2} \quad (3.127)$$

The full expression was (to get the one half)

$$\frac{1}{\alpha^2} (e^\alpha - 1 - \alpha) \approx \frac{1}{2} \frac{\alpha^2}{\alpha^2} \quad (3.128)$$

So the result for this problem is

$$\tau = \frac{1}{D} \left( \frac{L^2}{4} \right) (1 + \alpha^2) \quad (3.129)$$

(3.130)

and if  $\alpha = 1$  we recover the thermal ratchet result

$$\tau = \frac{1}{D} \frac{L^2}{2}. \quad (3.131)$$

- Problem 1

Form a dimensionless number and plug in actual numbers

$$\frac{\mu}{\rho v L} \sim 10^4 \quad (3.132)$$

- Problem 6

Estimate dimensionally  $\langle x^2 \rangle$  knowing

$$U = qx^6 \quad (3.133)$$

$$[q] = \frac{E}{l^2} \quad (3.134)$$

$$\lambda^{-6} = \frac{q}{k_B T} \quad (3.135)$$

$$\lambda = \left( \frac{k_B T}{q} \right)^{1/6}. \quad (3.136)$$

70 CHAPTER 3. NON-EQUILIBRIA AND TRANSPORT:  $J = \text{CONSTANT}$

This is similar to the homework problem where

$$U = \frac{1}{2} kx^2 \quad (3.137)$$

$$\lambda = \sqrt{\frac{k_B T}{k}} \quad (3.138)$$

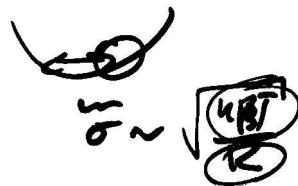


Figure 3.19: Spring trap

So “dimensionally estimate” means use the dimensions in the equation to estimate the requested quantity. For example, the homework problem asking how high you jumped knowing the time in the air. Then use  $t, g$  to get  $l$ . You get

$$[g] \sim \frac{l}{t^2} \quad (3.139)$$

## **Part II**

# **Systems Biology**



# Chapter 4

# Regulatory Dynamics

## 4.1 Kinetic models: 1-d dynamical systems

We finished off the last chapter talking about escape, and from that, two-state systems. In biology, frequently the transition from one state to another is facilitated by an enzyme, which wordnet defines as

any of several complex proteins that are produced by cells and act as catalysts in specific biochemical reactions.

For our purposes an enzyme is a thing that changes the energy barriers, but is not affected by the reaction.

Michaelis-Menten kinetics (mmk) for enzyme-triggered reactions  
 $\text{enzyme} + \text{substrate} \rightleftharpoons \text{complex} \rightarrow \text{enzyme} + \text{phosphate}$

### 4.1.1 Chemical kinetics

‘Chemical kinetics’ refers to the mapping between chemical reactions, as represented by chemical equations like  $A \leftrightarrow B \rightarrow C$ , and ordinary differential equations. Every arrow refers to the creation or destruction of some chemical species, each of which has a concentration which we denote in lowercase letters ( $\{a, b, c\}$ ).

Consider then  $dA/dt$ : the rate at which the concentration of  $A$  changes. There should be a negative contribution (meaning concentration is decreasing) at a rate proportional to the amount of  $A$  already present – the more copies of  $a$ , the higher the rate at which they are converted to  $b$ . Let’s call the constant of proportionality  $k_1$ :

$$\text{partial model : } \frac{dA}{dt} = -k_1 A + \dots \quad (4.1)$$

Clearly there should be a positive contribution, proportional to the amount of  $b$  in the system, or equivalently proportional to the concentration  $B$ . Since this reaction has the opposite effect, let’s call the constant of proportionality  $k_{-1}$ . These are all of the reactions

involving  $a$  (clearly, since there are no other lines pointing towards or away from  $a$  in the chemical reaction equation, so we know the whole dynamic for  $A$ :

$$\text{complete model : } \frac{dA}{dt} = -k_1 A + k_{-1} B \quad (4.2)$$

A complication occurs if two different species react, e.g.,  $A + B \rightarrow C$ . In this case, for example, the rate  $dc/dt$  at which  $C$  is created should be proportional not only to the amount of  $a$  present but also to the amount of  $B$  present. Defining a new constant of proportionality  $k_2$ , we have

$$\frac{dc}{dt} = k_2 ab. \quad (4.3)$$

The qualitative difference between these two types of kinetics is sometimes referred to as ‘first-order’ vs ‘second-order’ kinetics.

#### 4.1.2 Michaelis-Menton kinetics

[5],p433

Let’s think, then, about how we might represent the action of an enzyme in terms of a chemical equation, and then in terms of chemical kinetics. We imagine that the enzyme  $E$  and substrate  $S$  form a complex  $C$ , and that this reaction is reversible. Then, the enzyme chews up an ATP or some other energy source to facilitate a transition of the complex to some new product  $P$  while the enzyme itself remains intact, and able to perform subsequent reactions.

The reversibility of  $E + S \rightleftharpoons C$  will be parameterized by two reaction rates  $k_{\pm 1}$ , and the irreversible reaction  $C \rightarrow P + E$  will have a rate constant  $k_2$ . The enzyme uses the energy source to drive  $k_{-2} \rightarrow 0$ .

As a chemical equation we then have:



Let’s assume  $S$  is so huge that it is a constant in time and has no dynamic. Then  $\dot{c} = k_1 es - k_2 c - k_{-1} c$

1. Note that the total number of enzymes is a constant:  $e_t = e + c$
2. Note that the quantity of interest is the rate of production of product  $\dot{p}$  normalized by the total enzyme concentration  $e_t$
3. Note that  $k_1$  and  $k_2$  do not have the same units! We can form a quantity with units of concentration:  $K_m = (k_1 + k_2)/k_{-1}$

The steady state obeys

$$k_{-1} e_* s = (k_2 + k_1) c_* \quad (4.5)$$

$$= k_{-1} K_m c_* \quad (4.6)$$

$$e_* = \frac{K_m c_*}{s} \quad (4.7)$$

So the steady-state, normalized rate of production of product is

$$r \equiv \frac{\dot{p}}{e_t} = \frac{k_2 c_*}{e_* + c_*} \quad (4.8)$$

$$= \frac{k_2}{K_m c_* / s + c_*} \quad (4.9)$$

$$= \frac{k_2 s}{K_m + s} \quad (4.10)$$

which is the MMK law. However, there's also a dynamical systems way of looking at this.

$$\dot{e} = -k_{-1} e s + k_2 c + k_1 c \quad (4.11)$$

$$\dot{c} = k_{-1} e s - k_2 c - k_1 c \quad (4.12)$$

$$\dot{p} = k_2 c \quad (4.13)$$

How do I know the steady state is stable?

### 4.1.3 Hill functions and regulation models

hill kinetics model situations where a substrate (some multi-atomic “thing”) requires co-operative binding

mmk solutions have the form

$$\frac{\dot{p}}{e} \sim \frac{s}{s + k} \quad (4.14)$$

hill solutions have the form

$$\frac{\dot{p}}{e} \sim \frac{o^n}{o^n + k^n} \quad (4.15)$$

(note that both are quotients of low-order polynomials.)

## 4.2 Dynamical systems models for regulation

This sounds like chemistry. I thought this was a biology class!

Enzymatic models are often used to describe transcriptional regulation

### 4.2.1 Addendum to the parts list: informatic parts

Figure 4.1: a gene, showing regulatory region with “motifs” (really binding sites)

Figure 4.2: a transcription factor, in action (here, repressing), in the major groove

### 4.2.2 Mathematical models of transcriptional regulation

(e.g., the Griffiths model) (stability of fixed points in 1D and 2D) we can model the transcription/translation process as follows. say the rate of protein generation is given by

$$\dot{p} = km - Rp \quad (4.16)$$

where  $m$  is mRNA concentration and  $k$  and  $R$  are rate constants. in the absence of mRNA, proteins decay exponentially as

$$p(t) = p_0 e^{-Rt}. \quad (4.17)$$

if mRNA generation is modeled by hill kinetics, we have

$$\dot{m} = \alpha \frac{p^n}{p^n + k^n} - rm \quad (4.18)$$

### The algorithm of all dynamical systems

let's study this as a dynamical system in more depth via the following procedure:

1. identify degrees of freedom, coordinates, and parameters of the model
2. solve for fixed points
3. identify bifurcation points
4. characterize stability of fixed points

### Nondimensionalization

Really there's one proto-step before this algorithm: eliminating as many parameters as possible.

$$t = \tilde{t} \quad (4.19)$$

$$m = \tilde{m} \tilde{m}(\tilde{t}) \quad (4.20)$$

taking  $n = 2$  for simplicity, we can make this model dimensionless (do for hw) by re-scaling variables to obtain

$$\dot{p} = f(p(t)) \equiv \frac{p^2}{p^2 + 1} - \gamma p, \quad (4.21)$$

where our coordinate is  $t$ , degree of freedom is  $p$ , and parameter is  $\gamma$ .

Figure 4.3: A plot of  $\gamma p$  and  $\alpha(p)$ : decay and transcription/translation, indicating the existence of a “fixed point.”

now let's look for fixed points, those  $p_*$  for which  $f(p_*) = 0$ . this gives the condition:

$$\gamma p_* = \frac{p_*^2}{p_*^2 + 1} \quad (4.22)$$

which admits solutions

$$p_* = \frac{1 \pm \sqrt{1 - 4\gamma^2}}{2\gamma}. \quad (4.23)$$

looking at the argument of the square root, we see that the number of solutions for  $p_*$  change with the value of  $\gamma$ ; for  $0 < \gamma < \frac{1}{2}$ , we have 3 fixed points (call them  $p_0, p_{\pm}$ ), while for  $\gamma > \frac{1}{2}$  we have only 1 fixed point. the value of  $\gamma = \frac{1}{2}$  is known as a bifurcation point, a critical parameter value that effects the existence or stability of fixed points.

we can determine the stability of these fixed points by examining the system's behavior at points close to  $p_*$ , i.e.  $p_* + \eta$  for  $|\eta| \ll |p_*|$ . using a Taylor expansion in our equation of motion,

$$\frac{d(p_* + \eta)}{dt} = f(p_* + \eta) \approx f(p_*) + \eta f'(p_*). \quad (4.24)$$

by definition,  $\dot{p}_* = f(p_*) = 0$ , which gives

$$\dot{\eta} = \eta f'(p_*). \quad (4.25)$$

since  $f'(p_*)$  is just a number (independent of time), we can easily solve this differential equation to arrive at

$$\eta(t) = \eta_0 e^{f'(p_*)t}. \quad (4.26)$$

for  $f'(p_*) < 0$ ,  $\eta(t)$  exponentially tends towards zero, indicating that  $p_*$  is a stable fixed point, while for  $f'(p_*) > 0$ ,  $\eta(t)$  exponentially tends towards infinity, indicating that  $p_*$  is unstable.

### Homework:

: show that  $p_-$  is unstable,  $p_+$  stable.

$$m_t = \frac{p^2}{1 + p^2} - rm \quad (4.27)$$

$$p_t = m - Rp \quad (4.28)$$

## 4.3 Experiment: the toggle switch

### 4.3.1 Lambda phage

### 4.3.2 GFP

### 4.3.3 SGNs

## 4.4 Transduction

### 4.4.1 Lambda phage

### 4.4.2 Chemotaxis

### 4.4.3 MapK

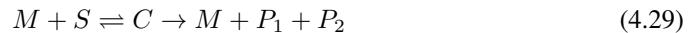
Chemical reaction



Figure 4.4: reaction diagram

How would one “break” detailed balance? One way is to use energy to bias the reaction rates. An enzyme is a molecule which facilitates a chemical reaction. We’re going to study now a mathematical model of one particular example, the motor protein kinesin. Kinesin cleaves ATP into two parts, ADP and the free phosphate group.

First let’s consider the energetic diagram we could write where the y-axis represents energy and the x-axis is a “chemical coordinate”.



Simple dynamical system example:

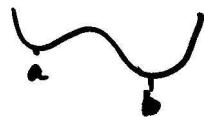
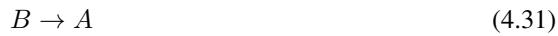


Figure 4.5: plot



In shorthand notation

$$A \rightleftharpoons B \quad (4.32)$$

is equivalent to the dynamical system

$$\dot{a} = -k_1 a + k_{-1} b \quad (4.33)$$

$$\dot{b} = -k_{-1} b + k_1 a \quad (4.34)$$

Real protein dynamics

$$\dot{m} = \dots \quad (4.35)$$

$$\dot{c} = \dots \quad (4.36)$$

$$\dot{s} = \dots \quad (4.37)$$

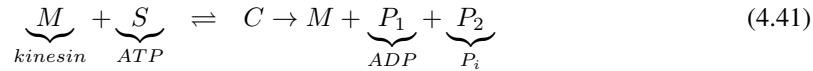
$$\dot{p}_1 = \dots \quad (4.38)$$

$$\dot{p}_2 = \dots \quad (4.39)$$

$$\dot{m} = \dots \quad (4.40)$$



Figure 4.6: ATP docks with protein binding site



$$\dot{c} = -k_2 c + k_1 s m - k_{-1} c \quad \text{concentration} \quad (4.42)$$

$$\dot{s} = -k_1 s m + k_{-1} c \quad \text{substrate} \quad (4.43)$$

$$\dot{m} = -k_1 s m + k_2 c + k_{-1} \quad \text{motor protein} \quad (4.44)$$

$$\dot{p} = k_2 c \quad (4.45)$$

$$\Rightarrow \dot{m} + \dot{c} = 0 \quad (4.46)$$

$$m_T = m + c \quad (4.47)$$

$$r = \frac{\dot{p}_i}{m_T} \quad (4.48)$$

Characteristic timescales

$$\begin{aligned} \dot{x} &= -\alpha x \\ \tau_m &= \frac{1}{k_1 s} & \tau_s &= \frac{1}{k_1 m} & \tau_c &= \frac{1}{k_{-1} + k_2} \\ k_1 \tau_m &= \frac{1}{s} & k_1 \tau_s &= \frac{1}{m} & k_1 \tau_c &= \frac{1}{k_{-1} + k_2} = \frac{1}{K} \end{aligned} \quad (4.49)$$

CHECK THIS (somewhere an “n” is missing)

$$S \rightleftharpoons F + F + F + F \quad (4.50)$$

$$\dot{s} = -k_3 s + k_3 f^n \quad (4.51)$$

$$\dot{f} = n k_3 s - n k_3 f^n \quad (4.52)$$

$$\dot{c} = -k_2 c + k_1 s m - k_1 c = 0 \quad (4.53)$$

$$\frac{m}{c} = \frac{k_{-1} + k_2}{k_1 s} = \frac{K}{s} \quad (4.54)$$

$$r = \frac{\dot{p}}{m_T} = \frac{k_2 c}{m + c} = \frac{k_2}{m/c + 1} = \frac{k_2}{K/s + 1} = \frac{k_2 s}{K + s} \quad (4.55)$$

$$(4.56)$$

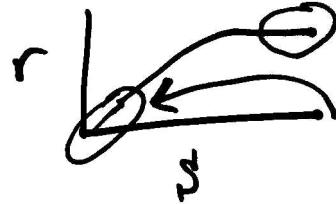


Figure 4.7:  $r(s)$

$$k_2 = \lim_{s \rightarrow \infty} r(s) \quad (4.57)$$

$$\frac{k_2}{K} \approx \left. \frac{dr}{ds} \right|_{s=0} \quad (4.58)$$

Opens the book “A Genetic Switch.”



Figure 4.8: Cells, lipod bilayers, proteins, ATP, free phosphate, nucleic acid

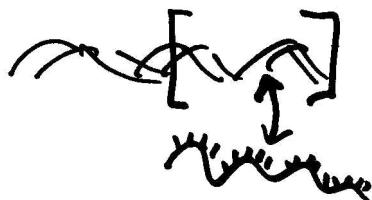


Figure 4.9: Transcription: DNA copied to nucleic acid

$$D \rightarrow R \rightarrow P \quad (4.59)$$

$$S \rightleftharpoons F + F + F + F \quad (4.60)$$

$$f_T = nS + f \quad (4.61)$$

$$\dot{s} = -k_3 S + k_{-3} f^n \quad (4.62)$$

$$\dot{f} = n k_3 S - n k_{-3} f^n \quad (4.63)$$

$$\text{like before except } S \rightarrow \left( \frac{k_{-3}}{k_3} \right) f^n \quad (4.64)$$

$$\frac{\dot{r}}{m_T} = \frac{k_2 \left( \frac{k_{-3}}{k_3} \right) f^n}{K + \left( \frac{k_{-3}}{k_3} \right) f^n} \quad (4.65)$$

(4.66)

Check above formula. Maybe no “dot” on r.

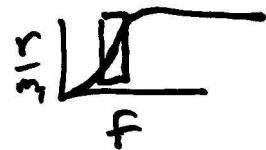


Figure 4.10:  $r/m_T$  versus  $f$

$$\frac{r}{m_T} = \frac{k_2 \left( \frac{k_{-3}}{k_3} \right) f^n}{K + \left( \frac{k_{-3}}{k_3} \right) f^n} \quad (4.67)$$

$$= \frac{k_2 f^n}{\tilde{K}^n + f^n} \quad (4.68)$$

#### 4.4.4 HW

1. Find  $\tilde{K}$  as a function of  $K, k_{-3}, k_3$ .
2.  $D + F \rightleftharpoons C \rightarrow D + F + M \dot{m}(D_T, F)$

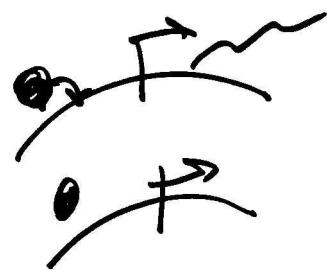


Figure 4.11: DNA

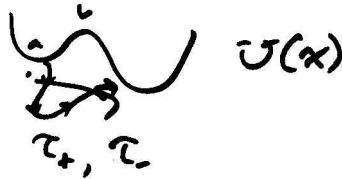


Figure 4.12: Double well

#### 4.4.5 Systems Biology

Recall Kromer's escape

Construct a model for rate of movement between states,  $A \rightleftharpoons B$ . Can define rate to left  $\tau_-$  and rate to the right  $\tau_+$ .

$$\dot{a} = -\tau_+ a + \tau_- b \quad (4.69)$$

$$\dot{b} = -\tau_- a + \tau_+ a \quad (4.70)$$

$$(4.71)$$

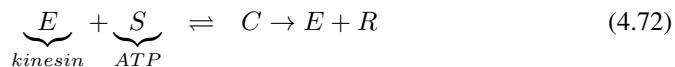
This corresponds to the chemical dynamics  $A \rightleftharpoons B$ .

But motor proteins don't just move back and forth. They use energy. So there is an enzyme which also has a role.



Figure 4.13: enzyme

The chemical equations of enzyme kinetics are of the form



which is written as the dynamical system for the concentrations in time

$$\dot{e} = -k_1 es + k_1 c + k_2 c \quad \text{enzyme} \quad (4.73)$$

$$\dot{c} = k_1 es - k_{-1} c - k_2 c \quad \text{complex} \quad (4.74)$$

$$\dot{p} = k_2 c \quad (4.75)$$

$$\dots \quad (4.76)$$

$$\dot{s} = -k_1 es + k_{-1} c \quad (4.77)$$

Characteristic timescales

$$k_1\tau_e = \frac{1}{s} \quad (4.78)$$

$$k_1\tau_c = \frac{k_1}{k_2 + k_{-1}} = \frac{1}{K} \quad (4.79)$$

$$k_1\tau_s = \frac{1}{e} \quad (4.80)$$

$$e + c = e_T \quad (4.81)$$

$$\dot{e}_T = \dot{e} + \dot{c} = 0 \quad (4.82)$$

$$r = \frac{\dot{p}}{e_T} = \frac{k_2 c}{e + c} \quad (4.83)$$

$$\dot{e} = 0 \Rightarrow e = \frac{(k_{-1} + k_2)c}{k_1 s} = \frac{cK}{s} \Rightarrow \frac{c}{e} = \frac{s}{K} \quad (4.84)$$

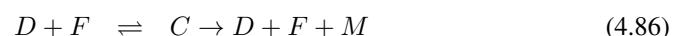
$$r = \frac{k_2 c/e}{1 + c/e} = k_2 \frac{s/K}{1 + s/K} = k_2 \frac{s}{K + s} \quad (4.85)$$

DNA has a major and minor groove.



Figure 4.14: DNA

DNA and transcription factor forms a complex yielding DNA, the transcription and mRNA



This can be written as a set of differential equations

$$\dot{d} = -k_1 d f + k_{-1} c + k_2 c \quad (4.87)$$

$$\dot{f} = -k_1 df + k_{-1} c + k_2 c \quad (4.88)$$

$$\dot{c} = k_1 df - k_{-1} c - k_2 c \quad (4.89)$$

$$\dot{m} = k_2 c \quad (4.90)$$

$$r = \frac{\dot{m}}{d_T} \quad (4.91)$$

$$\dot{c} = \dot{f} = \dot{d} = 0 \quad (4.92)$$

$$d = \frac{c(k_{-1} + k_2)}{k_1 f} = \frac{c}{K^{-1} f} \quad (4.93)$$

$$r = \frac{\dot{m}}{d_T} = \frac{\dot{m}}{d + c} = \frac{k_2 c}{d + c} = \frac{k_2 c / d}{1 + c / d} \quad (4.94)$$

$$\frac{c}{d} = K^{-1} f = \frac{f}{K} \quad (4.95)$$

$$r = \frac{k_2 f / K}{a + f / K} = k_2 \frac{f}{K + f} \quad (4.96)$$

What does plot of  $r$  look like?

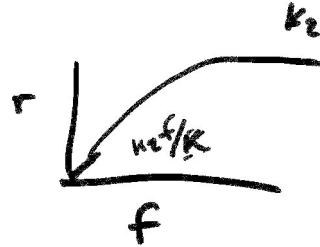


Figure 4.15:  $r(f)$

#### 4.4.6 Homework

- Recall the book “The Genetic Switch”. The transcription factors forms dimers and tetramers.

$$r = k_2 \frac{f}{K + f} \quad (4.97)$$

$$F \rightleftharpoons x + x + x + x \quad (4.98)$$

$$\dot{f} = -k_3 f + k_{-3} x^n \quad (4.99)$$

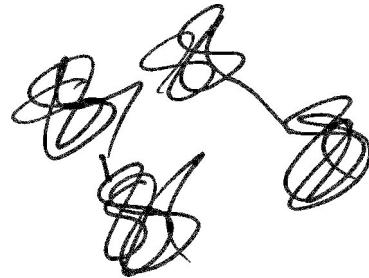


Figure 4.16: Sketch tetramer

$$\dot{x} = -nx^{nk-3} + nfk_3 \quad (4.100)$$

$$x + nf = x_T \quad (4.101)$$

$$\dot{x}_T = 0 \quad (4.102)$$

HW 1 Show  $r$  is in this form

$$\frac{x^n}{? + x^n} \quad (4.103)$$

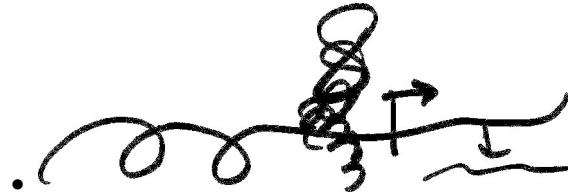


Figure 4.17: Transcription binding site



In equations

$$\dot{m} = k_2 d \quad (4.106)$$

$$\dot{d} = -k_1 df + k_{-1} c \quad (4.107)$$

$$d_T = d + c \quad (4.108)$$

$$\dot{d}_T = 0 \quad (4.109)$$

$$r = \frac{\dot{m}}{d + c} \quad (4.110)$$

$$\frac{c}{d} = \frac{k_1 f}{k_{-1}} = \frac{f}{\tilde{K}} \quad (4.111)$$

$$\tilde{K} = \frac{k_{-1}}{k_1} \quad (4.112)$$

$$r = \frac{\dot{m}}{d_T} = \frac{k_2 d}{d + c} = \frac{k_2}{1 + c/d} = \frac{k_2}{1 + f \tilde{K}} \quad (4.113)$$

Hw 2 and 3 are to show

$$r = k_2 \frac{\tilde{K}}{\tilde{K} + f} \quad (4.114)$$

What does plot of  $r$  look like?

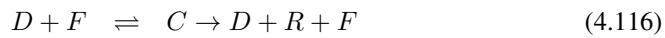


Figure 4.18:  $r(f)$

#### 4.4.7 HW: Transcription upregulation



Figure 4.19: transcription binding



In equations, at steady state

$$0 = \dot{d} = -k_1 d f + k_{-1} c + k_2 c \quad (4.117)$$

$$0 = \dot{c} = \dots \quad (4.118)$$

$$0 = \dot{f} = \dots \quad (4.119)$$

$$\frac{\dot{r}}{d_T} = \frac{\dot{r}}{d + c} = k_2 \frac{f}{K + f} \quad (4.120)$$

$$(4.121)$$

$$\dot{x} = 4k_{-3}f - 4k_3x^4 \quad (4.122)$$

The tetramer reaches equilibrium quickly, i.e.  $x$  reaches steady state quickly.

$$4k_{-3}f = 4k_3x^4 \quad (4.123)$$

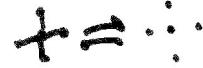


Figure 4.20: tetramer sketch

$$\frac{\dot{r}}{d_T} = \frac{k_2 \frac{k_3}{k_{-3}} x^4}{K + \frac{k_3}{k_{-3}} x^4} = k_2 \frac{x^4}{\tilde{K} + x^4} \quad (4.124)$$

where  $K \sim \frac{k_{-1}+k_2}{k_1}$ . So we have

$$\frac{\frac{k_3}{k_{-3}} x^4}{\frac{k_{-1}+k_2}{k_1} + \frac{k_3}{k_{-3}} x^4} = \frac{x^4}{\tilde{K}^4 + x^4} \quad (4.125)$$

where we define the new constant

$$\tilde{K}^4 \equiv \frac{k_{-1} + k_2}{k_1} \frac{k_{-3}}{k_3} \quad (4.126)$$

$$\dot{r} = -k_1 df + k_{-1} c k_2 c \quad (4.127)$$

$$(4.128)$$

Now we use the following dimensional analysis

$$[k_{-1}] \sim \frac{1}{t} = [k_2] \quad (4.129)$$

$$[k_1] \sim \frac{1}{c t} \quad (4.130)$$

$$[k_{-3}] \sim t^{-1} \quad (4.131)$$

$$[k_3] \sim \frac{1}{c^3} \frac{1}{t} \quad (4.132)$$

and find

$$[\tilde{K}^4] \sim \frac{t^{-1}}{\frac{1}{ct}} \frac{t^{-1}}{\frac{1}{c^3 t}} = c^4 \quad (4.133)$$

So we showed

$$\frac{\dot{r}}{d_T} = k_2 \frac{x^n}{\tilde{K}^n + x^n} \quad (4.134)$$

#### 4.4.8 HW given previously, inhibitor, down-regulation

By way of comparison, recall the process

$$D + F \rightleftharpoons C \quad (4.135)$$

$$D \rightarrow D + R \quad (4.136)$$

$$F \rightleftharpoons \underbrace{x + x + \cdots + x}_n \quad (4.137)$$

$$\dot{r}d_T = k_2 \frac{\tilde{K}^n}{\tilde{K}^n + x^n} \quad (4.138)$$

For simplicity, define

$$\alpha_n^\pm(y) = (1 + y^{\mp n})^{-1} \quad (4.139)$$

Then we can do two simultaneous problems and write the fractions on the right hand side of (4.134) and (4.138) as

$$\frac{1}{\left(\frac{\tilde{K}}{x}\right)^n + 1} = \left(1 + \left(\frac{x}{\tilde{K}}\right)^{-n}\right)^{-1} \quad (4.140)$$

$$\frac{1}{1 + \left(\frac{x}{\tilde{K}}\right)^n} = \left(1 + \left(\frac{x}{\tilde{K}}\right)^{+n}\right)^{-1} \quad (4.141)$$

We want to take the derivative of these terms. Using the shorthand

$$(\alpha_n^\pm)' = - (1 + y^\mp)^{-2} (\mp n y^{\mp n-1}) \quad (4.142)$$

$$= \pm \frac{y^{\mp n-1}}{(1 + y^\mp)^{-2}} \frac{(y^\pm)^2}{(y^\pm)^2} \quad (4.143)$$

$$= \pm \frac{y^{\pm n-1}}{(y^\pm + 1)^{-2}} \quad (4.144)$$

$$= -(\alpha_n^\mp)' \quad (4.145)$$

Let's look at an example with  $n = 1$

$$\alpha_1^+ = \frac{y}{1 + y} \quad (4.146)$$

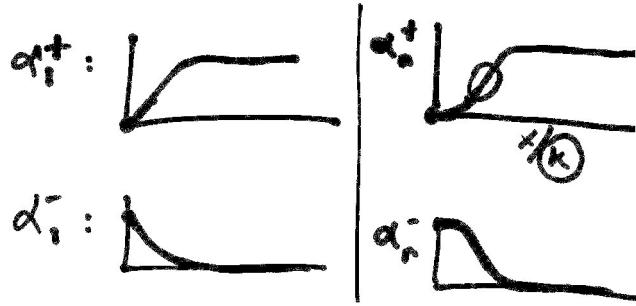
$$(\alpha_1^+)' = \frac{(1 + y) - y}{(1 + y)^2} = \frac{1}{y^2} \quad (4.147)$$

$$\alpha_1^- = \frac{1}{1 + y} \quad (4.148)$$

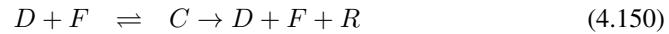
$$(\alpha_1^-)' = \frac{-1}{(1 + y)^2} = -\alpha_1^+ \quad (4.149)$$

Let's graph the  $\alpha$  curves

Notice the behavior for  $n = 1$  is very different from  $n \neq 1$ .

Figure 4.21: Graphs of  $\alpha_1^\pm$  and  $\alpha_n^\pm$ 

#### 4.4.9 Up-regulation



Or in equations

$$\dot{r} = d_T k_2 \alpha(x/\tilde{K}) - k_3 r \quad (4.155)$$

$$\dot{x} = -k_6 x + k_4 r \quad (4.156)$$

Say  $x$  rapidly establishes its own equilibrium. Then we have

$$x = \frac{k_4}{k_6} r \quad (4.157)$$

Then we have a one-dimensional dynamical system

$$\dot{r} = d_T k_2 \alpha\left(\underbrace{\frac{k_4}{k_6 \tilde{K}}}_\text{tilde{K}} r\right) - k_5 r \quad (4.158)$$

Define the rescaled dimensionless variables

$$y = \frac{r}{\tilde{K}} \quad (4.159)$$

$$\tau = \frac{t}{t_0} \quad (4.160)$$

$$(4.161)$$

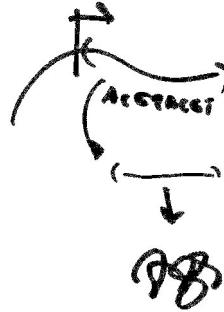


Figure 4.22: Transcription, translation sketch

Also, define derivatives

$$f' = \frac{d}{d\tau} f \quad (4.162)$$

$$\dot{f} = \frac{d}{dt} f \quad (4.163)$$

Then we can write the system as

$$\dot{r} = d_T k_2 \alpha \left( r / \tilde{K} \right) - k_5 r \quad (4.164)$$

$$y = r / \tilde{K} \quad (4.165)$$

$$t = \tau t_0 \quad (4.166)$$

- HW 1: Show

$$\frac{\tilde{K}}{t_0} y' = d_T k_2 \alpha(y) - k_5 y \tilde{K} \quad (4.167)$$

Letting

$$t_0 = k_5^{-1} \quad (4.168)$$

or

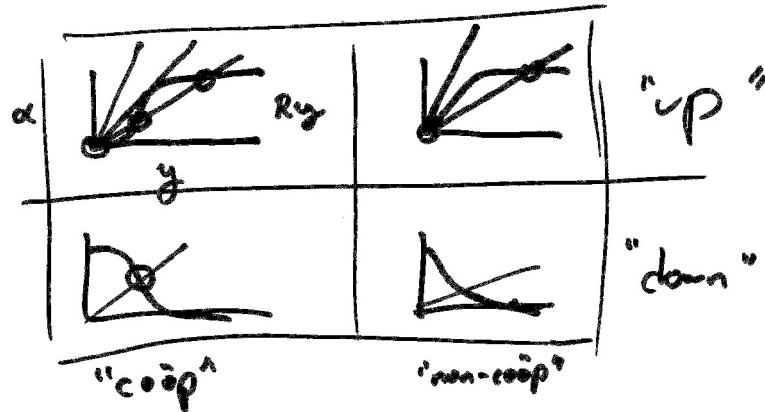
$$t_0 = k_2^{-1} \left( \frac{\tilde{K}}{d_T} \right) \quad (4.169)$$

- HW 2: Using the characteristic time scale (4.169), show

$$y' = \alpha(y) - Ry \quad (4.170)$$

where  $R$  is some characteristic decay rate.

Note the last equation is a first order ODE but highly nonlinear due to the form of  $\alpha$ . Thus, an analytic solution is not available. Instead, the equation is studied using dynamical systems techniques. We can also study an approximate long time solution where  $y' = 0$  corresponding to the intersection of  $\alpha(y)$  and  $Ry$ .

Figure 4.23:  $\alpha$ 

- HW 3:

$$\dot{x} = f(x) \quad (4.171)$$

$$f(x) = 0 \quad (4.172)$$

$$(4.173)$$

If  $x = x_0 + \eta$  where  $\eta$  is “small”, show

$$\dot{\eta} = \eta f'(x_0) \quad (4.174)$$

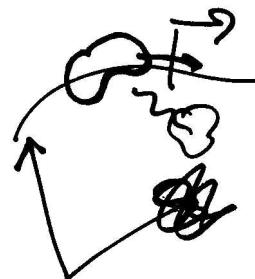


Figure 4.24: Upreregulating gene

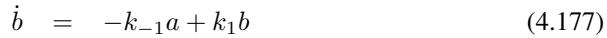
#### 4.4.10 Student question

Reviewed upregulation and especially which dynamical variables go to steady state.

### 4.5 Chemical kinetics review



Figure 4.25: double well



If we have parameters which rapidly reach equilibrium



If substrate is in equilibrium with the individual elements,  $S = nX$  we have



#### 4.5.1 Up- and down-regulation

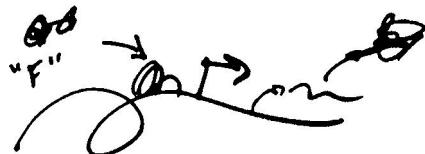


Figure 4.26: DNA transcription

$$D + F \rightleftharpoons C \rightarrow D + R + F \Rightarrow \frac{\dot{r}}{d_T} = k_2 \frac{F}{K + F} \quad (4.182)$$

$$D + F \rightleftharpoons C; D \rightarrow D + R \Rightarrow \frac{\dot{r}}{d_T} = k_2 \frac{K}{K + F} \quad (4.183)$$

Recall RNA decays

$$\dot{r} = k_2 d_T \frac{x^n}{K + x^n} - k_4 r \quad (4.184)$$

$$\dot{x} = k_5 r - k_6 x \quad (4.185)$$

where we have used  $K = x^n$ . If  $k_6$  dominates

$$\dot{r} = k_2 d_T \frac{r^n}{\tilde{K} + r^n} - k_4 r \quad (4.186)$$

$$\dot{y} = \alpha(y) - Ry \quad (4.187)$$

$$\alpha_n^\pm = \begin{cases} \frac{y^n}{1+y^n}, & + \\ \frac{1}{1+y^n}, & - \end{cases} \quad (4.188)$$

The steady state solutions can be enumerated graphically by finding the intersection of  $\alpha(y)$  and  $Ry$  in the different case of cooperative and non-cooperative up- and down-regulation.

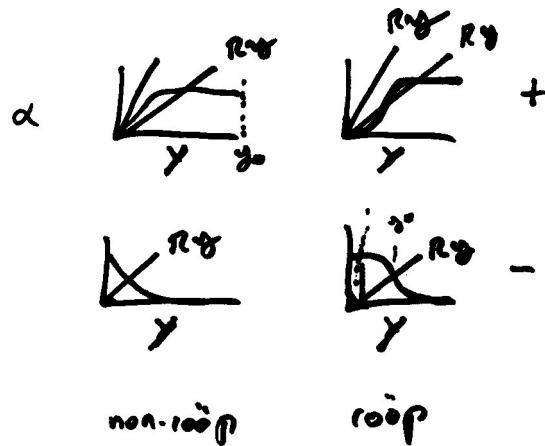
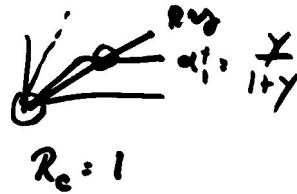


Figure 4.27:  $\alpha$

Consider the steady state of a protein which is auto-upregulating. The critical value of the parameter is  $R_c = 1$ .

Figure 4.28:  $\alpha$  for auto-upregulation

$$\dot{y} = \alpha(y) - Ry \quad (4.189)$$

We expand about one of the equilibrium solutions

$$\alpha(y_*) = Ry_* \quad (4.190)$$

$$y = y_* + \eta \quad (4.191)$$

$$y_* + \dot{\eta} = \alpha(y_* + \eta) - R(y_* + \eta) \quad (4.192)$$

Taylor expanding

$$\dot{\eta} \approx \alpha'(y_*)\eta + \alpha''(y_*)\eta^2 - Ry_* - R\eta \quad (4.193)$$

$$(4.194)$$

we get

$$\dot{\eta} = \eta(\alpha'(y_*) - R) \quad (4.195)$$

The solution to such an equation is of the form

$$y' = \alpha y \quad (4.196)$$

$$y = y_0 e^{\alpha t} \quad (4.197)$$

Or more generally,

$$\dot{x} = f(x) \quad (4.198)$$

The goal is to Find the fixed point satisfying

$$f(x_*) = 0 \quad (4.199)$$

and study

$$f'(x_*) \gtrless 0 \quad (4.200)$$

Negative implies linear stability and positive implies instability.

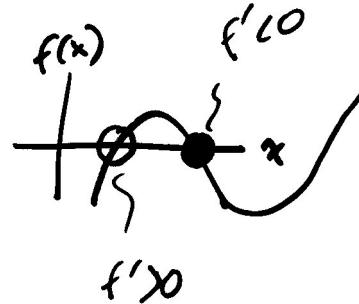


Figure 4.29: equilibrium points

So in the non-cooperative auto-upregulation example,  $R > 1$  implies stability.

$$\alpha = \frac{y}{1+y} \quad (4.201)$$

$$\alpha' = \frac{(1+y)-y}{(1+y)^2} = \frac{1}{(1+y)^2} \quad (4.202)$$

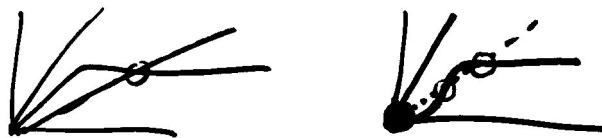
$$\alpha'(0) = 1 \quad (4.203)$$

$$\dot{\eta} = \eta(\alpha' - R) = \eta(\alpha'(0) - R) = \eta(1 - R) \quad (4.204)$$

- HW 1: Classify the stability for the nonzero fixed point where  $y_* \neq 0$

Figure 4.30:  $\alpha$  for auto-upregulation HW 1

Now consider the cooperative case. There are three possible fixed points.

Figure 4.31:  $\alpha$  for cooperative upregulation

Near a fixed point

$$\dot{\eta} = (\alpha' - R)\eta \quad (4.205)$$

$$\alpha = \frac{y^n}{1 + y^n} \quad (4.206)$$

$$(4.207)$$

At the origin,  $\alpha' = 0$  so

$$\dot{\eta} = -R\eta \quad (4.208)$$

- HW 2: Classify the stability for the nonzero fixed points where  $y_* \neq 0$

1. Solve for  $y_*$

2. Solve for  $\alpha'(y_*) - R$

Consider down-regulation. Now consider the cooperative case. There are three possible fixed points.



Figure 4.32:  $\alpha$  for down-regulation

$$\alpha' < 0 \quad (4.209)$$

$$\dot{y} = \alpha(y) - Ry \quad (4.210)$$

$$\partial_y \dot{y} = \alpha' - R \quad (4.211)$$

Thus,

$$y_* = \frac{\alpha(y_*)}{R} \quad (4.212)$$

So for a gene which is auto-repressive the one fixed point is stable. For the up-regulating gene there are three fixed points whose stabilities vary based on cooperation.

### 4.5.2 Regulation revisited

There were a number of simplifications used throughout the previous section. Here we elucidate those in reverse order.

First, we assumed that the dynamics of transcription and translation were separate and then combined them. The dynamics of transcription are

$$\dot{r} = \alpha(x/K) - k_5 r \quad (4.213)$$

$$\dot{x} = k_4 r - k_6 x \quad (4.214)$$

We want to consider a system with more degrees of freedom. Recall we have already seen one:

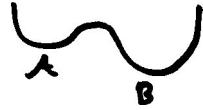


Figure 4.33: double well

$$\dot{a} = -k_1 a + k_{-1} b \quad (4.215)$$

$$\dot{b} = -k_{-1} b + k_1 a \quad (4.216)$$

Recall,  $\dot{a} + \dot{b} = 0$  so we have a conservation law. But what happens if you don't start at the steady state? We would like to show that the solution tends to these kinetic equilibria.

More generally, let

$$\dot{a} = f(a, b) \quad a = a_* + u \quad (4.217)$$

$$\dot{b} = g(a, b) \quad b = b_* + v \quad (4.218)$$

Then,

$$\underbrace{\dot{a}}_0 + \dot{u} = f(a_*, b_*, u, v) \quad (4.219)$$

$$\approx f(a_*, b_*) + u \partial_a f(a_*, b_*) + v \partial_b f(a_*, b_*) + O(u^2, v^2, uv) \quad (4.220)$$

$$\underbrace{\dot{b}}_0 + \dot{v} = g(a_*, b_*, u, v) \quad (4.221)$$

$$\approx g(a_*, b_*) + u \partial_a g(a_*, b_*) + v \partial_b g(a_*, b_*) + O(u^2, v^2, uv) \quad (4.222)$$

Recall a previous one-dimensional case,

$$\underbrace{\dot{y}_*}_0 + \dot{\eta} = \underbrace{f(y_*)}_0 + \eta f'(y_*) + \underbrace{\frac{1}{2} \eta^2 f''(y_*)}_{\approx 0} \quad (4.223)$$

Similarly, here we have Then,

$$\dot{u} = u \partial_a f(a_*, b_*) + v \partial_b f(a_*, b_*) \quad (4.224)$$

$$\dot{v} = u \partial_a g(a_*, b_*) + v \partial_b g(a_*, b_*) \quad (4.225)$$

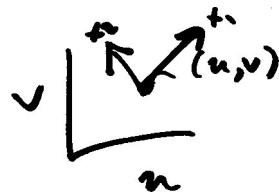


Figure 4.34: vectors

This is a pair of linear differential equations.

Since the system is linear, we can write it in matrix form:

$$\mathbf{x} = \begin{pmatrix} u \\ v \end{pmatrix} \quad (4.226)$$

$$\dot{\mathbf{x}} = \begin{pmatrix} \partial_a f & \partial_b f \\ \partial_a g & \partial_b g \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = M\mathbf{x} \quad (4.227)$$

We would like to get an equation like  $\dot{y} = \alpha y$ . For this we need to change coordinates using eigenvectors.

- HW 3: Consider the system

$$\dot{a} = k_{-1}b - k_1a \quad (4.228)$$

$$\dot{b} = k_1a - k_{-1}b \quad (4.229)$$

Letting  $\mathbf{x} = (a, b)^T$ , show that  $\dot{\mathbf{x}} \propto \mathbf{x}$ .



## Chapter 5

# Interactions: Dynamical systems

Agenda:

1 Encourage registration  
3 Encourage questions  
2 Review Checkup  
3 2D dynamical systems  
review the algorithm (some of these steps refer to algebra and/or drawing pictures)  
identify dof, coords, params  
identify fixed points  
identify bifurcations in existence  
characterize stability of all fixed points  
identify bifurcations in stability of fixed points  
review HW (classify stability of 3 FP)  
2D ODEs: what's different?  
return to MM kinetics  
connection: simple harmonic oscillator  
return to the Griffiths model  
3D: what's different?  
2D vs 3D  
importance of decay  
driven systems:  
1D  
2D  
linear (matrix) models  
connection: markov processes  
stochastics, return to FPE  
4 checkup  
5 HW  
show  $f'(P_-) > 0$   
show  $f'(P_+) < 0$

## 5.1 Dynamical systems and linear algebra

as an example of a 2-d dynamical system, look at a ball rolling at the bottom of bowl filled with a viscous fluid. Newton's 2nd law gives the second order ODE

$$m\ddot{x} = -kx - b\dot{x}. \quad (5.1)$$

letting  $v \equiv \dot{x}$ , we can transform this into a set of coupled first order ODEs:

$$\dot{x} = v \quad (5.2)$$

$$\dot{v} = -\frac{k}{m}x - \frac{b}{m}v \quad (5.3)$$

or, in matrix form

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{pmatrix} \begin{pmatrix} x \\ v \end{pmatrix}. \quad (5.4)$$

What are we to make of this?

### The simplest 2D dynamical system ever

$$\dot{x} = ax \quad (5.5)$$

$$\dot{y} = by \quad (5.6)$$

how would you characterize the stability of this (linear) dynamical system?

$a + b$

$ab$

### Return to the bowl

Sadly the bowl is not that simple. There is an off-diagonal term in the matrix which couples  $x$  and  $v$ . Another way of saying this is that the axes  $(x, v)$  are not a convenient basis for dealing with this dynamic, as the earlier basis  $(x, y)$  was. It would be great, wouldn't it, if there were some directions in  $(x, v)$  space such that the action of the matrix were as trivial as it is for just  $x$  and  $y$  above, namely multiplication by a constant.

Let's see if we can construct such magic directions, and see what the magic constants would be. That is, we will assert that there are magic directions  $\mathbf{u}$  such that  $M\mathbf{u} = \lambda\mathbf{u}$ , and we will strive to see what that says about the  $\lambda$ , since it is the behavior of the constants which tells you all you need to know about stability.

We can find eigenvectors  $\mathbf{x}$  satisfying  $A\mathbf{x} = \lambda\mathbf{x}$ , where  $A$  represents the 2-by-2 matrix above and the eigenvalues  $\lambda$  are scalars, by solving the characteristic equation

$$\det(A - \lambda\mathbb{1}) = 0 \quad (5.7)$$

which, for a 2-by-2 matrix A is equivalent to

$$\lambda^2 - \tau\lambda + \Delta = 0 \quad (5.8)$$

where  $\tau \equiv \text{tr}(A) = -\frac{b}{m}$  and  $\Delta \equiv \det(A) = \frac{k}{m}$ . from this the eigenvalues are found to be

$$\lambda_{\pm} = \frac{\tau \pm \sqrt{\tau^2 - 4\Delta}}{2} \quad (5.9)$$

$$= \frac{-\frac{b}{m} \pm \sqrt{\frac{b^2}{m^2} - \frac{4k}{m}}}{2}. \quad (5.10)$$

the eigenvectors are then determined (up to an overall multiplicative factor) from solutions of  $A\mathbf{v}_{\pm} = \lambda_{\pm}\mathbf{v}_{\pm}$ . time-dependent solutions to the original system are then given by

$$\mathbf{x}(t) = c_+ e^{\lambda_+ t} \mathbf{v}_+ + c_- e^{\lambda_- t} \mathbf{v}_-, \quad (5.11)$$

where the  $c_{\pm}$  are determined by boundary conditions.

## 5.2 Kinetic models: 2-d dynamical systems

returning to our system described by hill kinetics in the previous chapter, look at the original model:

$$\dot{m} = f_1(m(t), p(t)) \equiv \frac{p^2}{p^2 + 1} - rm \quad (5.12)$$

$$\dot{p} = f_2(m(t), p(t)) \equiv m - Rp \quad (5.13)$$

1. identify degrees of freedom, coordinates, and parameters of the model

we have 2 degrees of freedom ( $m, p$ ), one coordinate ( $t$ ), and two parameters ( $r, R$ )

2. solve for fixed points

solving for  $f_1(m_*, p_*) = 0 = f_2(m_*, p_*)$ ,

$$\frac{p_*^2}{p_*^2 + 1} = \gamma p_* \quad (5.14)$$

where  $\gamma \equiv rR$  and  $m_* = p_*/R$ . we see that this is the same condition that holds for the 1-d system.

3. identify bifurcation points

bifurcation points are same as 1-d system as well, with  $\gamma = \frac{1}{2}$  as the critical parameter value.

4. characterize stability of fixed points

look close to fixed points, at  $p_* + \eta_1$  and  $m_* + \eta_2$ , for  $|\eta_1| \ll |p_*|, |\eta_2| \ll |m_*|$ . a 2-d Taylor expansion gives

$$\dot{\eta}_1 = f_1(p_* + \eta_1) \approx f_1(p_*, m_*) + \left[ \eta_1 \frac{\partial f_1}{\partial \eta_1} + \eta_2 \frac{\partial f_1}{\partial \eta_2} \right]_{p_*, m_*} \quad (5.15)$$

$$\dot{\eta}_2 = f_2(p_* + \eta_2) \approx f_2(p_*, m_*) + \left[ \eta_1 \frac{\partial f_2}{\partial \eta_1} + \eta_2 \frac{\partial f_2}{\partial \eta_2} \right]_{p_*, m_*} \quad (5.16)$$

by definition  $f_1(m_*, p_*) = 0 = f_2(m_*, p_*)$ , which leaves the evolution of  $n_{1,2}$  in matrix form,  $\dot{\eta} = J\eta$ , as

$$\begin{pmatrix} \dot{\eta}_1 \\ \dot{\eta}_2 \end{pmatrix} = \begin{pmatrix} \partial_{\eta_1} f_1 & \partial_{\eta_2} f_1 \\ \partial_{\eta_1} f_2 & \partial_{\eta_2} f_2 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}. \quad (5.17)$$

evaluating the Jacobian<sup>1</sup> matrix, we have

$$J = \begin{pmatrix} -r & 2p_*(1+p_*^2)^{-2} \\ 1 & -R \end{pmatrix}. \quad (5.18)$$

to determine stability of fixed points, examine the real part of the eigenvalues of the Jacobian (

### Homework:

).

## 5.3 Synthetic genetic networks: oscillation

[8] [10]

### 5.3.1 HW: Transcriptional regulation review

Consider the production of mRNA. This process can be written as the dynamical system

$$\dot{r} = k_2 d_T \frac{x^n}{K^n + x^n} - k_6 r \quad (5.19)$$

$$\dot{x} = k_5 r - k_7 x \quad (5.20)$$

This comes from reactions of the form




---

<sup>1</sup>Named after Carl Gustav Jacob Jacobi (1804-1851), who wrote a long memoir De determinantibus functionalibus in 1841 devoted to this determinant (although it was first studied by Cauchy in 1815).



Figure 5.1: mRNA

If the abundance of protein goes quickly to steady state, ie  $k_7$  dominates, then we have

$$\dot{r} = k_2 d_T \frac{\left(\frac{k_5}{k_7 r}\right)^n}{K^n + \left(\frac{k_5}{k_7 r}\right)^n} - k_6 r \quad (5.23)$$

$$\dot{r} = k_2 d_T \alpha \left( \frac{r}{\bar{K}} \right) - k_6 r \quad (5.24)$$

Example:

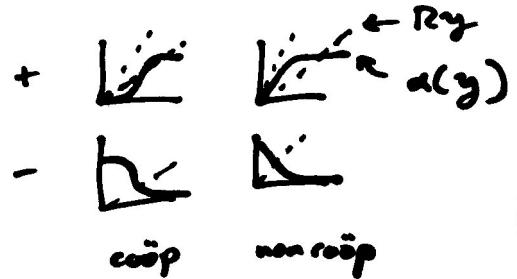
$$\dot{r} = k_2 d_T \alpha \left( \frac{r}{\bar{K}} \right) - k_6 r \quad (5.25)$$

$$y = r/\bar{K} \quad (5.26)$$

$$y' = \dot{y}/\tau \quad (5.27)$$

$$y' = \alpha(y) - Ry \quad ??? \text{maybe } \dot{y} \quad (5.28)$$

Recall the graphical method of analyzing the fixed points

Figure 5.2:  $\alpha$  plots

Stability of the fixed points  $y_*$  with  $f(y_*) = 0$  and  $\dot{y} = f(y)$  depends on

$$\frac{df}{dy}(y_*) \gtrless 0 \quad (5.29)$$

Let's look more closely. NOTE: THE DERIVATIVE NOTATION IS NOT CONSISTENT HERE

$$\dot{r} = k_2 d_T \alpha \left( \frac{r}{K} \right) - k_6 r = \frac{dr}{dt} \quad (5.30)$$

(5.31)

Introduce

$$y = r/K \quad (5.32)$$

$$\tau = t/t_0 \quad (5.33)$$

(5.34)

Then

$$K \frac{dy}{d\tau} \frac{1}{t_0} = k_2 d_T \alpha(y) - k_6 y K \quad (5.35)$$

If we choose

$$t_0 = \frac{K}{k_2 d_T} \quad (5.36)$$

we get

$$\frac{dy}{d\tau} = \alpha(y) - \underbrace{\left( \frac{k_6}{k_2 d_T} \right)}_R y \quad (5.37)$$

So  $\dot{y} \equiv dy/d\tau$ .

Consider

$$y = y_0 + \eta \quad (5.38)$$

$$\dot{\eta} = f'(y_*) \eta \quad (5.39)$$

Or graphically,

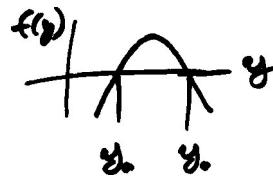


Figure 5.3:  $f(y)$  fixed points

Figure 5.4:  $\alpha$  plot

### 5.3.2 Previous homework

Recall upregulation stability

$$\dot{y} = \alpha(y) - Ry = f(y) \quad (5.40)$$

$$\alpha \equiv \alpha_2^+(y) = \frac{y^2}{1+y^2} \quad (5.41)$$

$$f' = \alpha' - R = n \frac{\alpha^2}{x^{n+1}} - R \quad (5.42)$$

NOTE:  $x = y$  notational switch.

Claim:  $\alpha' = n\alpha^2/x^{n+1}$

Proof:

$$\alpha = \frac{x^n}{1+x^n} = (1+x^{-n})^{-1} \quad (5.43)$$

$$\alpha' = -1(1+x^{-n})^{-2}(-nx^{-n+1}) = nx^{-(n+1)}\alpha^2 \quad (5.44)$$

Then stability implies

$$f' < 0 \quad (5.45)$$

$$\alpha' - R < 0 \quad (5.46)$$

$$n \frac{\alpha^2}{x^{n+1}} - R < 0 \quad (5.47)$$

$$n \frac{R^2 x^2}{x^{n+1}} < 0 \quad (5.48)$$

Now let  $n = 2$ , then stability criterion is

$$2 \frac{R}{x} < 1 \Rightarrow \frac{x}{2R} > 1 \quad (5.49)$$

Let's solve for the value of the fixed point

$$\alpha = Rx = \frac{x^2}{1+x^2} \quad (5.50)$$

$$\Rightarrow x = 0 \text{ or } R(1+x^2) = x \quad (5.51)$$

In the second case

$$Rx^2 - x + R = 0 \quad (5.52)$$

$$x_{\pm} = \frac{1 \pm \sqrt{1 - 4R^2}}{2R} \quad (5.53)$$

Thus for stability we require

$$\frac{1 \pm \sqrt{1 - 4R^2}}{4R^2} > 1 \quad (5.54)$$

If we define

$$\epsilon = 4R^2 \quad (5.55)$$

stability implies

$$1 \pm \sqrt{1 - \epsilon} > \epsilon \quad (5.56)$$

$$\pm \sqrt{1 - \epsilon} > -(1 - \epsilon) \quad (5.57)$$

To find the critical value of  $\epsilon$  and hence of  $R$  we graph both sides of the last inequality We

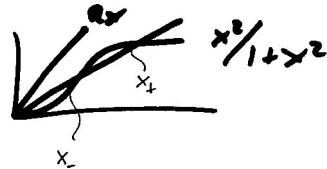


Figure 5.5:  $\epsilon$

find the critical value  $R_c = 1/2$ . We find if the decay is slow enough then the gene which is auto-upregulating is capable of two stable solutions.

So the possible trajectories are of two types depending on the initial value: What

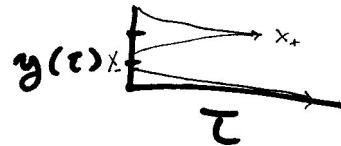


Figure 5.6:  $y(\tau)$  plot

happens right at the critical value? We have to look at the higher order terms.

$$y = y_* + \eta \quad (5.58)$$

$$\dot{\eta} = \underbrace{f'(y_*)}_{0} \eta + \frac{1}{2} \eta^2 f''(y_*) + \dots \quad (5.59)$$

So in this case, linear stability breaks down and we must consider a different dynamical system, with nonlinear terms.

Let's further explain some simplifications from earlier.

$$\dot{r} = k_2 d_T \frac{x^n}{K^n + x^n} - k_6 r \quad (5.60)$$

$$\dot{x} = k_5 r - k_7 x \quad (5.61)$$

- HW: Show this system can be nondimensionalized by putting it into the form:

$$\dot{u} = \alpha(v) - R_1 u \quad (5.62)$$

$$\dot{v} = u - R_2 v \quad (5.63)$$

You will have to choose characteristic values of  $R$  and  $x$ .

### 5.3.3 Simple two-dimensional dynamical systems

Let's consider a simple system:

$$\dot{x} = \alpha x = f(x, y) \quad (\Rightarrow x = x_0 e^{\alpha t}) \quad (5.64)$$

$$\dot{y} = \beta y = g(x, y) \quad (\Rightarrow y = y_0 e^{\beta t}) \quad (5.65)$$

First, find the fixed points and classify them. Here there is one fixed point at the origin. Hence, stability depends on the signs of  $\alpha$  and  $\beta$ .

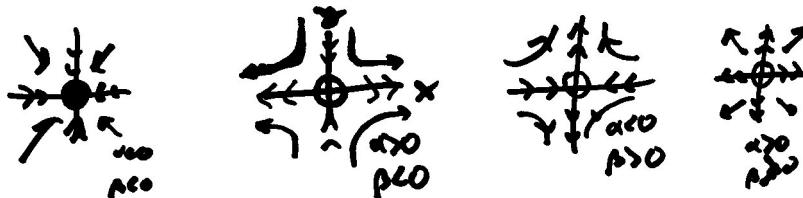


Figure 5.7: Fixed point at the origin

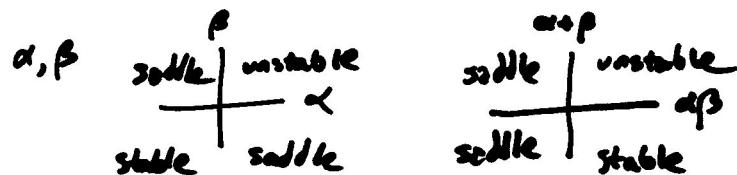


Figure 5.8: Stability in parameter space

### 5.3.4 A deeper look at dynamical systems

Consider

$$\dot{a} = -\tau_+ a + \tau_- b \quad (5.66)$$

$$\dot{b} = -\tau_- b + \tau_+ a \quad (5.67)$$

In matrix form

$$\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix} \quad (5.68)$$

$$\dot{\mathbf{x}} = \begin{pmatrix} \dot{a} \\ \dot{b} \end{pmatrix} = \begin{pmatrix} -\tau_+ & \tau_- \\ \tau_+ & -\tau_- \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} -\tau_+ a + \tau_- b \\ \tau_+ a - \tau_- b \end{pmatrix} \quad (5.69)$$

So if we think about a different coordinate system, the dynamics are in straight lines.

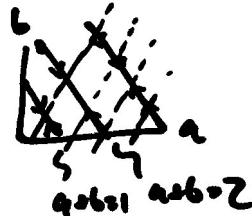


Figure 5.9:  $(a, b)$ -plane

$$(a, b) = c_0(a + b) + c_-(a + \gamma b) \quad (5.70)$$

Consider a general system with solutions which move in straight lines.

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix} \quad (5.71)$$

Solving,

$$Ax + By = \lambda x \Rightarrow By = \lambda x - Ax \quad (5.72)$$

$$Cx + Dy = \lambda y \Rightarrow BCx + (D - \lambda)(\lambda x - Ax) = 0 \quad (5.73)$$

The second equation

$$BCx + (D - \lambda)(\lambda - A)x = 0 \quad (5.74)$$

yields the solution  $x = 0$  and

$$BC + D\lambda - \lambda^2 - AD + A\lambda = 0 \quad (5.75)$$

$$\lambda^2 - \lambda(A + D) + AD - BC = 0 \quad (5.76)$$

We define the trace and determinant

$$\tau \equiv A + D \quad (5.77)$$

$$\Delta \equiv AD - BC \quad (5.78)$$

so that

$$\lambda^2 - \tau\lambda + \Delta \quad (5.79)$$

This induces a “better” coordinate system.

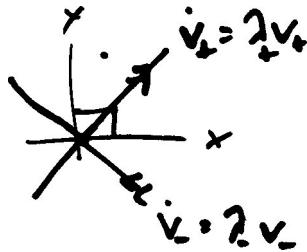


Figure 5.10: Eigenvector coordinates

$$\dot{x} = \alpha x \quad (5.80)$$

$$\dot{y} = \beta y \quad (5.81)$$

or

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (5.82)$$

where

$$\tau = \alpha + \beta \quad (5.83)$$

$$\Delta = \alpha\beta \quad (5.84)$$

and

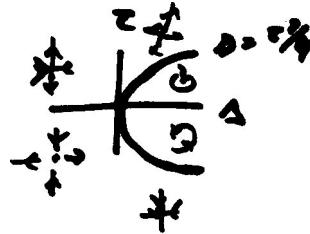
$$\lambda = \frac{\tau \pm \sqrt{\tau^2 - 4\Delta}}{2} \quad (5.85)$$

Note there is a change in behavior at  $4\Delta = \tau^2$ .

- HW: Find  $\lambda_+, \lambda_-$  for the damped harmonic oscillator equation

$$m\ddot{x} = -b\dot{x} - kx \quad (5.86)$$

Hint: introduce  $v = \dot{x}$  and write as a first order system.

Figure 5.11: Stability in  $(\tau, \Delta)$ -plane

### 5.3.5 HW: cooperative upregulation

A general ODE

$$\dot{x} = f(x) \quad (5.87)$$

$$f(x_*) = 0 \quad (5.88)$$

$$x = x_* + \eta \quad (5.89)$$

Plugging this into the ODE leads to

$$\underbrace{\dot{x}_*}_{0} + \dot{\eta} = f(x_* + \eta) \approx \underbrace{f(x_*)}_{0} + \eta f'(x_*) \quad (5.90)$$

$$\dot{\eta} = \eta f'(x_*) \quad (5.91)$$

$$(5.92)$$

which gives the solution

$$\eta = \eta_0 e^{tf'(x_*)} \quad (5.93)$$

We have stability at the fixed point if

$$f'(x_*) < 0 \quad (5.94)$$

In the case of cooperative upregulation

$$\dot{r} = k_{2d_T} \alpha(p/K) - k_7 r \quad (5.95)$$

$$\dot{p} = k_6 r - k_5 p \quad (5.96)$$

To nondimensionalize, introduce these constants and derivatives

$$y = r/\bar{r} \quad (5.97)$$

$$x = p/K \quad (5.98)$$

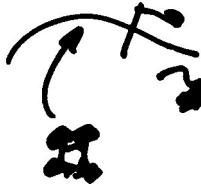


Figure 5.12: DNA

$$t = \tau \bar{t} \quad (5.99)$$

$$\dot{y} = \frac{dy}{d\tau} \quad (5.100)$$

$$\dot{r} = \frac{dy}{dt} \quad (5.101)$$

Making these substitutions

$$\frac{\bar{r}}{\bar{t}} \dot{y} = k_2 d_T \alpha(x) - k_7 \bar{r} y \quad (5.102)$$

$$\frac{K}{\bar{t}} \dot{x} k_6 \bar{r} - k_5 K x \quad (5.103)$$

Simplifying we have

$$\dot{y} = \left( \frac{k_2 d_T \bar{t}}{\bar{t}} \right) \alpha(x) - k_7 \bar{t} y \quad (5.104)$$

$$\dot{x} = \left( \frac{k_6 \bar{r} \bar{t}}{K} \right) y - k_5 \bar{t} x \quad (5.105)$$

Choose characteristic space and time scales conveniently. Let

$$\bar{r} \bar{t} = \frac{K}{k_6} \quad (5.106)$$

$$\frac{\bar{t}}{\bar{r}} = \frac{1}{k_2 d_T} \quad (5.107)$$

Hence, the differential equations for transcription and translation simplify to

$$\dot{y} = \alpha(x) - R_1 y \quad (5.108)$$

$$\dot{x} = y - R_2 x \quad (5.109)$$

Now assume  $R_1$  is very fast, ie  $R_1 \gg 1$  and thus  $y = \alpha/R_1$ . This eliminates the equation in  $y$ :

$$\dot{x} = \frac{\alpha}{R_1} - R_2 x \quad (5.110)$$

$$R_1 \dot{x} = \alpha - R_1 R_2 x \quad (5.111)$$

Introduce a new time scale  $s$  such that

$$R_1 \frac{d}{d\tau} = \frac{d}{ds} \quad (5.112)$$

Then,

$$\dot{x} = \alpha - \underbrace{R_1 R_2}_R x \quad (5.113)$$

Notice the five initial parameters collapse to just one by combining transcription and translation into one step. The one interesting parameter is  $R = k_5 k_7 \bar{t}^2$ . Now, we analyze the equation near fixed points.

$$\dot{x} = 0 = \alpha(x) - Rx \quad (5.114)$$

$$\frac{x^n}{1+x^n} = Rx \quad (5.115)$$

Solutions are

$$x = 0 \quad (5.116)$$

$$x^{n-1} = R(1+x^n) \quad (5.117)$$

In the case of dimerization,  $n = 2$  and the second fixed point equation is

$$x = R + Rx^2 \quad (5.118)$$

$$Rx^2 - x + R = 0 \quad (5.119)$$

$$x_* = \frac{1 \pm \sqrt{1 - 4R^2}}{2R} \quad (5.120)$$

Graphically,

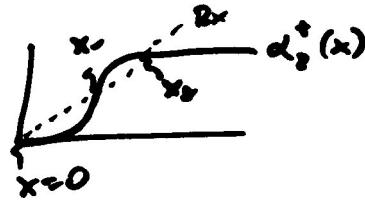


Figure 5.13: Two fixed points

$$\dot{x} = f(x) \quad (5.121)$$

$$f' \geq 0 \quad (5.122)$$

In our case

$$f = \alpha(x) - Rx \quad (5.123)$$

$$f' = \alpha' - R \quad (5.124)$$

$$\alpha = (1 + x^{-n})^{-1} \quad (5.125)$$

$$\alpha' = -(1 + x^{-n})^{-2}(-nx^{-n-1}) = \alpha^2 nx^{-(n+1)} \quad (5.126)$$

$$f' = \alpha' - R' = \alpha^2 nx^{-(n+1)} - R = R^2 nx^{1-n} - R \quad (5.127)$$

Then stability,  $f' < 0$  requires

$$R^2 nx^{1-n} < R \quad (5.128)$$

For  $n = 2$ ,

$$R^2 \frac{2}{x} < R \quad (5.129)$$

$$2R < x \quad (5.130)$$

$$2R < \frac{1}{2R} (1 \pm \sqrt{1 - 4R^2}) \quad (5.131)$$

$$4R^2 < 1 \pm \sqrt{1 - 4R^2} \quad (5.132)$$

$$-(1 - 4R^2) < \pm \sqrt{1 - 4R^2} \quad (5.133)$$

Now we need to investigate the existence of this fixed point. A bifurcation exists at  $R = 1/2$  where the two fixed points coalesce. The system goes to an “on” or “off” state depending on the initial conditions.

### 5.3.6 Genetic engineering and “synthetic biology”

Groups ordered some plasmid:

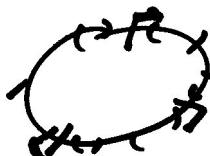


Figure 5.14: Plasmid

What genes to use? Look at mathematical models for a hint.  
N-cycles

### HW: Simple 2D dynamical systems

Finishing the two-dimensional dynamical system discussion we had

$$\dot{y} = \alpha(x) - R_1 x \quad (5.134)$$

$$\dot{x} = y - R_2 x \quad (5.135)$$



Figure 5.15: N-cycle

In general

$$\dot{x} = f(x, y) \quad (5.136)$$

$$\dot{y} = g(x, y) \quad (5.137)$$

$$(5.138)$$

We linearize about fixed points

$$f(x_*, y_*) = g(x_*, y_*) \quad (5.139)$$

$$x = x_* u \quad (5.140)$$

$$y = y_* v \quad (5.141)$$

$$\dot{u} = \underbrace{f(x_*, y_*)}_{0} + u \partial_x f|_* + v \partial_y f|_* \quad (5.142)$$

$$\dot{v} = \underbrace{g(x_*, y_*)}_{0} + u \partial_x g|_* + v \partial_y g|_* \quad (5.143)$$

This leads to the matrix equation

$$\begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = \underbrace{\begin{pmatrix} \partial_x f & \partial_y f \\ \partial_x g & \partial_y g \end{pmatrix}}_{\text{Jacobian, } \nabla(f,g)} \begin{pmatrix} u \\ v \end{pmatrix} \quad (5.144)$$

Recall, the simple system

$$\dot{x} = \alpha x \quad (5.145)$$

$$\dot{y} = \beta y \quad (5.146)$$

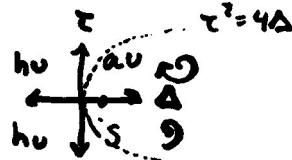
or

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (5.147)$$

where

$$\tau = \alpha + \beta \quad (5.148)$$

$$\Delta = \alpha \beta \quad (5.149)$$

Figure 5.16: Stability diagram in  $(\tau, \Delta)$ -plane

Stability can be determined with the help of a diagram.

Regions in the diagram can be absolutely stable, stable, or hyperbolically unstable (saddle points).

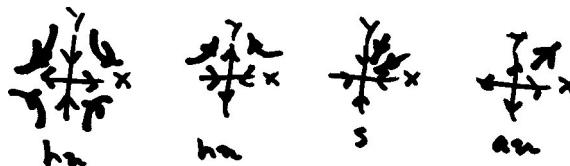


Figure 5.17: Stability cases

### HW: damped harmonic oscillator

The damped harmonic oscillator is a good example

$$F = ma = F_v + F_s \quad (5.150)$$

$$m\ddot{x} = -b\dot{x} - kx \quad (5.151)$$

$$v = \dot{x} \quad (5.152)$$

As a system

$$\begin{pmatrix} \dot{x} \\ v \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -k/m & -b/m \end{pmatrix} \begin{pmatrix} x \\ v \end{pmatrix} \quad (5.153)$$

So there is a fixed point at the origin. What are the trace and determinant?

$$\tau = -b/m \quad (5.154)$$

$$\Delta = k/m \quad (5.155)$$

For the case  $b = 0$ , there is no drag and trajectories are circles in the  $(x, v)$ -plane.

Let's look for the eigendirections for a general problem

$$\begin{pmatrix} \dot{x} \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix} \quad (5.156)$$

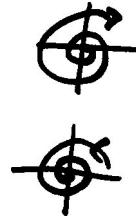


Figure 5.18: spiral trajectories

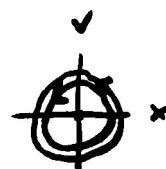


Figure 5.19: circle trajectories

We want to find  $y/x$  and  $\lambda$ .

$$ax + by = \lambda x \quad (5.157)$$

$$cx + dy = \lambda y \quad (5.158)$$

These equations imply

$$by = \lambda x - ax = (\lambda - a)x \quad (5.159)$$

$$bcx = -by(\lambda - d) \quad (5.160)$$

Solving

$$bcx = (\lambda - a)x(\lambda - d) \quad (5.161)$$

$$bc = \lambda^2 - (a + d)\lambda + ad \quad (5.162)$$

$$(5.163)$$

We get

$$\lambda^2 - \tau\lambda + \Delta = 0 \quad (5.164)$$

which has solutions

$$\lambda = \frac{\tau \pm \sqrt{\tau^2 - 4\Delta}}{2} \quad (5.165)$$

If there is no drag  $b = 0$  we have

$$\lambda = \pm i\sqrt{\Delta} = \pm i\sqrt{k/m} = \pm i\omega \quad (5.166)$$

So imaginary eigenvalues correspond to neutral stability or instability.



Figure 5.20: stability

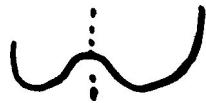
**HW: Double well potential dynamics**

Figure 5.21: double well

The equations are

$$\dot{p}_- = -\tau_+ p_- + \tau_- p_+ \quad (5.167)$$

$$\dot{p}_+ = -\tau_- p_+ + \tau_+ p_- \quad (5.168)$$

or

$$\begin{pmatrix} \dot{p}_- \\ \dot{p}_+ \end{pmatrix} = \begin{pmatrix} -\tau_+ & \tau_- \\ \tau_+ & -\tau_- \end{pmatrix} \begin{pmatrix} p_- \\ p_+ \end{pmatrix} \quad (5.169)$$

What are the eigenvalues? According to our intuition,  $\lambda_+ = 0$  and  $\lambda_- < 0$ . Let's see.

$$\tau = -(\tau_+ + \tau_-) \quad (5.170)$$

$$\Delta = \tau_+ \tau_- - \tau_- \tau_+ = 0 \quad (5.171)$$



Figure 5.22: phase plane for double well

$$\lambda^2 - \tau \lambda + \Delta = 0 \quad (5.172)$$

$$\lambda(\lambda - \tau) = 0 \quad (5.173)$$

So we have eigenvalues

$$\lambda_+ = 0 \quad (5.174)$$

$$\lambda_- = \tau < 0 \quad (5.175)$$

Now let's find the eigenvectors.



Figure 5.23: eigendirections

$$\begin{pmatrix} -\tau_+ & \tau_- \\ +\tau_+ & -\tau_- \end{pmatrix} \begin{pmatrix} a_- \\ b_- \end{pmatrix} = \underbrace{-(\tau_+ + \tau_-)}_{\lambda_-} \begin{pmatrix} a_- \\ b_- \end{pmatrix} \quad (5.176)$$

Solving the first equation

$$\underbrace{-\tau_+ a_- + \tau_- b_-}_{0} = \underbrace{-\tau_+ a_- - \tau_- a_-}_{0} \quad (5.177)$$

$$b_- = -a_- \quad (5.178)$$

The same can be done for the second equation. We find the eigenvector

$$\mathbf{v}_- = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad (5.179)$$

We can similarly compute the other eigenvector

$$\begin{pmatrix} -\tau_+ & \tau_- \\ \tau_+ & -\tau_- \end{pmatrix} \begin{pmatrix} a_+ \\ b_+ \end{pmatrix} = 0 \quad (5.180)$$

Solving as before

$$-\tau_+ a_+ + \tau_- b_+ = 0 \quad (5.181)$$

$$b_+ = \tau_+ / \tau_- \quad (5.182)$$

we find

$$\mathbf{v}_+ = \begin{pmatrix} 1 \\ \tau_+ / \tau_- \end{pmatrix} \quad (5.183)$$

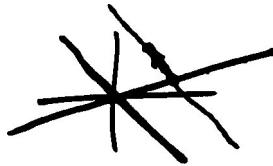


Figure 5.24: eigenvector plot

Let's verify the eigenvalues and eigenvectors.

$$\begin{pmatrix} p_+ \\ p_- \end{pmatrix} = c_+(t) \begin{pmatrix} 1 \\ \tau_+/\tau_- \end{pmatrix} + c_-(t) \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad (5.184)$$

$$\begin{pmatrix} \dot{p}_+ \\ \dot{p}_- \end{pmatrix} = \mu \begin{pmatrix} p_+ \\ p_- \end{pmatrix} = c_+(t) \begin{pmatrix} 1 \\ \tau_+/\tau_- \end{pmatrix} + c_-(t) \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad (5.185)$$

$$= \dot{c}_+(t) \begin{pmatrix} 1 \\ \tau_+/\tau_- \end{pmatrix} + \dot{c}_-(t) \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad (5.186)$$

$$\begin{pmatrix} -\tau_+ & \tau_- \\ \tau_+ & -\tau_- \end{pmatrix} \begin{pmatrix} 1 \\ \tau_+/\tau_- \end{pmatrix} = \begin{pmatrix} -\tau_+ + \tau_- (\tau_+/\tau_-) \\ \tau_+ - \tau_- (\tau_+/\tau_-) \end{pmatrix} \quad (5.187)$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (5.188)$$

So

$$\dot{c}_+ = 0 = \lambda_+ c_+ \quad (5.189)$$

Using the other eigenvector

$$\begin{pmatrix} -\tau_+ & \tau_- \\ \tau_+ & -\tau_- \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -\tau_+ - \tau_- \\ \tau_+ - \tau_- \end{pmatrix} \quad (5.190)$$

$$= -(\tau_+ + \tau_-) \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad (5.191)$$

So

$$\dot{c}_- = -(\tau_+ + \tau_-) c_- = \lambda_- c_- \quad (5.192)$$

### 5.3.7 HW: Transcription, translation

Recall the results for one-dimensional dynamics:

1. Origin is always stable
2. Fixed point  $x_-$  is unstable if and only if  $R_1 R_2 < 1/2$

3. Fixed point  $x_+$  is stable if and only if  $R_1 R_2 < 1/2$

Now let's consider the two-dimensional dynamics

$$\dot{y} = \alpha(x) - R_1 y \quad (5.193)$$

$$\dot{x} = y - R_2 x \quad (5.194)$$

$$\alpha(x) = \frac{x^2}{1+x^2} \quad (5.195)$$

The Jacobian is

$$J = \begin{pmatrix} -R_1 & \alpha' \\ 1 & -R_2 \end{pmatrix} \quad (5.196)$$

The trace and determinant are

$$\tau = -(R_1 + R_2) \quad (5.197)$$

$$\Delta = R_1 R_2 - \alpha' = R - \alpha' \quad (5.198)$$



Figure 5.25:  $\tau, \Delta$

In the one-dimensional case, stability required  $f' = \alpha' - R < 0$ . In two dimensions we require  $\tau < 0$  and  $\Delta > 0$  which is equivalent. Finishing the two-dimensional case

$$\alpha(x) = R_1 y \quad (5.199)$$

$$y = R_2 x \quad (5.200)$$

So

$$\alpha(x) = R_1 R_2 x = Rx \quad (5.201)$$

This leads to the same result as in the one-dimensional analysis.

$$\frac{x^2}{1+x^2} = Rx \quad (5.202)$$

has solutions

$$x = 0 \quad (5.203)$$

$$x = R(1+x^2) \quad (5.204)$$

$$x = \frac{1}{2R}(-1 \pm \sqrt{1-4R^2}) \quad (5.205)$$

The rest is the same as the one-dimensional analysis.

At the origin

$$\alpha' = 0 \quad \text{for } n > 1 \quad (5.206)$$

$$J = \begin{pmatrix} -R_1 & \alpha' \\ 1 & -R_2 \end{pmatrix} = \begin{pmatrix} -R_1 & 0 \\ 1 & -R_2 \end{pmatrix} \quad (5.207)$$

So

$$\tau = -(R_1 + R_2) \quad (5.208)$$

$$\Delta = R_1 R_2 > 0 \Rightarrow \text{stable} \quad (5.209)$$

- HW: N-repressor cycle

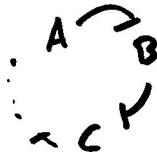


Figure 5.26: N-repressor cycle

Partial credit for  $N = 2$ .

Consider the system

$$\dot{x}_j = \alpha(x_{j+1}) - Rx_j \quad (5.210)$$

Choose

$$\alpha = \frac{1}{1 + x^n} \quad (5.211)$$

Write down the stability criterion. Consider the symmetric case:

$$x_1^* = x_2^* = x_3^* = \dots = x_N^* x_{N+1} = x_1 \quad (\text{N-cycle}) \quad (5.212)$$

Write expression of real part of eigenvalue

$$\lambda_r(n, N, R) \quad (5.213)$$

Stability requires  $\lambda_r < 0$



# Chapter 6

## Noise

### 6.1 Motivating example: Van Kampen's cat

### 6.2 Review of simple chemical kinetics

#### 6.2.1 Decay: $\dot{x} = -rx$

decay to 0

#### 6.2.2 Constitutive creation: $\dot{x} = \alpha_0$

#### 6.2.3 Two-state systems: $N_a + N_b = constant$

#### 6.2.4 Decay to a nonzero fixed point: $\dot{x} = \alpha_0 - rx$

#### 6.2.5 Nonlinearity: $\dot{x} = -rx + x^2/(1 + x^2)$

bistability or decay to 0, depending on r

#### 6.2.6 Transcription-translation models

$$\dot{x} = -Rx + y \quad (6.1)$$

$$\dot{y} = -ry + x^2/(1 + x^2) \quad (6.2)$$

#### 6.2.7 Oscillations

Consider two proteins which decay only when they couple and obey

$$\dot{x} = -x + y^2/(1 + y^2) - xy \quad (6.3)$$

$$\dot{y} = -y + 1/(1 + x^2) - xy \quad (6.4)$$

### 6.3 Stochastic kinetics

#### 6.3.1 Constitutive creation: $0 \rightarrow x$

#### 6.3.2 Decay: $0 \rightarrow x$

$$\dot{p}_n = \alpha_0 p_{n-1} - \alpha p_n \quad (6.5)$$

This is a dynamic we can solve completely.

#### 6.3.3 Decay: $x \rightarrow 0$

$$\dot{p}_n = -rnp_n + r(n+1)p_{n+1} \quad (6.6)$$

This dynamic is more difficult to solve, but we can still look at the macroscopic description for  $\langle n \rangle$ .

#### 6.3.4 Decay to a fixed point: $0 \leftrightarrow x$

$\dot{x} = \alpha_0 - rx$  Really two reactions are taking place here concurrently:  $0 \rightarrow x$  and  $x \rightarrow 0$

$$\dot{p}_n = \alpha_0 p_{n-1} - \alpha p_n - rnp_n + r(n+1)p_{n+1} \quad (6.7)$$

Consider the simple dynamic  $\dot{x} = \alpha - rx$ , corresponding to decay towards the fixed point  $x_* \equiv \alpha/r$ .

The corresponding master equation reads

$$\dot{p}_n = -rnp_n - \alpha_0 p_n + \alpha_0 p_{n-1} + r(n+1)p_{n+1} \quad (6.8)$$

where  $\alpha_0 = V\alpha$  and  $V$  is the volume of the cell.

Rather than solve for the full dynamic, let us consider the statistical steady state, for which, defining  $n_* = \alpha_0/r$ ,

$$0 = np_n - n_*p_n + n_*p_{n-1} + (n+1)p_{n+1} \quad (6.9)$$

$$0 = -n_*p_0 + (1)p_1. \quad (6.10)$$

The appearances of terms such as  $np_n$  and  $(n+1)p_{n+1}$  suggests a change of normalization; defining  $z_n$  below, defining  $z_n$  below,

$$p_n \equiv \frac{n_*^n}{n!} z_n \quad (6.11)$$

such that the Eqns. 6.9, 6.10 may be rewritten

$$0 = \frac{n_*^n z_n}{(n-1)!} + \frac{n_*^n z_{n-1}}{(n-1)!} - \frac{n_*^{n+1} z_n}{n!} + \frac{n_*^{n+1} z_{n+1}}{n!} \quad (6.12)$$

$$0 = -n_* z_0 + n_* z_1 \quad (6.13)$$

#### 6.4. THE GENERAL CASE: $\dot{P}_N = -J(N) + J(N - 1)$

Dividing Eqn. 6.12 by the common factor of  $n_*^n / (n - 1)!$  yields

$$z_n - z_{n-1} + \frac{n_*}{n} z_n = \frac{n_*}{n} z_{n+1} \quad (6.14)$$

$$z_{n+1} = z_n + \frac{n}{n_*} (z_n - z_{n-1}) \quad (6.15)$$

Since  $z_0 = z_1$  from Eqn. 6.13, we have  $z_n = z_0 \forall n$  and

$$p = z_0 \frac{n_*^n}{n!}. \quad (6.16)$$

Normalization then requires  $z_0 = e^{-n_*}$ ;  $z_0$  can also be interpreted as a moment generating function since

$$\begin{aligned} \langle n \rangle &= \sum n p_n = z_0 \sum n \frac{n_*^n}{n!} = z_0 n \partial_{n_*} \sum \frac{n_*^n}{n!} = e^{-n_*} n_* \partial_{n_*} e^{n_*} = n_* \\ \langle n(n-1) \rangle &= z_0 \sum n(n-1) \frac{n_*^n}{n!} = z_0 n_*^2 \partial_{n_*}^2 \sum \frac{n_*^n}{n!} = z_0 n_*^2 \partial_{n_*}^2 e^{n_*} = n_*^2 \\ \sigma_n^2 &= \langle n^2 \rangle - \langle n \rangle^2 = \langle n(n-1) \rangle + \langle n \rangle - \langle n \rangle^2 = n_*^2 + n_* - n_*^2 = n_* \end{aligned}$$

##### 6.3.5 Two-state systems: $N_a + N_b = constant$

$$x_- \leftrightarrow x_+$$

$$\dot{N}_\pm = -r_\pm N_\pm + r_\mp N_\mp \quad (6.17)$$

#### 6.4 The general case: $\dot{p}_n = -j(n) + j(n - 1)$

Many master equations are ultimately of the form  $\dot{p}_n = -j(n) + j(n - 1)$ . For example, for particle creation,  $\dot{p}_n = -\alpha_0 p_n + \alpha p_{n-1}$  for  $n > 0$ , and  $\dot{p}_0 = -\alpha_0 p_0$ . We can then demand  $p_n = 0$  for all  $n < 0$ . So that the equation holds for all  $n$ .

Or, for particle decay,  $\dot{p}_n = -r n p_n + r(n + 1)p_{n+1}$  for  $n \geq 0$ , and  $j(n) = -r(n + 1)p_{n+1}$ .

##### 6.4.1 Conservation

Preservation of normalization reads, if the set of allowed states runs from  $n_1$  to  $n_2$ ,

$$\frac{d}{dt} 1 = \frac{d}{dt} \sum_n p_n = - \sum_n j(n) + \sum_n j(n - 1) = 0 \quad (6.18)$$

$$0 = - \sum_{m=n_1}^{n_2} j(m) + \sum_{m=n_1-1}^{m=n_2-1} j(m) = j(n_2) - j(n_1 - 1). \quad (6.19)$$

Note the similarity with the continuum description,

$$\dot{p} = -\partial_x J \Rightarrow 0 = J(x_2) - J(x_1) \quad (6.20)$$

For example, for particle creation, take  $p(n_2) = p(\infty) = 0$ . Then, for  $n_1 < 1$  (e.g.,  $n_1 = 0$ ), probability is conserved.

Or, for particle decay,  $n_2 = \infty$  as before and  $n_1 = 0, j(n_1 - 1) = -r(0)p_0 = 0$ , and probability is conserved.

### 6.4.2 Moment dynamics

In general,

$$\frac{d}{dt} \langle n \rangle = \frac{d}{dt} \sum_n np_n = - \sum_n nj(n) + \sum_n j(n-1) \quad (6.21)$$

$$= - \sum_{n=n_1}^{n=n_2} nj(n) + \sum_{m=n_1-1}^{m=n_2-1} (m+1)j(m) \quad (6.22)$$

$$= -n_2 j(n_2) + n_1 j(n_1 - 1) + \sum_{n_1}^{n_2-1} j(n) \quad (6.23)$$

### Homework:

1. Read the paper “Control, exploitation and tolerance of intracellular noise” by Christopher V. Rao, Denise M. Wolf, and Adam P. Arkin. (Nature **420**, 14 November 2002, p 231.) [?]
2. Show that, in the limit  $q \ll 1, j \ll N, N \gg q$ , the binomial distribution

$$\binom{N}{j} q^j (1-q)^{N-j} \quad (6.24)$$

becomes the Poisson distribution

$$\frac{\lambda^j}{j!} e^{-\lambda} \quad (6.25)$$

where  $\lambda \equiv qN$ . You will need Stirling’s approximation for  $M \gg 1$

$$M! \approx M^M e^{-M} \sqrt{2\pi M} \quad (6.26)$$

as well as the limit

$$\lim_{M \rightarrow \infty} (1 + x/M)^M = e^x. \quad (6.27)$$

3. show that this distribution is normalized (over  $j$ )
4. Show that the dynamic  $\dot{p}_n = -j_n + j_{n-1}$  preserves normalization  $\iff j(-1) = 0$ .
5. Show that the master equation  $\dot{p}_n = -rnp_n + r(n+1)p_{n+1}$  implies  $A_t = -rA$  for  $A \equiv \langle n \rangle$ .

6. Show that the binomial distribution (given on p232 of Arkin's review article)

$$P(x, y; t) = p(x) = \frac{N!}{x!y!} k_2^x k_1^y \quad (6.28)$$

(a function of only one independent variable since  $y = N - x$ ) is a solution to the two-state master equation (given on p233 of Arkin's review article)

$$\frac{dP(n_a, n_b; t)}{dt} = -(k_1 n_a + k_2 n_b) P(n_a, n_b; t) + \quad (6.29)$$

$$k_1(n_a + 1) P(n_a + 1, n_b - 1; t) + \quad (6.30)$$

$$k_2(n_b + 1) P(n_a - 1, n_b + 1; t). \quad (6.31)$$

**Answer:**

First, note that

$$\begin{aligned} \frac{P(n_a + 1, n_b + 1; t)}{P(n_a, n_b; t)} &= \\ &= \frac{\frac{N!}{(n_a + 1)!(n_b + 1)!} k_2^{n_a+1} k_1^{n_b+1} \frac{1}{N!} (n_a)!(n_b)! k_2^{-n_a} k_1^{-n_b}}{(n_a + 1)!(n_b + 1)!} \quad (6.33) \end{aligned}$$

$$\begin{aligned} \frac{P(n_a - 1, n_b + 1; t)}{P(n_a, n_b; t)} &= \\ &= \frac{\frac{N!}{(n_a - 1)!(n_b + 1)!} k_2^{n_a-1} k_1^{n_b+1} \frac{1}{N!} (n_a)!(n_b)! k_2^{-n_a} k_1^{-n_b}}{(n_a - 1)!(n_b + 1)!} \quad (6.34) \end{aligned}$$

then the steady state condition reads, dividing through by  $P(n_a, n_b; t)$ ,

$$\begin{aligned} 0 &= -(k_1 n_a + k_2 n_b) + k_1(n_a + 1) \frac{k_2}{k_1} \frac{n_b}{n_a + 1} + k_2(n_b + 1) \frac{k_1}{k_2} \frac{n_a}{n_b + 1} \\ 0 &= -k_1 n_a - k_2 n_b + k_2 n_b + k_1 n_a \end{aligned} \quad (6.36)$$

## 6.5 Return to the cat

### 6.6 Lecture 14

#### 6.6.1 HW questions

- Repressor cycles There were a variety of experiments done with genes repressing each other. See Figure 6.1.

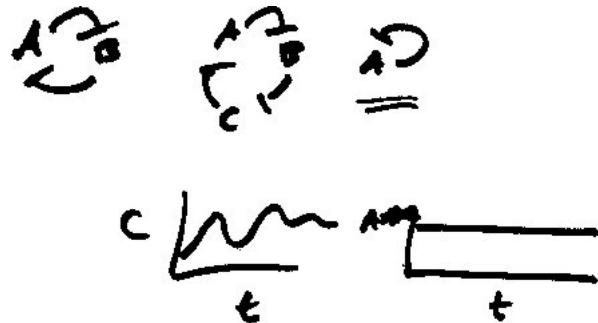


Figure 6.1: repressor cycles

HW problem was to study the general case in Figure 6.2 of an N-repressor cycle with

$$x_1 = x_2 = x_3 = \dots = x_N \quad (6.37)$$



Figure 6.2: N-repressor cycle

The equations are

$$\dot{x}_{j+1} = -x_{j+1} + \gamma \frac{1}{1 + x_j^n} \quad (6.38)$$

And schematically

$$x_j \dashv x_{j+1} \dashv x_{j+2} \dots \quad (6.39)$$

This is a cycle as in Figure 6.3.

We study this cycle as a dynamical system:

$$\dot{x} = f(x) \quad (6.40)$$

$$f(x_*) = 0. \quad (6.41)$$

We consider small perturbations from equilibrium:

$$x = x_* + \eta \quad (6.42)$$

$$\dot{\eta} = (\nabla f)\eta \quad (6.43)$$

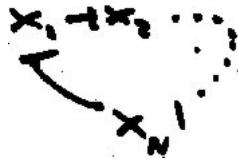


Figure 6.3: HW cycle

We have stability if the real parts of the eigenvalues are negative,  $\lambda_r < 0$  as in Figure 6.4.



Figure 6.4: Stable spiral

Consider the Jacobian

$$J_{n_j} = \partial_{x_k} \dot{x}_j \quad (6.44)$$

where

$$\dot{x}_j = -Rx_j + \alpha(x_{j+1}). \quad (6.45)$$

As a matrix it looks like

$$J = \begin{pmatrix} -R & \alpha' & & & \\ & -R & \alpha' & & \\ & & -R & \alpha' & \\ & & & \ddots & \ddots \\ & & & & -R & \alpha' \\ & & & & & -R \end{pmatrix}. \quad (6.46)$$

The  $N$ th element is

$$\dot{x}_N = -Rx_N + \alpha'(x_1). \quad (6.47)$$

The eigenvalue equation is

$$J\mathbf{v} = \lambda\mathbf{v} \quad (6.48)$$

or in components

$$-Rv_j + \alpha'v_{j+1} = \lambda v_j \quad (6.49)$$



Figure 6.5: values

for  $v_1, v_2, \dots$

The value of  $\alpha'$  may just be a constant as in Figure 6.5.

Rewriting the equations

$$v_{j+1} = \left( \frac{\lambda + R}{\alpha'} \right) v_j = q v_j \quad (6.50)$$

$$v_2 = q v_1 \quad (6.51)$$

$$v_3 = q v_2 = q^2 v_1 \quad (6.52)$$

$$v_N = q^N v_1 \quad (6.53)$$

$$v_1 = q^{N+1} v_1. \quad (6.54)$$

This means

$$q^{N+1} = 1 = e^{2\pi i m} \quad (6.55)$$

$$q^{N+1} = e^{2\pi i m} \quad (6.56)$$

$$q = e^{(2\pi i m)/(N+1)} = \frac{\lambda + R}{\alpha'} \quad (6.57)$$

So

$$\lambda = -R + \alpha' e^{(2\pi i m)/(N+1)} \quad (6.58)$$

$$\lambda_r = -R + \alpha' \cos \left( \frac{2\pi m}{N+1} \right). \quad (6.59)$$

We know the fixed point satisfies

$$x_* = -Rx_* + \alpha(x_*) \quad (6.60)$$

so

$$\alpha = \frac{1}{1+x^n} \quad (6.61)$$

$$\alpha' = -\left( \frac{1}{1+x^n} \right)^2 nx^{n-1} = -\alpha^2 nx^{n-1} \quad (6.62)$$

$$= -R^2 nx^{n+1}. \quad (6.63)$$

Then we have

$$\lambda_r = -R - R^2 n x^{n+1} \cos\left(\frac{2\pi m}{N+1}\right). \quad (6.64)$$

Consider  $\cos$  in Figure 6.6. For some  $m$ , we get  $\cos = -1$ .

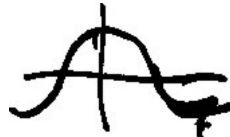


Figure 6.6: Cosine



Figure 6.7: N-cycle

Then let's investigate

$$\lambda_r \approx -R + R^2 n x^{n+1}. \quad (6.65)$$

We know at fixed point

$$Rx = \frac{1}{1+x^n}. \quad (6.66)$$

Then

$$Rx + Rx^{n+1} \approx 1 \quad (6.67)$$

$$x \approx R^{-\frac{1}{n+1}} \quad (6.68)$$

$$x \approx 1/R. \quad (6.69)$$

We have some MATLAB code to plot the eigenvalue for different Hill coefficients  $n$ . The code uses

$$x' = -x + \gamma \frac{1}{-1+x^n} \quad (6.70)$$

And the plot shows

$$-1 + \frac{nx^{n+1}}{\gamma}. \quad (6.71)$$

That is how to drive the system out of the boring state. This completes the discussion of Figure 6.8

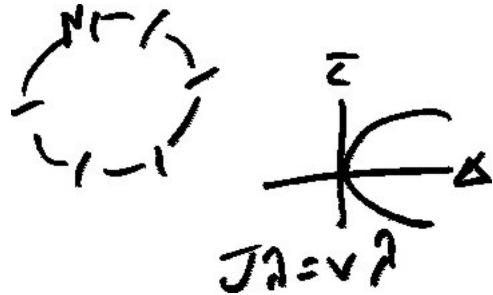


Figure 6.8: N-cycle

### 6.6.2 Viruses and life and death

Let's try to calculate virus count. Say the concentration is as in Figure 6.9.



Figure 6.9: Virus count

A molar is

$$M \sim \frac{6.2 \cdot 10^{23}}{L} \quad (6.72)$$

$$L = (10 \text{ cm})^3 \quad (6.73)$$

Recall

$$mL = (1 \text{ cm})^3 \quad (6.74)$$

$$10^3 L = 10^3 \text{ cm}^3. \quad (6.75)$$

So the total number of copies is

$$\# = 10 \text{ nM} \cdot v_{cell} = \frac{10 \cdot 10^{-9} \cdot 6 \cdot 10^{23}}{(10 \text{ cm})^3} (1\mu)^3 \left( \frac{10^{-4} \text{ cm}}{1\mu} \right)^3 \quad (6.76)$$

$$\sim 10^{1-9-3+23-12} = 10^0 \sim 1 \quad (6.77)$$



Figure 6.10: E. Coli. RNA

So it is very important at this scale to remember we are dealing with a discrete number of copies of these molecules, as with E. Coli. in Figure 6.10.

Recall the model of a pollen grain moving to the right or left on a discrete lattice as shown in Figure 6.11

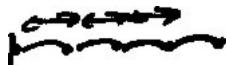


Figure 6.11: Pollen grain

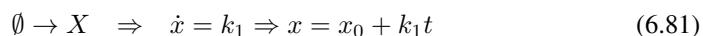
Let's discretize time as well. The model can be written

$$n_{t+\tau} = n_t + \eta_t \quad (6.78)$$

$$\eta_t^2 = 1 \quad (6.79)$$

$$p(\eta = 1) = p(\eta = -1) = 1/2 \quad (6.80)$$

Now we similarly consider the stochastic description of chemical kinetics. First there is nothing and then there is a protein.



We want the number of molecules as a function of time. At  $t = 0$  there are no proteins. Now what is the probability that there are  $n$  copies of the protein after time interval  $T$ ? Consider the case wherein the number of proteins increases is small (with probability  $\epsilon$ ).

$$p(\eta = 1) = \epsilon \quad (6.82)$$

$$P_n^T = \binom{T}{n} \epsilon^n (1-\epsilon)^{T-n} \quad (6.83)$$

When  $n = 1/2$  this distribution became a Gaussian. Recall,

$$n_{t+\tau} = n_t + \eta_t \quad (6.84)$$

$$\downarrow \quad (6.85)$$

$$p_{t+\tau}(n) = \sum_{n'} p(n|n') p_t(n') \quad (6.86)$$

$$\begin{array}{ccc} \swarrow & & \searrow \\ \dot{p} = Dp'' & \rightarrow & \mathcal{N} \end{array} \quad (6.87)$$

$$\quad (6.88)$$

We can also consider an unfair coin probability as well.

To form an update equation we use

$$n_{t+\tau} = n_t + \eta_t \quad (6.89)$$

$$p(\eta_t) \quad (6.90)$$

Then,

$$p_{t+\tau}(n) = \sum_{n'} p(n|n') = \sum_{n'} \sum_{\eta} p(n, n'|\eta) \quad (6.91)$$

$$= \sum_{n'} p(n|n') p_t(n') \quad (6.92)$$

$$= \sum_{n'} \sum_{\eta} p(n|n', \eta) p_t(n') p(\eta) \quad (6.93)$$

and

$$\delta_{n,n'+\eta} = p(n|n', \eta) \quad (6.94)$$

We get the obvious equations:

$$p_{t+\tau}(n) = \sum_{n'} \sum_{\eta=0,1} (\delta_{n,n'+\eta}) p_t(n') p(\eta) \quad (6.95)$$

$$= \sum_{n'} \delta_{n,n'+1} p_t(n') \epsilon + \delta_{n,n'} p_t(n') (1 - \epsilon) \quad (6.96)$$

$$p_{t+\tau}(n) = p_t(n-1)\epsilon + p_t(n)(1-\epsilon) \quad (6.97)$$

$$p_{t+\tau}(n) - p_t(n) = \epsilon [p_t(n-1) - p_t(n)] \quad (6.98)$$

If we assume small time and space steps we get the PDE

$$\tau \dot{p} = -\epsilon p' \quad (6.99)$$

By the method of characteristics to solve the previous partial differential equation, all solutions are of the form

$$p = f \left( t - \frac{\epsilon}{\tau} n \right). \quad (6.100)$$



Figure 6.12: Solution by characteristics

The initial condition is merely translated as shown in Figure 6.12.

We can prove that this is in fact a solution with the following computation.

$$\tau \dot{p} + \epsilon p' = 0 \quad (6.101)$$

$$p = f(n - ct) = f(z) \quad (6.102)$$

$$\dot{p} = \frac{df}{dz} \dot{z} = -c \frac{df}{dz} \quad (6.103)$$

$$p' = \frac{df}{dz} z' = \frac{df}{dz} \quad (6.104)$$

Now plugging this into the PDE

$$-\tau c \frac{df}{dz} + \epsilon \frac{df}{dz} \Rightarrow c = \frac{\epsilon}{\tau} \quad (6.105)$$

In the macroscopic model

$$\emptyset \xrightarrow{k_1} X \quad (6.106)$$

$$\dot{x} = k_1 \Rightarrow x = x_0 + k_1 t \quad (6.107)$$

$$\epsilon = k_1 V \tau \quad (6.108)$$

$$c = \epsilon / \tau = k_1 V \quad (6.109)$$

We see that the probability distribution just shifts as in Figure 6.13.

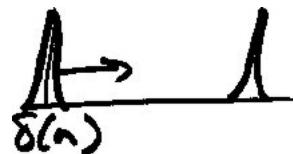


Figure 6.13: Shifted delta distribution

What if the dynamics are not that smooth and the approximation leading to the PDE above is not a good assumption?

$$\tau \dot{p}_n = -\epsilon p_n + \epsilon p_{n-1} \quad (6.110)$$

$$\frac{\tau}{\epsilon} \dot{p}_n = -p_n + p_{n-1}, \quad n > 0 \quad (6.111)$$

$$\frac{\tau}{\epsilon} \dot{p}_0 = -p_0 \quad (6.112)$$

We can solve the last equation:

$$p_0 = C e^{-t\epsilon/\tau} = e^{-t\epsilon/\tau} \quad (6.113)$$

Consider the example

$$\emptyset \xrightarrow{k_1} X \quad (6.114)$$

So at  $t = 0$  there are no particles:

$$p_0 = 1 \quad (6.115)$$

$$p_j = 0, \quad \forall j > 0 \quad (6.116)$$

To solve this we introduce a new variable

$$Q_j = p_j e^{t\epsilon/\tau}. \quad (6.117)$$

Using

$$\left(\frac{\tau}{\epsilon}\right) \dot{p}_n + p_n = p_{n-1} \quad (6.118)$$

we get

$$\dot{Q}_j = \dot{p}_j e^{t\epsilon/\tau} + p_j \frac{\epsilon}{\tau} e^{t\epsilon/\tau} \quad (6.119)$$

$$= e^{t\epsilon/\tau} \left( \dot{p}_j + p_j \frac{\epsilon}{\tau} \right) \quad (6.120)$$

$$= \frac{\epsilon}{\tau} e^{t\epsilon/\tau} \left( \frac{\tau}{\epsilon} \dot{p}_j + p_j \right) \quad (6.121)$$

$$= \frac{\epsilon}{\tau} e^{t\epsilon/\tau} p_{j-1} \quad (6.122)$$

$$\dot{Q}_j = \frac{\epsilon}{\tau} Q_{j-1} \quad (6.123)$$

$$Q_j = p_j e^{t\epsilon/\tau} \quad (6.124)$$

$$Q_0 = p_0 e^{t\epsilon/\tau} \Rightarrow Q_0(t=0) = 1 \quad (6.125)$$

$$(6.126)$$

So

$$\dot{Q}_0 = 0 \Rightarrow Q_0 = 1 \quad (6.127)$$

$$\dot{Q}_j = \frac{\epsilon}{\tau} e^{t\epsilon/\tau} \left( \frac{\tau}{\epsilon} \dot{p}_j + p_j \right). \quad (6.128)$$

Since

$$\frac{\tau}{\epsilon} \dot{p}_0 = -p_0 \quad (6.129)$$

we have

$$\dot{Q}_1 = \frac{\epsilon}{\tau} Q_0 = \frac{\epsilon}{\tau} \Rightarrow Q_1 = \frac{\epsilon t}{\tau} + \underbrace{C}_0 \quad (6.130)$$

$$\dot{Q}_2 = \frac{\epsilon}{\tau} Q_1 = \left( \frac{\epsilon}{\tau} \right)^2 t \Rightarrow Q_2 = \left( \frac{\epsilon}{\tau} \right)^2 \cdot \frac{1}{2} t^2 \quad (6.131)$$

The result is

$$\dot{Q}_j = \frac{\epsilon}{\tau} Q_{j-1} \quad (6.132)$$

$$Q_0 = 1, Q_1 = \left( \frac{\epsilon}{\tau} \right) t, Q_2 = \left( \frac{\epsilon}{\tau} \right)^2 \cdot \frac{1}{2} t^2 \quad (6.133)$$

$$Q_j = \left( \frac{\epsilon}{\tau} \right)^j \cdot \frac{1}{j!} t^j \quad (6.134)$$

$$p_j = \frac{1}{j!} t^j e^{-\epsilon t/\tau} \left( \frac{\epsilon}{\tau} \right)^j \quad (6.135)$$

$$\lambda = \epsilon t / \tau \quad (6.136)$$

$$p_j = \frac{1}{j!} \lambda^j e^{-\lambda} \quad (6.137)$$

This distribution is called a Poisson distribution with Poisson parameter  $\lambda$  and is shown in Figure 6.14.

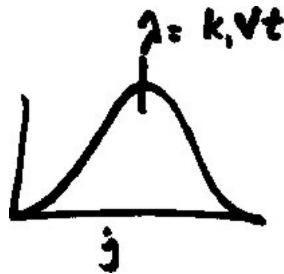


Figure 6.14: Poisson distribution

### 6.6.3 HW

1. Show

$$\binom{T}{n} \epsilon^n (1-\epsilon)^{T-n} \rightarrow \text{Poisson distribution} \quad (6.138)$$

Hint: Use Sterling's approximation and other tricks:

$$T! \approx T^T e^{-T} \sqrt{2\pi T} \quad (6.139)$$

$$(T-n)! \approx (T-n)^{T-n} e^{-(T-n)} \sqrt{2\pi(T-n)} \quad (6.140)$$

$$\left(1 - \frac{x}{T}\right)^T \rightarrow e^{-x} \quad (6.141)$$

2. Prove

$$\dot{p}_j = -rp_j + rp_{j-1} \quad (6.142)$$

is solved by a Poisson distribution.

3. Find  $\langle j \rangle$  for  $p_j = \frac{1}{j!} \lambda^j e^{-\lambda}$

## 6.7 Lecture 15

### 6.7.1 Transcriptional regulatory networks

The dynamics are intrinsically noisy due to the finiteness of the constituents. Recall the setup in Figure 6.15.



Figure 6.15: Transcriptional regulation

$$\underbrace{\partial_t[x]}_0 \rightarrow \partial_t p_n \quad (6.143)$$

We write

$$p_{t+\tau}(n) = \sum_{n'} \sum_{\eta} p(n, n', \eta) \quad (6.144)$$

$$n_{t+\tau} = n + \eta_t \Rightarrow p(n|n', \eta) = \delta_{n,n'+\eta} \quad (6.145)$$

$$p(\eta_t) \quad (6.146)$$

This is always true of a marginal with a joint distribution. For example,

$$p(a) = \sum_b \sum_c p(a, b, c) = \sum_b \sum_c p(b|a, c)p(c|a)p(a) \quad (6.147)$$

To derive a stochastic description write

$$p_{t+\tau}(n) = \sum_{n'} \sum_{\eta} p(n, n', \eta) p(\eta|n') p(n') \quad (6.148)$$

Recall some general probability.

$$1 = \sum_a p(a|b, c, d, \dots) \quad (6.149)$$

$$= \sum_a \frac{p(a, b, c, d)}{p(b, c, d)} \cdots \quad (6.150)$$

$$= \frac{p(b, c, d)}{p(b, c, d)} \cdots = 1 \quad (6.151)$$

$$(6.152)$$

So

$$p(a) = \sum_{b,c} p(b|a, c)p(c|a)p(a) \quad (6.153)$$

$$= \sum_c p(c|a)p(a) \quad (6.154)$$

$$= p(a) \quad (6.155)$$

In our case we have

$$p_{t+\tau} = \sum_{n'} \sum_{\eta} \delta_{n, n' + \eta} p(\eta|n') p_t(n') \quad (6.156)$$

$$= \sum_{\eta} p(\eta|n - \eta) p_t(n - \eta) \quad (6.157)$$

We want to turn this into ODEs.

$$p_{t+\tau} - p_t(n) = \sum_{\eta} p(\eta|n - \eta) p_t(n - \eta) - p_t(n) \quad (6.158)$$

$$\tau \frac{d}{dt} p(n) = \left[ \lim_{\tau \rightarrow 0} \sum_{\eta} p(\eta|n - \eta) p_t(n - \eta) \right] - p_t(n). \quad (6.159)$$

So for proteins we want

$$X \Rightarrow \emptyset \quad (6.160)$$

but we have only discussed

$$\emptyset \xrightarrow{k_1} X \quad (6.161)$$

$$\dot{x} = k_1 \quad (6.162)$$

$$p(\eta = 0)|n - \eta) = 1 - \epsilon \quad (6.163)$$

$$p(\eta = 1)|n - \eta) = \epsilon = k_1 V \tau. \quad (6.164)$$

So to derive the ODE

$$\tau \frac{d}{dt} p(n) = \sum_{\eta=0,1} p(\eta|n - \eta) p_t(n - \eta) - p_t(n) \quad (6.165)$$

$$= (1 - \epsilon)p_t(n) + \epsilon p_t(n - 1) - p_t(n) \quad (6.166)$$

$$= \epsilon [p_t(n - 1) - p_t(n)] \quad (6.167)$$

If we assume  $p$  is smooth in  $n$  we get a PDE

$$\tau \dot{p} + \epsilon p' = 0 \quad (6.168)$$

which had solutions of the form

$$p = f \left( n - \frac{\epsilon}{\tau} t \right) \quad (6.169)$$

which just translates with  $\epsilon/\tau = k_1 V$  as in Figure 6.16.



Figure 6.16: Solution of probability PDE

Alternatively,

$$\tau \dot{p}_n = \epsilon [p_t(n - 1) - p_t(n)] \quad (6.170)$$

and fixing up notation gives

$$\tau \dot{p}_n = \epsilon (p_{n-1} - p_n). \quad (6.171)$$

Now introduce a nondimensional time

$$s = \frac{t\epsilon}{\tau} \quad (6.172)$$

so that the equation is

$$\frac{dp_n}{ds} = p_{n-1} - p_n. \quad (6.173)$$

For the homework you will show

$$p_n = \frac{s^n}{n!} e^{-s} \text{ "Poisson"} \quad (6.174)$$

$$Q_n = p_n e^s \quad (6.175)$$

$$\frac{dQ_n}{ds} = Q_{n-1} \quad (6.176)$$

$$\frac{dQ_0}{ds} = 0. \quad (6.177)$$

Hint:

$$p_n = \frac{s^n}{n!} e^{-s} \quad (6.178)$$

$$\langle n^j \rangle = \text{"jth moment of } n\text{"} \quad (6.179)$$

Define moment generating function to get the moments. Given

$$\sum_{n=0}^{\infty} p_n \quad (6.180)$$

You can always define

$$\mathcal{Z}(\lambda) = \sum_{n=0}^{\infty} p_n e^{\lambda n} \quad (6.181)$$

$$\partial_{\lambda}^j \mathcal{Z} \Big|_{\lambda=0} = \sum_{n=0}^{\infty} n^j p_n = \langle n^j \rangle \quad (6.182)$$

Can be more general

$$\mathcal{Z}(\lambda) = \sum_{n=-\infty}^{\infty} p_n e^{\lambda n} \quad (6.183)$$

where  $\lambda$  may need to be complex.

$$p_n = \frac{s^n}{n!} e^{-s} \quad (6.184)$$

gives

$$\langle n^j \rangle = \sum_{n=0}^{\infty} \frac{n^j s^n e^{-s}}{n!} \quad (6.185)$$

$$= e^{-s} \sum_{n=0}^{\infty} \frac{n^j s^n}{n!} \quad (6.186)$$

$$= e^{-s} (s \partial_s)^j \sum_{n=0}^{\infty} \frac{s^n}{n!} \quad (6.187)$$

$$= e^{-s} (s \partial_s)^j e^s. \quad (6.188)$$

For the Poisson distribution let's calculate

$$\langle n^1 \rangle = e^{-s} s \partial_s e^s \quad (6.189)$$

$$= e^{-s} s e^s \quad (6.190)$$

$$= s \quad (6.191)$$

$$\langle n^2 \rangle = e^{-s} s \partial_s s \partial_s e^s \quad (6.192)$$

$$= e^{-s} s \partial_s e^s \quad (6.193)$$

$$= e^{-s} s (e^s + s e^s) \quad (6.194)$$

$$= s(1+s) \quad (6.195)$$

$$= s + s^2 \quad (6.196)$$

Now we compute

$$\sigma_n^2 = \langle n^2 \rangle - \langle n \rangle^2 \quad (6.197)$$

$$= s + s^2 - s^2 \quad (6.198)$$

$$= s. \quad (6.199)$$

So that

$$\frac{\sigma_n}{\langle n \rangle} = \frac{s^{1/2}}{s} \quad (6.200)$$

$$= s^{-1/2} \quad (6.201)$$

$$= \frac{1}{\sqrt{\langle n \rangle}}. \quad (6.202)$$

Recall we wanted to understand the reaction



$s$  is proportional to time

$$s = \frac{t\epsilon}{\tau} = tk_1 V. \quad (6.204)$$

This reaction is given by the ODE

$$\dot{x} = k_1 \quad (6.205)$$

with solution

$$x = k_t \quad (6.206)$$

$$= \frac{\langle n \rangle}{V}, \quad \text{since } \langle n \rangle = tk_1 V \quad (6.207)$$

Consider the converse reaction



The equation is

$$\dot{x} = -k_{-1}x \quad (6.209)$$

$$n_{t+\tau} = n_t + \eta_t \quad (6.210)$$

$$p(\eta = -1|n') = \tau k_{-1} n' \quad (6.211)$$

Recall from the earlier part of class Figure 6.17 where the equations was

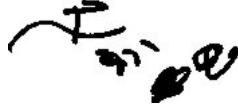


Figure 6.17: transcription

$$\dot{x} = \alpha(x) - rx. \quad (6.212)$$

So things are qualitatively different between decay and creation as we can see in both the micro and macro descriptions. There is one way for creation and  $n$  for annihilation.

$$p_n(t + \tau) = \sum_n \sum_{\eta} \underbrace{p(n|n', \eta)}_{\delta_{n,n'+\eta} \equiv \delta_{n',n-\eta}} p(\eta|n') p_{n'}(t) \quad (6.213)$$

$$= \sum_{\eta} \sum_{n'} \delta_{n',n-\eta} p(\eta|n') p_{n'}(t) \quad (6.214)$$

$$= \sum_{n'} \delta_{n',n+1} \tau k_{-1} n' p_{n'}(t) \quad (6.215)$$

$$+ \sum_{n'} \delta_{n',n} (1 - \tau k_{-1} n') p_{n'}(t) \quad (6.216)$$

$$= \tau k_{-1} [(n+1)p_{n+1} - np_n] + p_n(t) \quad (6.217)$$

So we get

$$p_n(t + \tau) - p_n(t) = \tau k_{-1} [(n+1)p_{n+1} - np_n] + p_n(t) \quad (6.218)$$

$$\tau \frac{dp}{dt} = \quad (6.219)$$

$$(x \xrightarrow{k_{-1}} \emptyset) : \dot{p}_n = k_{-1} [(n+1)p_{n+1} - np_n] + p_n \approx \partial_n(np) \quad (6.220)$$

$$(\emptyset \xrightarrow{k_1} x) : \dot{p}_n = k_1(p_{n-1} - p_n) \approx -\partial_n p \quad (6.221)$$

$$\frac{d}{dt} \langle n \rangle = \sum n \frac{dp_n}{t} \quad (6.222)$$

$$= k_{-1} \sum_{n=0}^{\infty} n [(n+1)p_{n+1} - np_n] \quad (6.223)$$

$$= k_{-1} \left[ \sum_{m=1}^{\infty} (m-1)mp_m - \sum_{n=0}^{\infty} n^2 p_n \right] \quad (6.224)$$

$$= k_{-1} \sum_{m=0}^{\infty} m^2 p_m - mp_m - m^2 p_m \quad (6.225)$$

We get

$$\frac{d}{dt} \langle n \rangle = k_{-1} \sum_{m=0}^{\infty} (-mp_m) \quad (6.226)$$

$$= -k_{-1} \langle n \rangle \quad (6.227)$$

and using

$$x = \langle n \rangle / V \quad (6.228)$$

$$\langle n \rangle = xV \quad (6.229)$$

we get

$$V \frac{d}{dt} x = -k_{-1} V x \quad (6.230)$$

$$\dot{x} = -k_{-1} x \quad (6.231)$$

(\* Dimensional analysis to find the extra  $V$  in the lecture

$$[k_1] \sim \frac{1}{t} \frac{1}{V} \quad (6.232)$$

$$[k_{-1}] \sim \frac{1}{t} \quad (6.233)$$

\*) These results have been written as a discrete form of a conservation law. To see this more clearly, recall the rule about conditional distributions and its consequences.

$$\sum_a p(a|b) = 1 \quad (6.234)$$

$$p_n(t + \tau) = \sum_n \sum_{\eta} p(n|n', \eta) p(\eta|n') p_{n'}(t) \quad (6.235)$$

We want

$$\sum_n p_n(t + \tau) = \sum_n p_n(t) \quad (6.236)$$

so we need

$$\sum_n p(n|n', \eta) = 1 \quad (6.237)$$

$$\sum_n p_n(t + \tau) = \sum_n \sum_{\eta} p(\eta|n') p_{n'}(t) \quad (6.238)$$

$$\sum_n p(\eta|n') = 1 \quad (6.239)$$

$$\sum_n p_n(t + \tau) = \sum_n p_{n'}(t) \quad (6.240)$$

- HW 1: Show dynamics in equations (6.220) preserve normalization. I.e. show  $\sum_n \dot{p}_n = 0$  for  $\emptyset \rightarrow x$  and  $x \rightarrow \emptyset$ .
- HW 2: Show

$$\frac{d}{d\tau} \langle n \rangle = \begin{cases} -k_{-1} \langle n \rangle, & x \rightarrow \emptyset, \\ k_1 V, & \emptyset \rightarrow x. \end{cases} \quad (6.241)$$

Let's put both dynamics together:

$$\emptyset \rightleftharpoons x \quad (6.242)$$

which yields the differential equation

$$\dot{x} = \alpha - rx. \quad (6.243)$$

Let's look for a fixed point.

$$\dot{x} = f(x) \quad (6.244)$$

$$f(x_*) = 0 \quad (6.245)$$

$$x_* = \alpha/r \quad (6.246)$$

At the microscopic level,

$$p_n(t + \tau) = \sum_{\eta} \sum_n p(\eta|n') p(n|n', \eta) p_{n'}(t) \quad (6.247)$$

- HW 3: Show

$$\dot{p}_n = -(\alpha p_n + \alpha p_{n-1})V - rnp_n + r(n+1)p_{n+1} \quad (6.248)$$

To do this you compute  $p(\eta = -1, 0, 1)$ . In the macroscopic system:

$$\dot{x} = f(x) = \alpha - rx \quad (6.249)$$

$$x_* = \alpha/r \quad (6.250)$$

To determine stability compute

$$\frac{d}{dx} (\alpha - rx)|_{x_*} = -r \Rightarrow \text{stable} \quad (6.251)$$

- HW 4: See if all eigenvalues are negative, i.e.  $-r = f'|_{x_*} = \lambda_r < 0$ ?
- HW 5: Read and enjoy the papers.

## 6.8 Lecture 16

### 6.8.1 Review of stochastic description of chemical kinetics

Master equations for creation and destruction

$$\begin{array}{lll} \text{micro scale} & & \text{macro scale} \\ x \xrightarrow{r} \emptyset : \quad \dot{p}_n = -rnp_n + r(n+1)p_{n+1} : & \dot{x} = -rx & (6.252) \end{array}$$

$$\emptyset \xrightarrow{\alpha} x : \quad \dot{p}_n = (-\alpha p_n + \alpha p_{n-1}) \underbrace{V}_{\text{volume}} : \quad \dot{x} = \alpha & (6.253)$$

$$(6.254)$$

$$x = \langle n \rangle \quad (6.255)$$

$$= \sum_{n=0}^{\infty} np_n \quad (6.256)$$

Possion distribution:

$$p_n(t) = \frac{1}{n!} (\alpha V t)^n e^{-\alpha V t} \quad (6.257)$$

$$= P \quad (6.258)$$

$$\langle n \rangle = \lambda = \sigma^2 \quad (6.259)$$

$$\frac{\sigma}{\langle n \rangle} = \frac{1}{\sqrt{\langle n \rangle}} \quad (6.260)$$

Consider both reactions



$$\dot{p}_n = -rnp_n + r(n+1)p_{n+1} - (\alpha p_n V) + (\alpha p_{n-1} V) \quad (6.262)$$

$$\dot{x} = -rx + \alpha = f(x) \quad (6.263)$$

$$f' = -r < 0 \quad (6.264)$$

$$x_* = \alpha/r \quad (6.265)$$

The method for determining stability is

$$\dot{x} = f(x) \quad (6.266)$$

$$f(x_*) = 0 \quad (6.267)$$

$$f'(x_*) \quad \left\{ \begin{array}{ll} < 0 : \text{stable} & > 0 : \text{unstable} \end{array} \right. \quad (6.268)$$

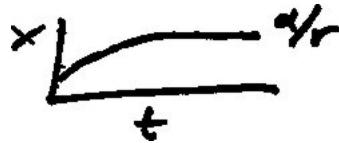


Figure 6.18: Stability of macroscopic system

So we compute

$$f' = -r < 0 \quad (6.269)$$

and find the fixed point is stable.

Now for the microscopic description.

$$0 = -rnp_n + r(n+1)p_{n+1} - \alpha V p_n + \alpha V p_{n-1} \quad (6.270)$$

$$(6.271)$$

Let's divide by  $r$  and define

$$\lambda = \frac{\alpha V}{r}. \quad (6.272)$$

What are its dimensions? Return to the equation

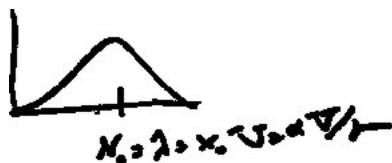
$$\dot{x} = -rx + \alpha \quad (6.273)$$

$$x_* = \alpha/r \quad (6.274)$$

So  $\lambda$  is nondimensional. We can think of it as

$$\mathcal{N}_* = \lambda = x_* V \quad (6.275)$$

as in Figure 6.19

Figure 6.19:  $\lambda$  plot

Then we have:

$$0 = -np_n + (n+1)p_{n+1} - \lambda p_n + \lambda p_{n-1} \quad (6.276)$$

$$p_n = \frac{x^n}{n!} c_n, \quad (\text{ansatz}) \quad (6.277)$$

$$\begin{aligned} 0 &= -n \frac{x^n}{n!} c_n + (n+1) \frac{x^{n+1}}{(n+1)!} c_{n+1} - \lambda \frac{x^n}{n!} c_n \\ &\quad + \lambda \frac{x^{n-1}}{(n-1)!} c_{n-1} \end{aligned} \quad (6.278)$$

$$= -\frac{x^n c_n}{(n-1)!} + \frac{x^{n+1} c_{n+1}}{n!} - \frac{\lambda x^n c_n}{n!} + \frac{\lambda x^{n-1} c_{n-1}}{(n-1)!} \quad (6.279)$$

$$= \frac{1}{(n-1)!} [-\lambda^n + \lambda \cdot \lambda^{n-1}] + \frac{1}{n!} [\lambda^{n+1} - \lambda \cdot \lambda^n] \quad (6.280)$$

where we have set  $x = \lambda$  and  $c_n = 1$ .

- HW 1: Show the Poisson distribution  $P(\lambda)$  solves  $\emptyset \rightleftharpoons x$ .



Figure 6.20: Poisson distribution

$$\langle n \rangle = \sigma^2 = \lambda = \mathcal{N}_* = x_* V = \alpha V / r \quad (6.281)$$

$$\frac{\sigma}{\langle n \rangle} = \frac{1}{\sqrt{\langle n \rangle}} \quad (6.282)$$

We could also include regulation.

$$\dot{x} = \alpha(x) - rx \quad (6.283)$$

Take  $\alpha(n)$ . Some examples are in as in Figure 6.21.

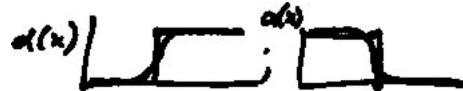


Figure 6.21:  $\alpha(x)$

If  $\alpha$  is a function of number as

$$\alpha(n) = \begin{cases} \alpha_+ & n \geq n_* = KV \\ \alpha_- & n < n_* \end{cases} \quad (6.284)$$

Recall the continuous case.

$$\dot{x} = -rx + (\alpha_+ - \alpha_-)H(x - x_*) + \alpha_- \quad (6.285)$$

$$= -rx + \alpha(x) \quad (6.286)$$

where  $x_*$  is the transition value.

We have a fixed point  $x = \alpha/r$ . There are cases.

$$x < x_* = \alpha_-/r \quad (6.287)$$



Figure 6.22: Fixed point diagram

Also, there is

$$x = \alpha_+/r > x_* = KV \quad (6.288)$$

$$x \rightarrow \alpha/r \quad (6.289)$$



Figure 6.23: Distribution

In terms of the micro system

$$0 = -rnp_n + r(n+1)p_{n+1} - \alpha(n)Vp_n + \alpha(n-1)Vp_{n-1} \quad (6.290)$$

The analogues are

$$n-1 \geq n_* \quad \lambda = \alpha_+V/r \quad (6.291)$$

$$n+1 < n_* \quad \lambda = \alpha_-V/r \quad (6.292)$$

$$(6.293)$$

- HW 2: Show

$$p_n = \begin{cases} c_- \lambda_-^n / n! & n \leq n_* \\ c_+ \lambda_+^n / n! & n > n_* \end{cases} \quad (6.294)$$

Solve for  $c_+/c_-$ . We use the constraint

$$1 = \sum_{n=0}^{n_*} c_- \frac{\lambda_-^n}{n!} + \sum_{n=0}^{n_*} c_+ \frac{\lambda_+^n}{n!} \quad (6.295)$$

Figure 6.24 shows some fixed point plots.



Figure 6.24: Fixed point cases

MATLAB discussion.

Student Question on dynamical systems

$$\dot{x} = f(x) = \dot{\eta} \quad (6.296)$$

$$f(x_*) = 0 \quad (6.297)$$

$$f(x_* + \eta) \approx (x_*) + \eta f'(x_*) + \frac{1}{2} \eta^2 f''(x_*) \quad (6.298)$$

$$\dot{\eta} = \frac{1}{2} \eta^2 f''(x_*) \quad (6.299)$$

$$\frac{d\eta}{\eta^2} = \frac{1}{2} f''(x_*) dt \quad (6.300)$$

$$-\eta^{-1} = \frac{1}{2} f'' t + C \quad (6.301)$$

$$\eta = \frac{-1}{t^{\frac{1}{2}} f'' + C} \quad (6.302)$$

See Figure 6.25.

For nonlinear systems stability is more complicated. For instance, you may find a nondiagonalizable Jacobian

$$J = \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \quad (6.303)$$

Back to MATLAB comparing deterministic and stochastic dynamics. If things are not regulating each other

$$x \xrightarrow{\frac{r}{\alpha}} \emptyset \quad (6.304)$$

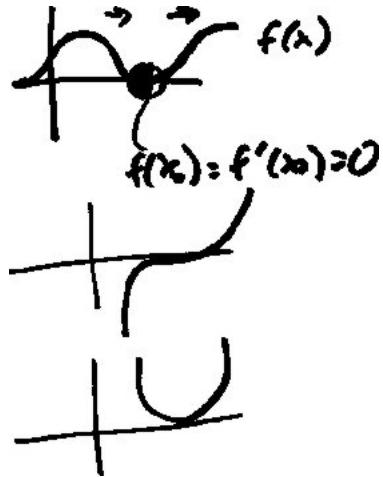
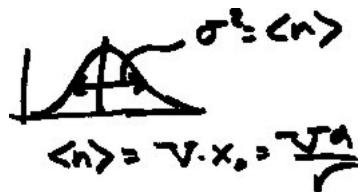


Figure 6.25: Fixed point

Figure 6.26: Distribution with  $\langle n \rangle = Vx_* = V\alpha/r$ Figure 6.27:  $\alpha(x)$ 

which we see in Figure 6.26.

We have choices for  $\alpha(x)$ . See Figure 6.27.

What if you have an arbitrary regulation function? We replace differences by derivatives. When we discussed pollen grains we had

$$(\delta t)\dot{p}_n = \frac{1}{2}p_{n-1} + \frac{1}{2}p_{n+1} \approx (\delta x)^2 p'' \quad (6.305)$$

For  $\emptyset \rightarrow x$ ,

$$\dot{p}_n = V\alpha(-p_n + p_{n-1}) \quad (6.306)$$

$$\dot{p}_n = -V\alpha \partial_n p_n \quad (6.307)$$

$$p_n = f(n - V\alpha t) \quad (6.308)$$

We see the propagating solution in Figure 6.28.

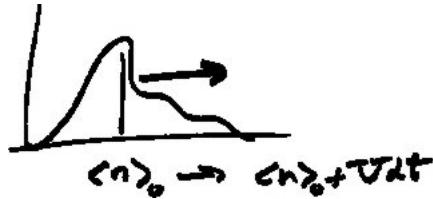


Figure 6.28: Propagating average

More carefully, we write

$$q(x+y) = \sum_{j=0}^{\infty} y^j \frac{1}{j!} \partial_x^j q|_x = e^{y \cdot \partial_x} q(x) \quad (6.309)$$

$$p_{n+1} = e^{\partial_n} p_n \quad (6.310)$$

$$P(x) : P(nV) \propto p_n \quad (6.311)$$

$$P(nV + V) = e^{\frac{1}{V} \partial_x} P(nV) \quad (6.312)$$

since

$$\partial_n = \frac{dx}{dn} \partial_x \quad (6.313)$$

$$n = xV \quad (6.314)$$

$$\left( \frac{dn}{dx} \right)^{-1} = V^{-1} \quad (6.315)$$

Take

$$\dot{p}_n = V\alpha(-p_n + p_{n-1}) \quad (6.316)$$

$$\dot{p}_n = V\alpha(-1 + e^{-\partial_n})p_n \quad (6.317)$$

$$\dot{P}(x) = V\alpha(-1 + e^{-\frac{1}{V} \partial_x})P(x) \quad (6.318)$$

$$= V\alpha(-1 + 1 - \frac{1}{V} \partial_x)P(x) + O(1/V^2) \quad (6.319)$$

$$= -\alpha \partial_x P(x) \quad (6.320)$$

This has a solution of the form

$$f(x - \alpha t) \quad (6.321)$$

solving

$$(\partial_t + \alpha \partial_x)P = 0 \quad (6.322)$$

Aside: The two way wave equation can be written

$$(\partial_t + \alpha \partial_x)(\partial_t - \alpha \partial_x)P = 0 \quad (6.323)$$

So we get

$$\dot{p}_n = -rnp_n + r(n+1)p_{n+1} \quad (6.324)$$

$$= -rnp_n + re^{0n}np_n \quad (6.325)$$

$$\dot{P} = -rxVP + re^{\frac{1}{V}\partial_x}xVP \quad (6.326)$$

$$= -rxVP + r\left(1 + \frac{1}{V}\partial_x\right)xVP + O(1/V^2) \quad (6.327)$$

$$\dot{P} = r\partial_x xP \quad (6.328)$$

$$\dot{P} = r\partial_x xP - \alpha \partial_x P \quad (6.329)$$

$$0 = rP + rxP' - \alpha P' \quad (6.330)$$

We reach a problem

$$P' = \frac{rP}{\alpha - rx} = \frac{P}{\alpha/r - x} = \frac{P}{x_* - x} \quad (6.331)$$

The derivative diverges when  $x = x_*$ . We made an incorrect assumption

$$\frac{1}{V}\partial_x P \ll 1. \quad (6.332)$$

Being more careful

$$\dot{p}_n = -rnp_n + r(n+1)p_{n+1} \quad (6.333)$$

$$= r(-1 + e^{\partial_n})nP \quad (6.334)$$

$$\dot{P} = r\left(-1 + e^{\frac{1}{V}\partial_x}\right)xVP \quad (6.335)$$

$$= r\left(-1 + 1 + \frac{1}{V}\partial_x + \frac{1}{2}V^{-2}\partial_x^2\right)xVP \quad (6.336)$$

$$= \partial_x \left[ \frac{r}{V} + \frac{1}{2}rV^{-2}\partial_x \right] VxP \quad (6.337)$$

$$= \partial_x \left[ r + \frac{1}{2}rV^{-1}\partial_x \right] xP \quad (6.338)$$

$$(6.339)$$

So

$$\dot{p}_n = -\alpha p_n + \alpha p_{n-1} \quad (6.340)$$

$$= \alpha(-1 + e^{-\partial_n})p_n \quad (6.341)$$

$$\dot{P} = \alpha \left( -1 + e^{-\frac{1}{V}\partial_x} \right) P \quad (6.342)$$

$$\approx \alpha \left( -1 + 1 - \frac{1}{V}\partial_x + \frac{1}{2}\frac{1}{V^2}\partial_x^2 \right) P \quad (6.343)$$

$$= \partial_x \left( -\frac{\alpha}{V} + \frac{1}{2}\frac{\alpha}{V^2}\partial_x \right) P \quad (6.344)$$

$$= \partial_x \left[ -\frac{\alpha}{V} + \frac{1}{2}\frac{\alpha}{V^2}\partial_x + rx + \frac{1}{2}rV^{-1}\partial_x \right] P \quad (6.345)$$

$$\dot{P} = -\partial_x J \quad (6.346)$$

Now in the case  $J = 0$

$$\left( -\frac{\alpha}{V} + rx + \frac{1}{2}\frac{r}{V} \right) P + \left( \frac{1}{2}\frac{\alpha}{V^2} + \frac{1}{2}\frac{r}{V} \right) P' = 0 \quad (6.347)$$

(mistake above)

- HW: Derive correct expression for  $J$  in terms of  $P$  and  $P'$ .

The result is:

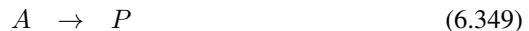
$$P' = \frac{(x - x_* + \frac{1}{2}\frac{r}{V}) P}{\frac{1}{2}(\alpha + \frac{r}{V})} = \frac{vP}{D} \quad (6.348)$$

This looks like the diffusion-drift equations from the earlier part of the class.  
See Figures 6.29 and 6.30.

## 6.9 Lecture 17

### 6.9.1 Chemical kinetics review

Changing chemical kinetics into a differential equation. The chemical reactions



lead to differential equations

$$\dot{a} = -k_1 a - k_3 ab \quad (6.352)$$

$$\dot{b} = k_2 - k_3 ab \quad (6.353)$$

$$\dot{c} = k_3 ab \quad (6.354)$$

$$\dot{p} = k_3 ab. \quad (6.355)$$

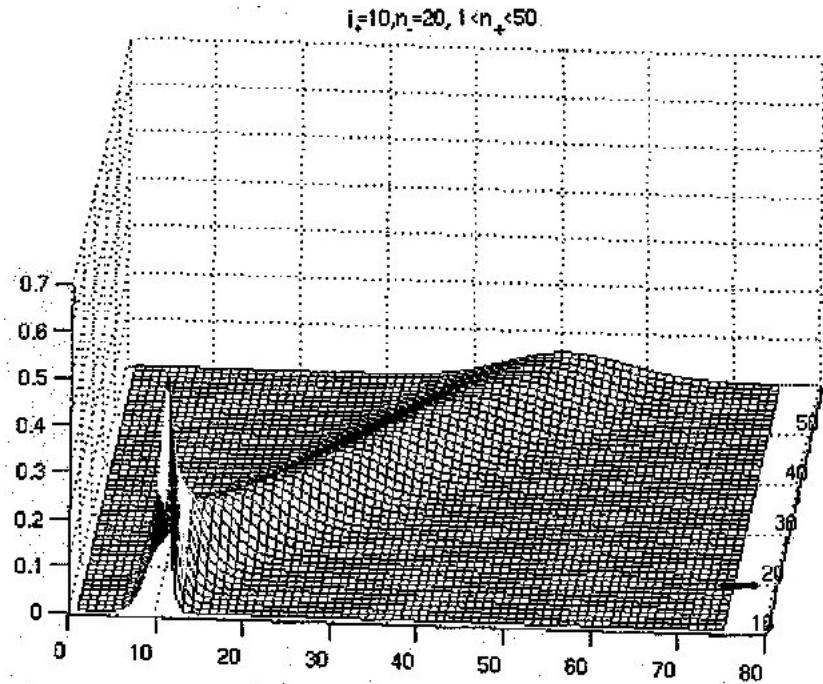


Figure 6.29: MATLAB

Also, the equation

$$A \rightleftharpoons \emptyset \quad (6.356)$$

yields the macroscopic equation and the microscopic master equation:

$$\dot{a} = -k_4 a + k_{-4} \quad (6.357)$$

$$\dot{p}_n = -k_4 n p_n + k_4(n+1)p_{n+1} + k_{-4}Vp_{n-1} - k_{-4}Vp_n \quad (6.358)$$

Similarly

$$\emptyset \rightarrow A \quad (6.359)$$

leads to

$$\dot{a} = k_5. \quad (6.360)$$

Answering a student question we write

$$\dot{p}_0 = -k_5 p_0 V \quad (6.361)$$

$$\dot{p}_j = -k_5 p_j V + k_5 p_{j-1} V, \quad j \geq 1 \quad (6.362)$$

$$1 = \sum_{j=0}^{\infty} p_j = p_0 + p_1 + \dots \quad (6.363)$$

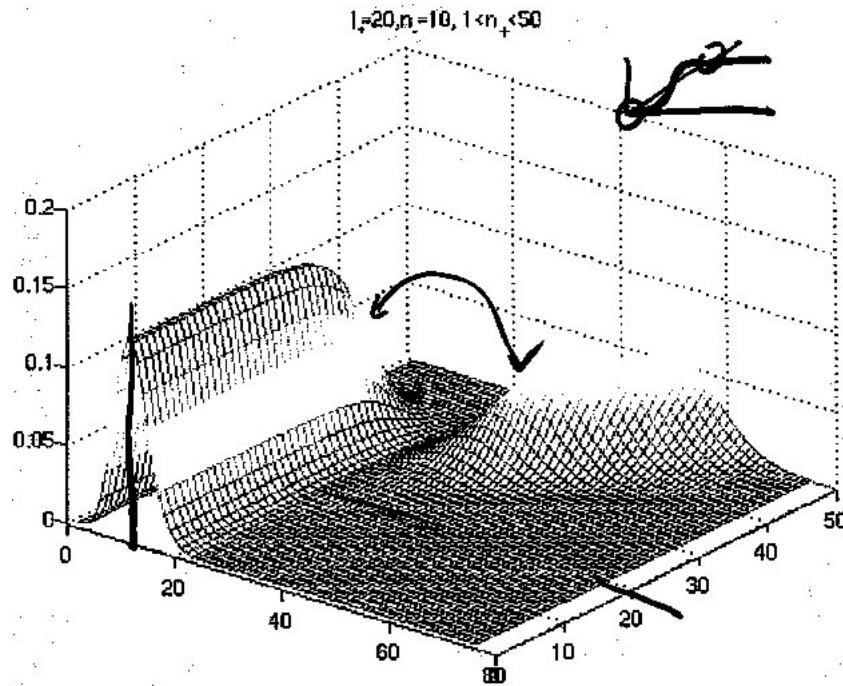


Figure 6.30: MATLAB

In general probability is conserved:

$$1 = \sum_{j=0}^{\infty} p_j \quad (6.364)$$

$$\dot{p}_j = \lim_{\tau \rightarrow 0} \sum_{k=0}^{\infty} p_{\tau}(j|k) p_k^{t-\tau} - p_j \quad (6.365)$$

$$0 = \sum_{j=0}^{\infty} \dot{p}_j = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} p_{\tau}(j|k) p_k^{t-\tau} - \sum_{j=0}^{\infty} p_j \quad (6.366)$$

$$= \sum_{k=0}^{\infty} p_k - \sum_{j=0}^{\infty} p_j = 0 \quad (6.367)$$

Consider the case of stuff being created.

$$\dot{p}_j = -p_j + p_{j-1} \quad (6.368)$$

$$\dot{p}_0 = -p_0 \quad (6.369)$$

$$\sum_{j=0}^{\infty} \dot{p}_j = - \sum_{j=0}^{\infty} \dot{p}_j + \sum_{j=1}^{\infty} \dot{p}_{j-1} \quad (6.370)$$

Or more carefully

$$\sum_{j=0}^{\infty} p_j = p_0 + \sum_{j=1}^{\infty} p_j \quad (6.371)$$

$$\sum_{j=0}^{\infty} \dot{p}_j = \dot{p}_0 + \sum_{j=1}^{\infty} \dot{p}_j = -p_0 + \sum_{j=1}^{\infty} (-p_j + p_{j-1}) \quad (6.372)$$

$$0 = -p_0 - \sum_{j=1}^{\infty} p_j + \sum_{j=1}^{\infty} p_{j-1} \quad (6.373)$$

$$0 = - \sum_{j=0}^{\infty} p_j + \sum_{j=0}^{\infty} p_k. \quad (6.374)$$

$$A \rightarrow \emptyset \quad (6.375)$$

$$\tau \dot{p}_j = -jp_j + (j+1)p_{j+1}, \quad j \geq 0 \quad (6.376)$$

## 6.9.2 Dynamics and nondimensionalization review

Also we have

$$\frac{dx}{dt} = \dot{x} = d_t k_2 \frac{x^n}{k^n + x^n} - k_7 x. \quad (6.377)$$

To avoid carrying constants through the calculations we nondimensionalize. Let

$$x = \bar{x}\tilde{x} \quad (6.378)$$

$$t = \bar{t}\tilde{t}. \quad (6.379)$$

Then the time derivative becomes

$$\frac{d}{dt} = \frac{d\tilde{t}}{dt} \frac{d}{d\tilde{t}} = \frac{1}{\bar{t}} \frac{d}{d\tilde{t}}. \quad (6.380)$$

So the nondimensional equation is

$$\frac{\bar{x}}{\bar{t}} \frac{d\tilde{x}}{d\tilde{t}} = d_t k_2 \frac{\bar{x}^n \tilde{x}^n}{k^n + \bar{x}^n \tilde{x}^n} - k_7 \bar{x} \tilde{x} \quad (6.381)$$

We want to find a nondimensional parameter to simplify the equation. Rewrite the equation as

$$\frac{d\tilde{x}}{d\tilde{t}} = d_t k_2 \frac{\bar{t}}{\bar{x}} \frac{\tilde{x}^n}{(k/\bar{x})^n + \tilde{x}^n} - k_7 \frac{\bar{x}\tilde{t}}{\bar{x}} \tilde{x} \quad (6.382)$$

Now let's take  $\bar{x} = k$ . Then, either

$$\bar{t} = \frac{\bar{x}}{d_t k_2} \quad \text{or} \quad \bar{t} = k_7^{-1}. \quad (6.383)$$

In the first case,

$$\frac{d\tilde{x}}{d\tilde{t}} = \frac{\tilde{x}^n}{1 + \tilde{x}^n} - \underbrace{\left( \frac{k_7 k}{d_t k_2} \right)}_R \tilde{x}. \quad (6.384)$$

And in the second,

$$\frac{d\tilde{x}}{d\tilde{t}} = \underbrace{\frac{d_t k_2}{k k_7}}_\gamma \frac{\tilde{x}^n}{1 + \tilde{x}^n} - \tilde{x}. \quad (6.385)$$

In both cases the equations are simpler with the constants  $R$  and  $\gamma$  defined above. This gives the two cases

$$\frac{d\tilde{x}}{d\tilde{t}} = \frac{\tilde{x}^n}{1 + \tilde{x}^n} - R\tilde{x} \quad (6.386)$$

$$\frac{d\tilde{x}}{d\tilde{t}} = \gamma \frac{\tilde{x}^n}{1 + \tilde{x}^n} - \tilde{x}. \quad (6.387)$$

In fact, the equations are even simpler if we notice

$$\gamma = \frac{d_t k_2}{k k_7} = R^{-1}. \quad (6.388)$$

Now we can consider fixed points, stability, etc. We now know that there is only one nondimensional parameter which controls the dynamics.

A fixed point is stable provided the eigenvalues of the Jacobian have negative real part. The Jacobian is

$$J [f_j(x_i)] = \begin{pmatrix} \partial_{x_1} f_1 & \partial_{x_2} f_1 & \dots \\ \partial_{x_1} f_2 & \partial_{x_2} f_2 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (6.389)$$

There are two dimensional dynamical systems such as

$$m\ddot{x} + b\dot{x} + kx = 0, \quad (6.390)$$

$$\begin{cases} \dot{x} = \frac{y^2}{1+y^2} - Ry \\ \dot{y} = x - R_2 y \end{cases} \quad (6.391)$$

Stability is determined by the Jacobian

$$J = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (6.392)$$

and especially its trace and determinant

$$\tau = a + d \quad (6.393)$$

$$\Delta = ad - bc. \quad (6.394)$$

Figure 6.31 shows the stability diagram.

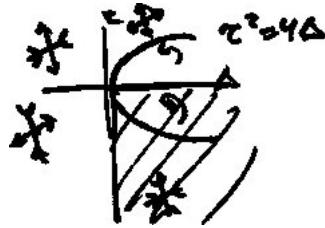


Figure 6.31: Stability diagram

We have stability if and only if

$$\tau < 0 \quad (6.395)$$

$$\Delta > 0. \quad (6.396)$$

You can have decay and oscillations together as in Figure 6.32

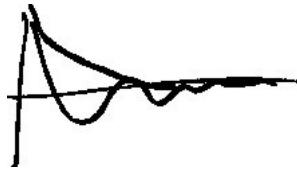


Figure 6.32: Decaying oscillation

### 6.9.3 Master equation kinetics

$$\dot{p}_n \rightarrow \frac{d}{dt} \langle n \rangle = \dot{x} V \quad (6.397)$$

$$p_n - p_{n-1} \approx \frac{d}{dn} p \quad (6.398)$$

$$p_n \propto P(x) \text{ such that } P(n/V) = p_n \quad (6.399)$$

$$p_{n\pm} = e^{\pm \partial_n} p_n = e^{\pm \frac{1}{V} \partial_x} P(x) \quad (6.400)$$

$$\approx \left[ 1 \pm \frac{1}{V} \partial_X + \frac{1}{2} \frac{1}{V^2} \partial_X^2 \right] P(x) + O(V^{-3}). \quad (6.401)$$

Recall

$$\dot{p}_j = -k_1 j p_j - V k_2 p_j + k_1(j+1)p_{j+1} + k_2 p_{j-1}V \quad (6.402)$$

$$A \xrightarrow{k_1} 0 \quad (6.403)$$

$$0 \xrightarrow{k_2} A \quad (\dot{a} = k_2) \quad (6.404)$$

$$\dot{p}_j = k_1 [e^{\partial_j} - 1] j p_j + k_2 V [e^{-\partial_j} - 1] p_j \quad (6.405)$$

We can work this out as previously.

$$p_j \propto P(x = jV). \quad (6.406)$$

Now letting  $x = j/V, j = x/V$  we have

$$\dot{P} = k_1 \left[ e^{\frac{1}{V}\partial_x} - 1 \right] xVP + k_2 V \left[ e^{-\frac{1}{V}\partial_x} - 1 \right] P \quad (6.407)$$

$$= k_1 \left[ \frac{1}{V} \partial_x + \frac{1}{2} \frac{1}{V^2} \partial_x^2 \right] xVP + k_2 V \left[ -\frac{1}{V} \partial_x + \frac{1}{2} \frac{1}{V^2} \partial_x^2 \right] P \quad (6.408)$$

$$= \partial_x \left\{ k_1 \left[ 1 + \frac{1}{2V} \partial_x \right] xP + k_2 \left[ -1 + \frac{1}{2V} \partial_x \right] \right\} \quad (6.409)$$

Now let  $-J$  be the quantity in braces.

$$J = -k_1 \left[ 1 + \frac{1}{2V} \partial_x \right] xP - k_2 \left[ -1 + \frac{1}{2V} \partial_x \right] \quad (6.410)$$

$$= -k_1 xP - \frac{k_1}{2V} P - \frac{k_1}{2V} xP' + k_2 P - \frac{k_2}{2V} P' \quad (6.411)$$

Recall the equation for diffusion with drift.

$$J = pv - Dp' \quad (6.412)$$

So we have

$$v = -k_1 x - \underbrace{\frac{-k_1}{2V}}_0 + k_2 \Rightarrow -k_1(x - x_0) \quad (6.413)$$

$$D = \frac{k_1 x}{2V} + \frac{k_2}{2V} = \frac{k_1}{2V}(x + x_0) \quad (6.414)$$

The macroscopic descriptions is

$$\dot{x} = k_2 k_1 x \quad (6.415)$$

So the stable fixed point is

$$x_* = \frac{k_2}{k_1} \quad (6.416)$$

What about the solution to the master equation?

$$0 = \dot{p}_j = -jp_j + (j+1)p_{j+1} - j_*p_j + j_*p_{j-1} \quad (6.417)$$

$$p_j = \frac{(j^*)}{j!} e^{-j_*} \quad (6.418)$$

$$\langle j \rangle = j_* \quad (6.419)$$

For an up-regulating gene

$$\dot{x} = \frac{x^n}{1+x^n} - Rx = H(x-1) - Rx. \quad (6.420)$$

A stochastic description of this is

$$\dot{p}_j = \underbrace{-\alpha(j)p_j + \alpha(j-1)p_{j-1}}_{\text{creation}} - \underbrace{jp_j + (j+1)p_{j+1}}_{\text{decay}}. \quad (6.421)$$

Assume

$$\alpha(j) = \begin{cases} \alpha_+, & j \geq j_0 \\ \alpha_-, & j < j_0 \end{cases}. \quad (6.422)$$

So

$$0 = -\alpha_+ P_{j_0} + \alpha_- p_{j_0-1} - j_0 p_{j_0} + (j_0+1) p_{j_0+1}, \quad j = j_0, \quad (6.423)$$

$$p_j = \frac{\alpha_+^j}{j!} c_+, \quad j \geq j_0, \quad (6.424)$$

$$p_j = \frac{\alpha_-^j}{j!} c_-, \quad j < j_0. \quad (6.425)$$

$$(6.426)$$

Now we match the terms at  $j_0$  to get  $c_+$  and  $c_-$ .

$$0 = -\alpha_+ c_+ \frac{\alpha_+^{j_0}}{j_0!} + \frac{\alpha_- \alpha_-^{j_0-1} c_-}{(j_0-1)!} - \frac{j_0 \alpha_+^{j_0}}{j_0!} c_+ + (j_0+1) \frac{c_+ \alpha_+^{j_0+1}}{(j_0+1)!} \quad (6.427)$$

So

$$\frac{c_-}{c_+} = \frac{(j_0-1)!}{\alpha_-^{j_0}} \left[ \frac{\alpha_+^{j_0}}{(j_0-1)!} + \frac{\alpha_+^{j_0+1}}{j_0!} - \frac{\alpha_+^{j_0+1}}{j_0!} \right] \quad (6.428)$$

$$= \left( \frac{\alpha_+}{\alpha_-} \right)^{j_0} \quad (6.429)$$

So for an autoregulating gene it looks like a Poisson distribution on either side of the transition and these distributions can be matched.

Let's discuss some MATLAB code to determine fixed points.

$$\dot{x} = H(x-1) - Rx + \alpha_0 \quad (6.430)$$

$$x_* = \frac{H(x-1)}{R} \quad (6.431)$$

We can schematically see where the fixed points will be in Figure 6.33 and 6.34.

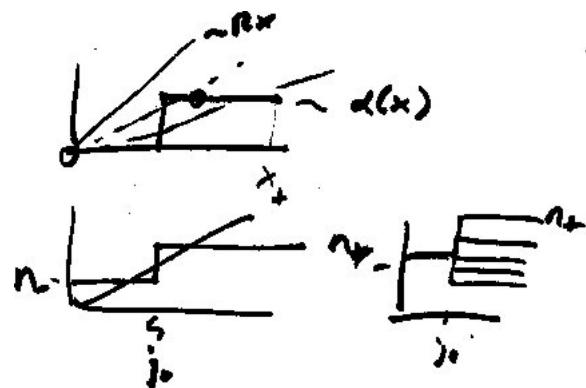


Figure 6.33: Fixed points

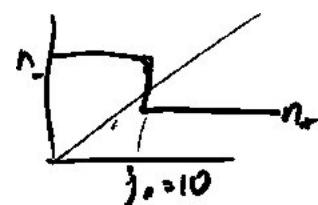


Figure 6.34: Fixed points

## Chapter 7

# Biological networks

orphan snippet:

“Scale-free” is a term used by some researcher to refer to a *property* if a network. It is sometimes misused as describing a *Pmodel*; but in fact there are many many models which can reproduce this one property.

A first statistics is the one-point statistic on nodes: what is the probability of finding a node with connectivity  $k$ ?  $P(k) = \sum_i \delta_{k_i, k} / N$

A second statistic is the clustering coefficient, a type of three-point statistic defined on edges:  $C = p(a - c | a - b, b - c)$  is a conditional probability on edges This quantity has a long history in social science, for example, asking what is the probability that two of your friends are friends with each other? Questions such as these led to a highly mathematical literature in the social sciences dating back at least to the 50s.

---

When people started looking at real-world networks in 1998 and 1999 drawn from the real world, and, in particular—

there were a pair of events. Applied mathematicians interested in dynamical systems on networks asked what shapes of networks could lead to synchronized dynamics? This is an example of macroscopic biomathematics – For example, if there are many fireflies oscillating (blinking on and off) what connectivity diagram among the fireflies could make possible global synchrony?

1999 as computational power became for advances,d and technological/information networks (principally the Internet, a physical network, and the www, an ideational network) became automatically surf-able (and thus measurable — not that this is without problems of statistical biases [?]), the was an increase in the literature

So there was this creation of new lit in applied mathematics math and cs and physics of people looking at networks from the word and asking about these particular statistics. These statistics also have the nice property that at least for very simple models, these quantities can be calculated analytically. (as we will see below)

When looking at the real world, physicists and computer scientists looked at the Internet and the world-wide web [?][?][?] looked at the Internet and the www and noticed a particular shape for the statistic  $P(k)$ : namely  $P(k) \sim k^{-\alpha}$ . This property was dubbed

“scale free” and inspired models to try to explain this property. It’s useful to distinguish between two general classes of such models, “preferential attachment” and “duplication mutation” although several authors have noted that they both create a bias in the growth of the network which then generates a scale-free distribution.

In the “preferential attachment” models, we imagine a network evolving by creation of new nodes which are attached to existing nodes randomly, with a probability proportional to the degree of the existing nodes.<sup>1</sup>

— Examples include

transcriptional regulatory networks, a.k.a “genetic networks” protein protein interaction networks

## 7.1 Questions and methods

### 7.1.1 Network statistics

### 7.1.2 Form vs function

handout: Guet

punchline: form does not imply function

If you only know form, you only know form.

## 7.2 What is to be measured?

### 7.2.1 Qualitative questions

directed or not?

unipartite or multipartite?

### 7.2.2 Quantitative features

**Number of nodes**

**Number of edges**

**Size**

**Lattices**

**Trees**

---

<sup>1</sup>Much has been written [?, 12] about how, although this was the first time such a model had been proposed for description of the Internet, the mathematical and mechanistic roots of the model can be traced to 1977, or even 1956, or even 1026, or even the gospel of Matthew. As such, there is little reason to expound on this history here.

## Statistics

### One-point: $P(k)$

One of the nice features of degree distributions is that, for many simple models, we can calculate these analytically.

However, this particular statistic may not be useful for the inverse problem – that is, it might be the case that many different models can produce a particular degree distribution, making degree distributions of limited utility for assessing to what extent one model is better than any other as a description of a particular network.

are many many models which could produce a scale-free distribution.

### Two-point: $c$ and $C$

## 7.3 Modeling

‘Random models or real world networks’<sup>2</sup> is becoming a more and more common problem in biological modeling, with contributions from physicists, applied mathematicians, discrete mathematicians, evolutionary biologists, and computer scientists. (handout: Keller)

The models constructed are largely motivated by simple considerations of some evolutionary mechanism, for example, duplication and mutation, or the preferential interaction of newly-created proteins (or genes) with highly connected existing ones. The goal of these models is then to study to what extend the graph structure<sup>3</sup> of the network can be explained by simple evolutionary mechanisms, perhaps quantified by a few parameters, as we study below.

A research question then is, given a real-world biological network, what is to be modeled? What attributes of the network should be considered relevant?

It’s useful to answer the above question to ask what is relevant. If we were to try to build a model of the form  $y = f(x)$ , where  $x$  is some attribute of the network, what is  $y$ ? How about:  $y$ =What evolutionary mechanism could best describe that graph structure?

### 7.3.1 Graph drawing: a caveat

Graph drawing is a black art. Frequently one draws a graph of a network. The coordinates here may be chosen via a number of algorithms, but are generally meaningless<sup>4</sup>. This is bothersome if we think of a networks as something that should exist in two or three dimensions, or even some metric space with a distance defined between every pair of notes. This frustration is lifted, however, on realizing that there is nothing dimensionful about a network – there is no reason for a distance measure to exist between every pair of nodes,

---

<sup>2</sup>the phrase is in quotes because it is taken from the name of a recent workshop on the subject:

<sup>3</sup>a.k.a ‘topology’ or even ‘shape’

<sup>4</sup>an exception of spectral embedding, where the coordinates have a mathematical meaning

and the picture is an art project – an aid to comprehension, but not necessarily the unique solution to any well-defined optimization problem. [12]

Sometimes all you can say is statistical.

**Homework:**

if you have 3 friends, and they have 3 friends, and so on, and so on, how many degrees of separation should there be between you and every one of the 6,416,362,206 or so people on the planet?

**7.3.2 Subgraph census****7.3.3 Price's model**

5

---

<sup>5</sup>this discussion taken from mark Newmann's review article

## **Part III**

# **Learning models from data**



Reality is created out of confusion and contradiction, and if you exclude those elements, you're no longer talking about reality. You might think that – by following language and a logic that appears consistent — you're able to exclude that aspect of reality, but it will always be lying in wait for you, ready to take its revenge.

—Haruki Murakami



# Chapter 8

## Complexity Control

### 8.1 Motivating example: inferring networks

#### 8.1.1 Genetic networks

Our first model of transcriptional regulation featured a single, auto-inducing gene. We also studied networks in which only two or three genes interact with each other.

Finally, in the last chapter, we looked at statistical descriptions of networks, in the case of many nodes. But how many genes do we expect to interact with each other *in vivo*?

#### 8.1.2 Scales of genomes

- λ phage: 50 genes
- E coli: 3000 genes
- Yeast: 6000 genes
- man, mouse: 25,000 genes

Wow. well, that's how many nodes there are, but how many edges?

#### 8.1.3 Densities of inferred genetic networks

- E coli: 423 nodes and 519 edges
- uh... who says?

#### network inference

Many of these individual interactions are learned through single experiments on individual pairs. However, many of these interactions are inferred statistically within the context of a given model. In these cases, an “edge” may not represent the binary outcome of

an individual experiment (*e.g.*, “the protein bound to the DNA”) but something far more complicated.

When someone shows you a network, always ask what an edge means, and in the context of what modeling.

Let’s take a simple example of how we might form a correspondence between several observables and a mathematical model

handout: reduce

Q: when should we stop?

### 8.1.4 Overfitting data

## 8.2 ‘Model selection’

### 8.2.1 The Bayesian approach

#### Regression

#### In general

(handout: Schwartz 1978) [14]

#### Conditional dependence tables and “Bayes nets”

Consider an experiment where you can tune the concentration of 7 different transcription factor. We have GFP-labeled a naturally-occurring gene in an organism and want to know which of these transcription factors bind to the target gene.

All we can do is perform a number of experiments. More typically, we download the dataset from an experimentalist who has already done the experiment and asks us to infer which proteins are controlling the target gene.

You might start out with a simple model as in REDUCE.

But what if all of the genes are potential targets?

handout: Heckerman [11]

handout: Nir Friedman 04 [9]

### 8.2.2 The machine learning approach

Cross validation

## 8.3 Statistical significance

### 8.3.1 Use and abuse of $p$ -values

### 8.3.2 Bootstrap vs biology

## 8.4 Data models

### 8.4.1 Multinomial density estimation

### 8.4.2 Mixture model density estimation

### 8.4.3 Learning continuous distributions

## 8.5 Applications

### 8.5.1 Microarrays

The third revolution in biology in the last decade was the massively-parallel upscaling of traditional, one-pair-at-a-time studies to study thousands of genes or proteins at once. The most famous of these technological advances are DNA microarrays.

### 8.5.2 Binding sites as “motifs”

Must assume a background to asses a p-value

### 8.5.3 Binding sites as binding sites

Must assume a loss function; then can check CV

handout: Nir Friedman



## Chapter 9

# Complexity Control

In Part II of this book we studied how the dynamics of transcriptional regulatory networks may be modeled, even in the presence of noise. A reasonable question to ask is how we know which proteins regulate which genes. After all, there are thousands or tens of thousands of genes in the model organisms studied in biology; how can we be sure we have found the regulatory interactions which are the most important in building our models? For that matter, how do we quantify ‘most important’?

In Part III of this book we will begin to study how we learn models of biology from data. One challenge which makes biology different from many other sciences is its tremendous complexity, i.e., for many biological phenomena of interest it is not obvious what is the right or best model. Often the phenomenon of interest involves many distinct players which can not be treated as identical parts (e.g., in the diffusing pollen grains of Part I). For these and other reasons this part of the book will be in part about what it means to build a ‘good’ model, including not just how to learn (a.k.a ‘infer’ or ‘fit’) parameters in a model to data, but more importantly when to stop adding parameters. In the models of Part I and Part II, the number of parameters was set by how complicated a model we were willing to calculate with; armed with a finite data set, however, we may ask instead how many parameters we should be ‘allowed’ to learn.

We will build, initially, on the unifying mathematical theme of the book: using probability to think precisely about uncertainty in modeling. In Part I and Part II we focused on dynamics, modeling  $p_{t+1}(x)$  as a functional of  $p_t(x)$ . Here we will be interested in models which relate the data we have observed to parameters we wish to learn from these data.

### 9.1 Learning regulatory interactions from data

The models of Part II rested on knowing which protein concentrations (or copy numbers!) determine the production rates of the other protein-encoding genes. How might we learn which such possible interactions are to be included in our models? In the late 20th century a number of experimental groups began developing massively-parallel experiments to reveal

the ‘on-off’ state of all the genes in an organism at once<sup>1</sup>. Beginning with the model organism *Saccharomyces cerevisiae* AXXX, these experiments measured the abundance of mRNA for thousands of genes simultaneously by patterning a small array – called a microarray – with nucleic acid sequences specific to each of the genes. The resulting data changed the relationship between biology and quantitative modeling.

What made this change possible? First, one needs to know the genomic sequence of the organism, which for many organisms only became available at the end of the 20th century. Secondly, the World Wide Web facilitated sharing the resulting data with experimental and mathematical researchers across the globe, encouraging researchers from previously-isolated fields to begin collaborating.

As we discussed in Part II, the rate of production of mRNA for one gene is a function of the abundance of transcription factors which bind to the regulatory region associated with the gene, provided that this region contains the correct sequence element for the transcription factor protein to bind. Knowing the sequence data associated with all the genes in an organism, then, along with the abundance of mRNA for all of these genes, we should be able to correlate presence or absence of different sequence elements with expression data. Many groups began modeling this relationship; for the purposes of this chapter we will limit ourselves to discussing one such model, called “REDUCE”, which implements the simple model

$$y_g = \sum_{\mu} N_{\mu g} c_{\mu} \equiv f_g(\mathbf{c}), \quad (9.1)$$

that is, the expression signal  $y_g$  of the gene  $g$  is posited to be linearly related to the number  $N$  of occurrences of the motif  $\mu$  in the regulatory region of  $g$ , multiplied by an unknown ‘charge’  $c_{\mu}$  associated with each motif. The expression data  $y_g$  are provided by the experiment; the number  $N_{\mu g}$  is known once one has sequenced the organism, identified the location of all the genes in the genome and defined what one means by ‘regulatory region’; and the ‘charges’  $c_{\mu}$  are the parameters one wishes to learn from the data.

### 9.1.1 Models, data, and uncertainty

Of course, we do not expect the data and the model to agree perfectly, so as in Part I and Part II we are really interested in modeling probabilities. In particular we may treat the experimental data as noisy — i.e., described by a probability distribution. The most common way of modeling our uncertainty is one which first asserts an additive noise  $\xi_g$  for each observation (here, for each gene):

$$y_g = f_g(\mathbf{c}) + \xi_g ?? \quad (9.2)$$

and, most importantly, we assert that we know the distribution from which this noise is drawn. The most common assumption, one which leads to the simplest algebra and which

---

<sup>1</sup>For the purposes of this text we will omit the fascinating history of the technologies producing these data. Students are encouraged to investigate online resources such as <http://dnai.org> for less cursory explanations of the technologies which produced the data types we will be discussing

can be derived under a relatively benign set of assumptions<sup>2</sup>, is the normal distribution  $p(\xi_g) = N(\xi_g | 0, \sigma)$ , where we must posit an error bar for the measured signal. Just as in Part I and II, our model can be incorporated by a conditional distribution which allows one and only one value for the random variable:

$$p(y_g | \mathbf{c}, \xi) = \delta(y_g - f_g(\mathbf{c} - \xi_g)). \quad (9.3)$$

Now, instead of a delta function for a discrete variable (the ‘Kroneker delta’), we have a delta function defined on the real-valued expression signal (the ‘Dirac delta’). Following the now-familiar sum-product rules of marginalization, we have

$$p(y_g | \mathbf{c}) = \int d\xi_g p(y_g | \mathbf{c}, \xi_g) p(\xi_g) = N(y_g - f_g(\mathbf{c}) | 0, \sigma). \quad (9.4)$$

For  $N$  independent observations, these probabilities multiply and we have<sup>3</sup>

$$p(\vec{y} | \mathbf{c}) = \frac{e^{-\frac{1}{2\sigma^2} \sum_{g=1}^N (y_g - f_g(\mathbf{c}))^2}}{\sqrt{2\pi\sigma^2}^N}. \quad (9.5)$$

This quantity, the probability of observed data given the set of unknowns we wish to find, is called the *likelihood*  $L$ ; maximizing this likelihood serves as the basis for almost all statistical analysis of experimental data in the natural sciences. Students who have taken experimental lab courses will recognize the term in the exponent as ‘ $\chi^2$ ’ or the ‘residual sum of squared errors’  $S$ , i.e.,

$$L = \quad (9.6)$$

and it will be clear that, since  $L$  is monotonically decreasing in  $\chi^2$  and  $S$  we have the relationship

$$\begin{array}{lll} \mathbf{c}^* = \underset{\mathbf{c}}{\operatorname{argmax}} & (9.7) & (9.9) \\ (9.8) L = \underset{\mathbf{c}}{\operatorname{argmin}} & (9.10) \chi^2 = \underset{\mathbf{c}}{\operatorname{argmin}} & (9.11) \\ (9.12) S \end{array}$$

That is, the familiar act of minimizing  $\chi^2$  is a special case of the more general problem of ‘maximizing likelihood’.

### 9.1.2 Learning parameters from data

## 9.2 Lecture 18 - Data

Regulatory sequence, transcription factors.

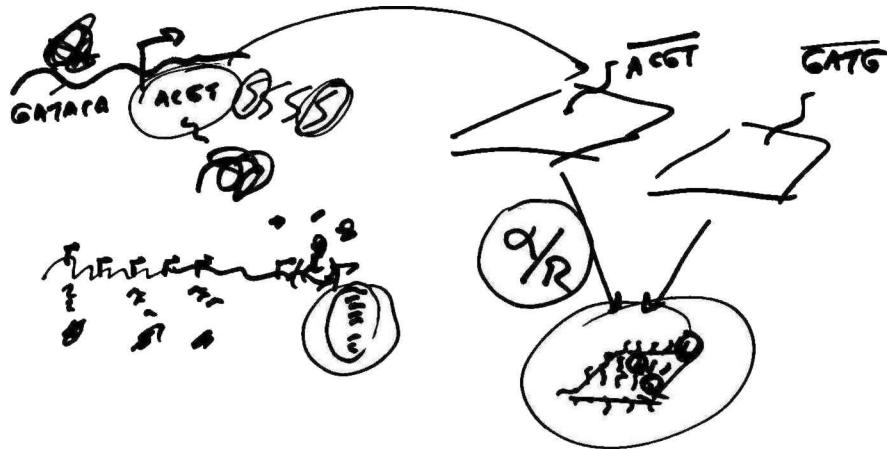


Figure 9.1: GATACA

### 9.2.1 Finding keywords in RNA sequence

Querying genome.

Models from data.

$$x_l = N \tanh\left(\frac{F_l}{K_a T}\right)^l$$

$F$

$$\text{Figure 9.2: } \frac{x}{l} = N \tanh\left(\frac{Fl}{K_a T}\right)$$

transcription factor, regulatory binding sites

Correlate words with genes that have word.

Define the extension  $x(F, l)$ :

$$x = Nl \tanh\left(\frac{Fl}{K_a T}\right) \quad (9.14)$$

---

<sup>2</sup>This derivation is usually called the ‘central limit theorem’.

<sup>3</sup>although  $y$  and  $c$  are both vectors, we use the mixed notations  $\vec{y}$  and  $c$  to emphasize that the former is a vector over the  $N$  observations and the latter over the  $M$  unknowns (i.e., the sequence elements).



Figure 9.3: GATACA

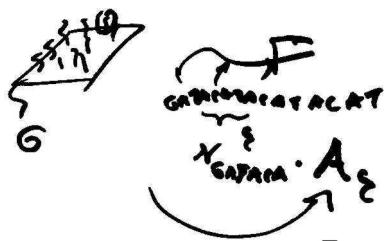


Figure 9.4: mRNA microarray

$$x = f(F, l) \quad (9.15)$$

Find best  $l$  to fit  $(x, F)$  data.



Figure 9.5: data

Linear Model:

$$A_e = N_e \theta \quad (9.16)$$

This is analogous to  $x = f(F, l) \approx \langle x \rangle_0 + \frac{F}{T} \langle x^2 \rangle_0$  from part 2 of the course. Measure  $A$  and  $N$ . Then we can use model to find  $\hat{\theta} = A/N$ . But you have an ensemble of measurements. How do you find the best value? Averages?

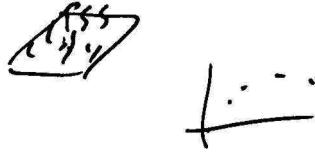


Figure 9.6: microarray data

### 9.2.2 Minimize squared error

$$\hat{\vartheta} = \frac{\langle A_g \rangle}{\langle N_g \rangle} \quad (9.17)$$

Minimize error:

$$\mathcal{E}[\theta] = (A_g - N_g\theta)^2 \quad (9.18)$$

$$\hat{\vartheta} = \arg \min_{\theta} \mathcal{E}[\theta] \quad (9.19)$$

$$\mathcal{E} = (A - N\vartheta)^2 \Rightarrow \hat{\vartheta} = \frac{A}{N} \quad (9.20)$$

$$0 = \partial_{\vartheta} \mathcal{E} = 2(A - N\vartheta)(-N) \Rightarrow \vartheta = \frac{A}{N} \quad (9.21)$$

$$\mathcal{E} = \sum_g \mathcal{E}_g = \sum_g (A_g - N_g\vartheta)^2 \quad (9.22)$$

$$0 = \partial_{\vartheta} \mathcal{E} = \sum_g 2(A_g - N_g\vartheta)(-N_g) = -2 \sum_g A_g N_g + 2\vartheta \sum_g N_g^2 \quad (9.23)$$

$$\hat{\vartheta} = \frac{\sum_g A_g N_g}{\sum_g N_g^2} \equiv \frac{\mathbf{A} \cdot \mathbf{N}}{\mathbf{N} \cdot \mathbf{N}} \equiv \frac{\langle A | N \rangle}{\langle N | N \rangle} \quad (9.24)$$

$$N_g = \text{number of times "GATACA" appears in promoter region of gene } g \quad (9.25)$$

### 9.2.3 Finding “important” words

Suppose you don't know what word you are interested in. Really have a matrix of times you have a particular word,  $N_g^\mu$ . Now consider a more complicated model. We have

$$A_g = \sum_{\mu} N_g^\mu \vartheta_{\mu} \quad (9.26)$$

$$\mathcal{E} = \sum_g (A_g - \sum_{\mu} \mu N_g^\mu \vartheta_{\mu})^2 = \mathcal{E}[\vartheta] \quad (9.27)$$

$$\vartheta_1 = \text{GATACA} \quad (9.28)$$

$$\vartheta_2 = \text{TACAT} \quad (9.29)$$

$$\partial_{\vartheta_\nu} \mathcal{E} = \sum_g 2(A_g - \sum \mu N_g^\mu \vartheta_\mu)(-N_g^\nu) \quad (9.30)$$

$$= -2 \sum_g N_g^\nu A_g + 2 \sum_\mu \sum_g N_g^\mu N_g^\nu \vartheta_\mu \quad (9.31)$$

$$M\vartheta = \mathbf{b} \quad (9.32)$$

$$M_{\mu,\nu} = \sum_g N_g^\mu N_g^\nu \quad (9.33)$$

$$b_\nu = \sum_g N_g^\nu A_g \quad (9.34)$$

$$N\vartheta = A \Rightarrow \vartheta \neq \frac{A}{N} \quad (9.35)$$

$$\vartheta = \frac{\sum_g N_g A_g}{\sum_g N_g^2} \quad (9.36)$$

If  $M^{-1}$  exists we have

$$\hat{\vartheta} = M^{-1}\mathbf{b} = (\sum_g N_g^\mu N_g^\nu)^{-1}(\sum_g N_g^\nu A_g) \quad (9.37)$$

Things that can go wrong when fitting the best statistically independent length given a set of  $n$  observations.

Question: how many unknowns can we learn from these data? Mathematically, need  $M$  invertible.

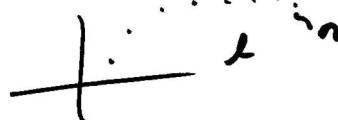


Figure 9.7: data plot

$$M = \sum_g N_g^\mu N_g^\nu \quad (9.38)$$

$$= \sum_g |N_g^\mu\rangle\langle N_g^\nu| \quad (9.39)$$

$$= \sum_{g=1}^n \mathbf{N}^\mu \mathbf{N}^{\nu T} \quad \text{???? since } N^\mu \in \mathbb{R}^k \quad (9.40)$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \begin{pmatrix} q_1 & q_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \begin{pmatrix} u_1 & u_2 \end{pmatrix} \quad (9.41)$$

Why did we write down the loss function. Recall in beginning of class the example of the pollen grain. Let the grain take a random walk.

$$x_{t+1} = x_t + \eta_t \quad (9.42)$$

$$P(\eta_t) = \begin{cases} 1 = \eta \Rightarrow P = 1/2 \\ -1 = \eta \Rightarrow P = 1/2 \end{cases} \quad (9.43)$$

$$A = N\vartheta + \eta \quad (9.44)$$

$$P(\eta) = P(-\eta) \quad (9.45)$$

$$P(\eta) = G(\eta|0, \sigma) \text{ normal distribution} \quad (9.46)$$

$$= \frac{e^{-\frac{1}{2} \frac{\eta^2}{\sigma^2}}}{\sqrt{2\pi\sigma^2}} \quad (9.47)$$

Now we can reinterpret “best” fit.

$$\hat{\vartheta} = \arg \min_{\vartheta} \mathcal{E} = \arg \max_{\vartheta} P(\vartheta|A) \quad (9.48)$$

$$(9.49)$$

We have  $P(A|\eta, \vartheta)$  and so

$$x_{t+1} = x_t + \eta \Leftrightarrow P(x_{t+1}|x_t, \eta) = \delta_{x_{t+1}, x_t + \eta} \quad (9.50)$$

$$A = N\vartheta + \eta \Leftrightarrow \underbrace{P(A|\vartheta, \eta)}_{\delta_{x_{t+1}, x_t + \eta}} = \delta(A - N\vartheta - \eta) \quad (9.51)$$

$$P(x_{t+1}) = \sum_{x_t} \sum_{\eta_t} \underbrace{P(x_{t+1}|x_t, \eta)}_{\delta(A - N\vartheta - \eta)} \underbrace{P(\eta)}_{G(\eta)} P(x_t) \quad (9.52)$$

$$(9.53)$$

$$P(A|\vartheta) = G(A - N\vartheta) = \frac{e^{-\frac{1}{2\sigma^2}(A - N\vartheta)^2}}{\sqrt{2\pi\sigma^2}} \quad (9.54)$$

$$\hat{\vartheta} = \arg \min_{\vartheta} \mathcal{E}[\vartheta] = \arg \max_{\vartheta} \frac{e^{-\frac{1}{2\sigma^2}\mathcal{E}^2}}{\sqrt{2\pi\sigma^2}} = \arg \max_{\vartheta} P(A|\vartheta) \quad (9.55)$$

Why we fit. Have an error model of the distribution that error is drawn. Have a model for how data and an unknown parameter, an error, are related to each other. We are trying to find value of unknown parameter which maximizes likelihood of the data given the unknown parameter.

Now consider a vector of data.

$$A_g = N_g \vartheta + \eta_g \quad (9.56)$$

$$P(\eta) = P(\eta_1)P(\eta_3)P(\eta_3)\dots \quad (9.57)$$

$$= \frac{e^{-\frac{1}{2} \frac{\eta_1^2}{\sigma^2}}}{\sqrt{2\pi\sigma^2}} \frac{e^{-\frac{1}{2} \frac{\eta_2^2}{\sigma^2}}}{\sqrt{2\pi\sigma^2}} \dots \quad (9.58)$$

$$= (2\pi\sigma^2)^{-N/2} e^{-\frac{1}{2} \sum_{g=1}^N \frac{\eta_g^2}{\sigma^2}} \quad (9.59)$$

$$P(\mathcal{E}) = \sum_g \frac{\eta_g^2}{\sigma^2}, \quad \text{Distribution of total error, } \chi^2 \text{ or Boltzman} \quad (9.60)$$

$$P(\mathbf{A}|\vartheta) = \int d^N \eta \underbrace{P(\mathbf{A}|\vartheta, \eta)}_{\delta(\mathbf{A} - (\mathbf{N}\vartheta + \eta))} P(\eta) \quad (9.61)$$

$$= \frac{e^{-\frac{1}{2} \sum_{g=1}^N (\mathbf{A} - \mathbf{N}\vartheta)^2 / \sigma^2}}{(\sqrt{2\pi\sigma^2})^N} \quad (9.62)$$

Finding minimum error  $\vartheta$  is the same as finding maximum likelihood  $\vartheta$ . Again how many motifs can we learn from data? What is  $P(k|A)$ ?

- Hw 1: read reduce
- Hw 2:  $P(\mathbf{A}|\vartheta)$  where  $\vartheta \in \mathbb{R}^k$  given  $\mathbf{A} = \sum N\vartheta + \eta$  and  $P(\eta) = G(\eta)$ .

## 9.3 Lecture 19

### 9.3.1 Review

How many unknowns should you be allowed to learn from a finite amount of data? Suppose you have the data in the figure:

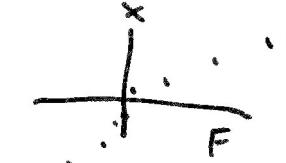


Figure 9.8: data

Simple model with quadratic error to fit data:

$$\langle x \rangle = Nl \tanh\left(\frac{Fl}{K_B T}\right) \quad (9.63)$$

$$\mathcal{E}[\theta] = (y_i - f(x_i))^2 \quad (9.64)$$

$$\hat{\vartheta} = \arg \min_{\theta} \mathcal{E}[\theta] \quad (9.65)$$

$$y_i = f_\vartheta(x_i) + \eta_i \Rightarrow p(y, \vartheta, \eta) = \delta(y - (f + \eta)) \quad (9.66)$$

$$p(\eta_i) = \mathcal{G}(\eta_i | 0, \sigma) \quad (9.67)$$

This is like the pollen grain from the beginning of the class

$$x_{n+1} = x_n + lx_n \quad (9.68)$$

$$p(\eta = 1) = 1/2 \quad (9.69)$$

A marginal distribution:

$$p(x_{t+1}) = \sum_{x_t} \sum_{\eta_t} p(x_{t+1}|x_t, p_t) p(x_t) p(\eta_t) \quad (9.70)$$

$$p(a) = \sum_{b,c} p(a|b,c) p(b,c) \quad (9.71)$$

We do the same thing here.

$$p(\mathcal{A}|\vartheta) = \sum_{\eta} p(\mathcal{A}|\vartheta, \eta) p(\eta) \quad (9.72)$$

$$= \int d\eta \delta(\mathcal{A} - (f_\vartheta + \eta)) \mathcal{G}(\eta) \quad (9.73)$$

$$= \mathcal{G}(\mathcal{A} - f_\vartheta) \quad (9.74)$$

$$= \frac{e^{-\frac{1}{2\sigma^2}(\mathcal{A}-f_\vartheta)^2}}{\sqrt{2\pi\sigma^2}} \text{ "likelihood"} \quad (9.75)$$

Now the regression.

$$f_\vartheta(x_i) = \sum_{\mu} \vartheta_{\mu} \mathcal{N}_{\mathcal{E}}^{\mu} \approx \mathcal{A}_{\mathcal{E}} \quad (9.76)$$

$$y \approx \sum_{j=0}^{K-1} c_j(x_i)^j = \sum_{j=1}^K \vartheta_j \Phi_j(x_i) \quad (9.77)$$

$$\Phi_j(x_i) = (x_i)^{j-1} \quad (9.78)$$

For example

$$y = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \dots \quad (9.79)$$

$$= \sum_{j=0}^3 c_j(x_i)^j = f_{\mathbf{c}} x_i \quad (9.80)$$

All linear regression can be written

$$\mathbf{y} = H\vartheta \iff y_i = \sum_{j=1}^K H_{ij} \vartheta_j = \sum_{j=1}^K \Phi_j(x_i) \vartheta_j \quad (9.81)$$



Figure 9.9: Heavisides

The error is

$$\mathcal{E} = \sum_{i=1}^n (y_i - f_\vartheta(x_i))^2 = (\mathbf{y} - H\vartheta)^2 \quad (9.82)$$

$$= |y|^2 - 2\mathbf{y}^T H\vartheta + (H\vartheta)^2 \quad (9.83)$$

$$\partial_{\vartheta_l} \mathcal{E} = -2\mathbf{y}^T H \hat{\mathbf{e}}_l + \left[ \sum_{i,k=1}^N \sum_{j_1,j_2} H_{ij_1} \vartheta_{j_1} H_{kj_2} \vartheta_{j_2} \right]_{\vartheta_l} \quad (9.84)$$

$$= \sum_{i,k=1}^N \sum_{j_1,j_2} H_{ij_1} \delta_{j_1 l} H_{kj_2} \vartheta_{j_2} + H_{ij} \vartheta_{j_1} H_{kj_2} \delta_{j_1 l} \quad (9.85)$$

Should get

$$2H^T H\vartheta \quad (9.86)$$

- HW 1: Show if  $f = H\vartheta$  then  $\hat{\vartheta}_{ML}$  (*ML*=“Maximum likelihood”) solves  $y^T H = H^T H\vartheta$ .

Then

$$y^T H = (H^T H)\vartheta \Rightarrow \vartheta_{ML} = (H^T H)^{-1}(y^T H) \quad (9.87)$$

The invertibility of the matrix depends on the basis functions. For example, given a model, if you write

$$f = c_1 \sin(x) + c_2 \cos(x) + c_3 4 \sin(x) \quad (9.88)$$

Then the resulting matrix does not have linearly independent columns and is thus not invertible. Doing linear regression says a lot about the solvability of the problem.

$$H^T H = \sum_{i=1}^N \phi_{K1}(x_i) \phi_{K2}(x_i) \quad (9.89)$$

$$= \sum_{i=1}^N |\phi_{K1}\rangle \langle \phi_{K2}| \quad (9.90)$$

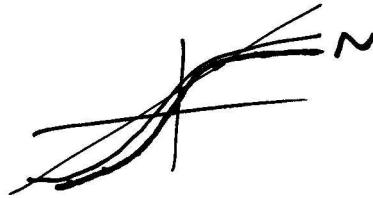


Figure 9.10: fit plot

This tells us how much we can fit the data.

Bayes' rule:

$$\mathcal{E} \rightarrow p(\mathcal{A}|\vartheta) \neq p(\vartheta|\mathcal{A}) \quad (9.91)$$

$$p(a, b) \equiv p(a|b)p(b) \quad (9.92)$$

$$\equiv p(b|a)p(a) \quad (9.93)$$

Given

$$\mathcal{A} = f_\vartheta + \eta \quad (9.94)$$

$$p(\eta) = \dots \quad (9.95)$$

then

$$\underbrace{\frac{p(\mathcal{A}|\vartheta)}{\text{likelihood}}}_{\text{posterior evidence}} = \frac{\underbrace{p(\vartheta|\mathcal{A})}_{\text{prior}} \underbrace{p(\mathcal{A})}_{Z}}{\underbrace{p(\vartheta)}_{\text{prior}}} \quad (9.96)$$

$$\arg \max_{\vartheta} p(\vartheta|\mathcal{A}) = \hat{\vartheta}_{MAP}, \quad \text{maximum a posteriori} \quad (9.97)$$

$$p(\vartheta|\mathcal{A}) = \frac{p(\mathcal{A}|\vartheta)p(\vartheta)}{\underbrace{p(\mathcal{A})}_{Z}} \quad (9.98)$$

Say you believe values of  $\vartheta$  are small. Then

$$p(\vartheta) = \frac{e^{-\frac{1}{2l_1^2}\vartheta_1^2}}{\sqrt{2\pi l_1^2}} \frac{e^{-\frac{1}{2l_2^2}\vartheta_2^2}}{\sqrt{2\pi l_2^2}} \dots \quad (9.99)$$

$$= \frac{e^{-\frac{1}{2}\vartheta^T L^{-2}\vartheta}}{\sqrt{|2\pi L^2|}} \quad (9.100)$$

$$p(\mathcal{A}|\vartheta) = \frac{e^{-\frac{1}{2}x^2}}{\left(\sqrt{2\pi\sigma^2}\right)^N} \quad (9.101)$$

Then

$$p(\vartheta|\mathcal{A}) = \frac{e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}\vartheta^T L^{-2}\vartheta}}{\underbrace{\sqrt{|2\pi\sigma^2|}}_{N \times N} \underbrace{\sqrt{|2\pi L^2|}}_{K \times K} p(\mathcal{A})} \quad (9.102)$$

Max posterior value of  $\vartheta$

$$L^K p(\vartheta|\mathcal{A}) = \frac{e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}\vartheta^T L^{-2}\vartheta}}{Z} \quad (9.103)$$

Why is there an  $L^K$ ? Think about the dimensions of

$$p(x), \quad \int p(a) da = 1. \quad (9.104)$$

So

$$[p(\vartheta|a_1, a_2, a_3)] \sim \frac{1}{[\vartheta]} = \frac{1}{[\vartheta_1]} \frac{1}{[\vartheta_2]} \frac{1}{[\vartheta_3]} \quad (9.105)$$

Also,

$$\hat{\vartheta}_{MAP} = \arg \max_{\vartheta} p(\vartheta|\mathcal{A}) = \arg \min_{\vartheta} x^2 + \frac{1}{2} \vartheta^T L^{-2} \vartheta \quad (9.106)$$

$$\nabla_{\vartheta} x^2 = H^T \vartheta - y^T H \quad (9.107)$$

- HW 2: Show

$$\nabla_{\vartheta} \frac{1}{2} \vartheta^T L^{-2} \vartheta = L^{-2} \vartheta \quad (9.108)$$

Then

$$(H^T H + L^{-2}) \vartheta = y^T H \quad (9.109)$$

How do we quantify how close to invertible a matrix is?

$$M^{-1} = \sum_{\alpha=1}^N \frac{1}{\lambda_\alpha} v_\alpha v_\alpha^T = \frac{1}{\|\lambda\|_\infty} \sum_{\alpha=1}^N \frac{\|\lambda_\alpha\|}{\lambda_\alpha} v_\alpha v_\alpha^T \quad (9.110)$$

$$cond(M) = \frac{\lambda_{smallest}}{\lambda_{biggest}} \quad (9.111)$$

We can regularize a problem to make it well-posed.

$$(H^T H + L^{-2}) v = (\lambda + L^{-2}) v \quad (9.112)$$

$$cond = \frac{\lambda_S + L^{-2}}{\lambda_B - L^{-2}} \rightarrow 1 \quad (9.113)$$

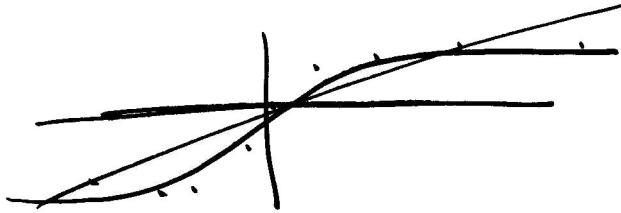


Figure 9.11: data plot

So we change the model when we introduce the regularization (or parameterization). But this allows us to fit data.

Since you can't know all the parameters you make assumptions.

$$p(\mathcal{A}|\vartheta) \rightarrow \hat{\vartheta}_{ML} \quad (9.114)$$

Or based on some belief

$$p(\vartheta|\mathcal{A}) \rightarrow \hat{\vartheta}_{MAP}[p(\vartheta)] \quad (9.115)$$

$$\hat{\vartheta}_{MAP} = (H^T H + L^{-2})^{-1} y^T H \quad (9.116)$$

This assumes data is distributed nicely. On the other hand, if the distribution is wide you might want to use a high parameter fit.

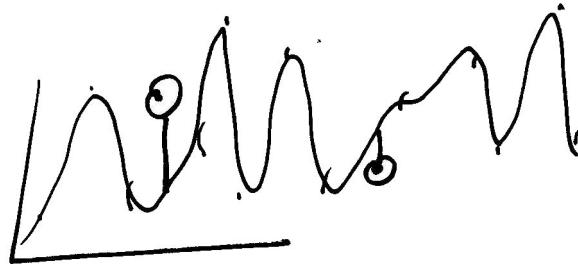


Figure 9.12: polynomial fit plot

This model should not have good predictive power. This is a problem with choosing the right complexity of a model.

Consider the probability  $p(\mathcal{A}, \vartheta, k)$ . Let's compute

$$\arg \max_k p(k|\mathcal{A}) = \hat{k}. \quad (9.117)$$

We know

$$p(\mathcal{A}|\vartheta) \quad (9.118)$$

from  $\mathcal{A} = f + \eta$  and  $p(\eta)$ .

$$p(k|\mathcal{A}) = \int d\vartheta d\eta p(k|\mathcal{A}, \vartheta, \eta)p(\vartheta)p(\eta) \quad (9.119)$$

$$p(\mathcal{A}|k) = \int d\vartheta \underbrace{p(\mathcal{A}|\vartheta)}_{\text{likelihood}} \underbrace{p(\vartheta|k)}_{\text{prior}} \quad (9.120)$$

$$= \int d^k \vartheta \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi\sigma^2}^N} p(\vartheta|k) \quad (9.121)$$

Say the integrand is sharply peaked, i.e.  $x^2 = (y - \vartheta x)^2$ .

$$\delta(x) = \lim_{\sigma \rightarrow 0} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi\sigma^2}} \quad (9.122)$$

$$p(\mathcal{A}|\vartheta) = \int d^k \vartheta \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi\sigma^2}^N} p(\vartheta|k) \quad (9.123)$$

We use the Taylor series to simplify the integral

$$x^2 \approx \underbrace{\hat{x}^2}_{0} + \vartheta^T \cdot \nabla \underbrace{x^2}_{0} + \frac{1}{2} \vartheta^T (\nabla \nabla x^2) \vartheta \quad (9.124)$$

where we have recalled a similar problem  $\int e^{-g(x)} dx$  using

$$g(x) \approx g(x_*) + (x - x_*) \underbrace{g'(x_*)}_{0} + \frac{1}{2}(x - x_*)^2 g''(x_*) \quad (9.125)$$

and substituted  $g = x^2$ .

$$p(\mathcal{A}|k) = \int d^k \vartheta \frac{e^{-\frac{1}{2}\hat{x}^2} e^{-\frac{1}{2}\vartheta^T (\nabla \nabla x^2) \vartheta}}{Z} p(\vartheta|k) \quad (9.126)$$

- HW 3: Show

$$p(\mathcal{A}|k) \approx \frac{p(\hat{\vartheta}|k)}{Z} \frac{e^{-\frac{1}{2}\hat{x}^2}}{\sqrt{|\nabla \nabla x^2|}} \quad (9.127)$$

## 9.4 Lecture 20

### 9.4.1 HW questions- Linear Regression

Say you are fitting data.

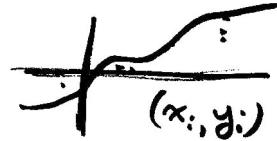


Figure 9.13: data

Look for a polynomial fit:

$$y_i = c_0 + c_1 x + c_2 x^2 \quad (9.128)$$

$$= \sum_{j=1}^K \vartheta_j \Phi_j(x_i) = \sum_{j=1}^K H_{ij} \vartheta_j = (H\vartheta)_i \quad (9.129)$$

where  $\Phi_j : X \rightarrow \mathbb{R}$ ,  $y \in \mathbb{R}$ , and the matrix  $(H_{ij} = \Phi_j(x_i)) \in \mathbb{R}^{K \times N}$ . For example,  $\Phi_1 = x^0$ ,  $\Phi_2 = x^1$ .

For the previously given paper,

$$\mathcal{A}_i = \sum_{\mu=1}^K \mathcal{N}_i^\mu \vartheta_\mu \quad (9.130)$$

Back to our case.

$$y_i = \sum_{j=1}^K H_{ij} \vartheta_j \Rightarrow \mathbf{y} = H\vartheta \quad (9.131)$$

Now, we continue to discuss fitting. We want to maximize the probability of the data given the unknowns.

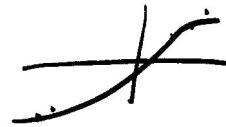


Figure 9.14: data

$$p(\mathbf{y}|\vartheta) \quad (9.132)$$

$$\vartheta \in \mathbb{R}^K, \quad (\text{unknowns}) \quad (9.133)$$

$$\mathbf{y} \in \mathbb{R}^N, \quad (\text{examples}) \quad (9.134)$$

It's easiest to think of the values of  $y$  as a vector  $\mathbf{y} = H\vartheta + \eta$ . So, we are looking for  $p(\mathbf{y}|\vartheta, \eta)$  which will be written  $p(\eta)$ .

$$p(\mathbf{y}|\vartheta) = \int d\eta p(\mathbf{y}|\vartheta, \eta)p(\eta) \quad (9.135)$$

$$p(\eta) = \mathcal{G}(\eta; 0, \sigma) \quad (9.136)$$

$$p(\mathbf{y}|\vartheta, \eta) = \delta(\mathbf{y} - (H\vartheta\eta)) \quad (9.137)$$

$$p(\mathbf{y}|\vartheta) = \prod_{i=1}^N \mathcal{G}(y_i - (H\vartheta)_i) \quad (9.138)$$

$$= \frac{e^{-\chi^2}}{(\sqrt{2\pi\sigma^2})^N} \quad (9.139)$$

$$\hat{\vartheta}_{ML} = \arg \max_{\vartheta} p(\mathbf{y}|\vartheta) = \arg \min_{\vartheta} \chi^2 \quad (9.140)$$

The last equation has a name. Recall Bayes' rule

$$p(a|b) = \frac{p(b|a)p(a)}{p(b)} \quad (9.141)$$

$$p(a|b) = p(a|b)p(b) \quad (9.142)$$

$$= p(b|a)p(a) \quad (9.143)$$

In statistics, probabilities have names.

$$\underbrace{p(\vartheta|y)}_{\text{posterior}} = \frac{\overbrace{p(y|\vartheta)}^{\text{likelihood}} \overbrace{p(\vartheta)}^{\text{prior}}}{\underbrace{p(y)}_{\text{evidence}}} \quad (9.144)$$

This becomes controversial since the “prior” is on the right hand side. This is the Bayesian way. So minimizing error is the same as the maximum likelihood of  $\vartheta$ . We fit because we believe in Gaussian additive noise.

Problems arise when the matrix is not invertible.

$$\chi^2 = \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - (H\vartheta)_i)^2 = \frac{1}{2\sigma^2} (\mathbf{y} - H\vartheta)^T \vartheta \quad (9.145)$$

$$\hat{\vartheta}_{ML} = \arg \min_{\vartheta} -2\mathbf{y}^T H\vartheta + \vartheta^T \underbrace{H^T H}_{\nabla \nabla \chi^2} \vartheta = \chi^2 \quad (9.146)$$

The homework was to show the minimum obeys

$$H^T H\vartheta = \mathbf{y}^T H \quad (9.147)$$

$$\vartheta = (H^T H)^{-1} \mathbf{y}^T H \quad (9.148)$$

$$H^T H = \sum_{i=1}^N \Phi_K(x_i) \Phi_{K'}(x_i) \quad (9.149)$$

We also compute

$$\hat{\vartheta}_{MAP} = \arg \max_{\vartheta} p(\mathbf{y}|\vartheta) p(\vartheta) \frac{e^{\frac{1}{2}\vartheta^T L^{-2}\vartheta}}{\sqrt{|2\pi L^2|}} \quad (9.150)$$

So that

$$\hat{\vartheta}_{ML} = \arg \min_{\vartheta} \chi^2 + \frac{1}{2}\vartheta^T L^{-2}\vartheta \quad (9.151)$$

In  $\hat{\vartheta}_{MAP}$ ,

$$p(\mathbf{y}|\vartheta) p(\vartheta) = \frac{e^{-\chi^2}}{\left(\sqrt{2\pi\sigma^2}\right)^N} \frac{e^{-\frac{1}{2}\vartheta^T L^{-2}\vartheta}}{\sqrt{|2\pi L^2|}} \quad (9.152)$$

The equations are

$$(H^T H + L^{-2}) \vartheta = \mathbf{y}^T H \quad (9.153)$$

where we recall,  $H^T H = \sum_{i=1}^N \Phi_{K_1}(x_i) \Phi_{K_2}(x_i)$ , and this term dominates for large data sets. So the prior is not so important compared to the likelihood. The matrix on the left hand side is now invertible. This is called Bayesian regularization or ridge regression.

Furthermore, the Bayesian approach can work with complexity explicitly.

$$p(y|k) = \int d\vartheta \underbrace{p(y|\vartheta, k)}_{\frac{e^{-\frac{1}{2}\chi^2}}{\left(\sqrt{2\pi\sigma^2}\right)^N}} p(\vartheta) \quad (9.154)$$

We want to integrate using a Gaussian integral. Take a case with one datum.

$$\chi^2 = (a - x\vartheta)^2 \quad (9.155)$$

$$\int d\vartheta \frac{e^{-\chi^2}}{Z} \equiv \int d\vartheta \frac{e^{-(a-x\vartheta)^2}}{Z} = \sqrt{2\pi} \frac{e^{-\hat{\chi}^2}}{\sqrt{2\pi/x^2}} \quad (9.156)$$

We can do a similar thing in our case.

$$p(\mathbf{y}|k) = \int d^K \vartheta \frac{e^{-\frac{1}{2}\chi^2}}{\left(\sqrt{2\pi\sigma^2}\right)^N} \sim p(\vartheta) = \int d^K \vartheta p(\vartheta) e^{-g(\vartheta)} \quad (9.157)$$

$$g(\vartheta) \approx g(\vartheta_*) + \underbrace{\vartheta_* q'(\vartheta_*)}_{0} + \frac{1}{2} \vartheta_*^T g''(\vartheta_*) \vartheta_* + \dots \quad (9.158)$$

$$= \int d^K \vartheta p(\vartheta) e^{-\frac{1}{2}\hat{\chi}^2 - \vartheta^T (\nabla \nabla \chi^2) \vartheta + \dots} \quad (9.159)$$

Approximate at dominant values.

$$p(\mathbf{y}|k) = p(\hat{\vartheta}) e^{-\frac{1}{2}\hat{\chi}^2} \sqrt{\left| \frac{2\pi}{\nabla\nabla\chi^2} \right|} \quad (9.160)$$

Or written as,

$$-\ln p(\mathbf{y}|k) = -\ln p(\hat{\vartheta}) + \frac{1}{2}\hat{\chi}^2 + \frac{1}{2} \ln \left| \frac{\sum_{i=1}^N \Phi_{K_1}(x_i)\Phi_{K_2}(x_i)}{2\pi} \right| \quad (9.161)$$

Now,

$$|H^T H| = |NC| \quad (9.162)$$

$$C = \frac{1}{N} \sum_{i=1}^N \Phi_{K_1} \Phi_{K_2} \quad (9.163)$$

We can estimate scaling of the sum  $C$  by the determinant. For example,

$$\left| \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \right| = \alpha^2 \quad (9.164)$$

$$\left| \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{pmatrix} \right| = \alpha^3. \quad (9.165)$$

So we have

$$-\ln p(\mathbf{y}|k) = -\ln p(\hat{\vartheta}) + \frac{1}{2}\hat{\chi}^2 + \frac{1}{2} \ln \frac{N^K}{Z} \quad (9.166)$$

To compare models of different complexities we use the statistic

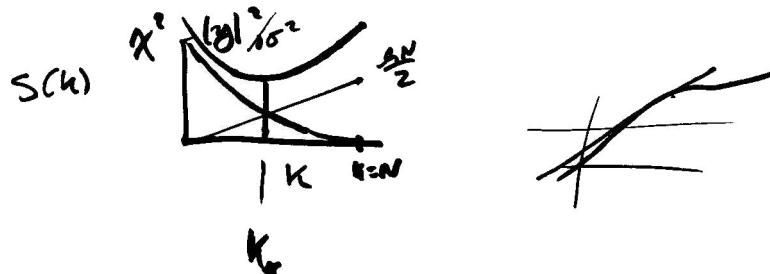
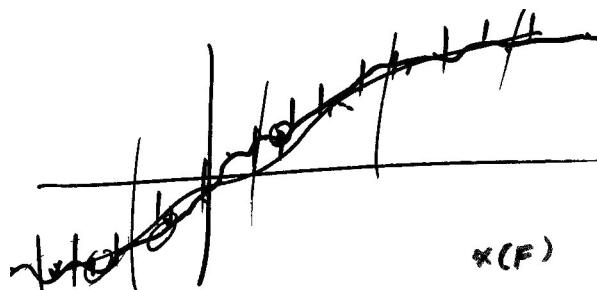
$$S(K) = \frac{1}{2}\chi^2 + \frac{K}{2} \ln N \quad (9.167)$$

Let's plot the two terms in this function and the function itself.

The importance of this statistic is that the value of  $K$  giving the minimum of  $S(K)$  is the desired complexity for the model. This minimum is the balance between fidelity and simplicity of the model.

Let's review our assumptions:

- Noise is normally distributed
- Noise is additive distributed
- We know the value of  $\sigma$ , giving error bars.

Figure 9.15:  $S(K)$ Figure 9.16: data fits for  $x(F)$ 

Recall our goal. We have data points and can fit a high order polynomial to it passing through all data points.

What is the problem with this? We said it has bad predictive power because it is too complicated.

To see how well this model predicts we can leave out a point and then do the best fit. Then we see how well the fit predicts the missing point. We must do this for each point.

Pseudocode for this test (cross-validation):

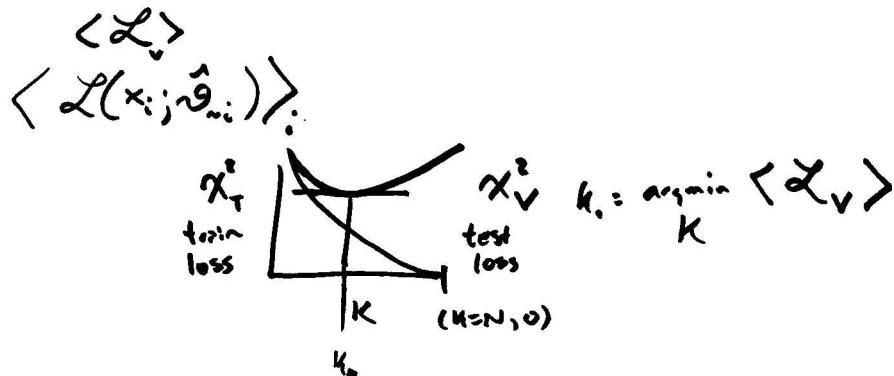
$$\langle \mathcal{L}_v \rangle, \quad \text{Learn on validation data} \quad (9.168)$$

$$\langle \mathcal{L}(x_i, \hat{\vartheta}_i) \rangle_i, \quad \text{Fit without the } i^{\text{th}} \text{ data point} \quad (9.169)$$

But we do this in many different ways. For instance we can omit, say, one-fifth of the points at once, and fit with the remaining points (5-fold cross-validation).

A Bayesian would say an error in choosing basis functions comes at the point of approximation:

$$|H^T H| = |\nabla \nabla \chi^2| = CN^K \quad (9.170)$$

Figure 9.17:  $\chi^2$ 

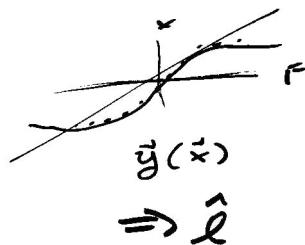
- HW 1: Play with the MATLAB cross-validation code
- HW 2: Read Gideon Schwartz's paper.

## 9.5 Lecture 21

### 9.5.1 Modeling

The purpose of modeling is to predict, explain, reproduce.

We try to form a model which fits data.

Figure 9.18:  $y(x) \Rightarrow \hat{l}$ 

To predict means to demand  $\chi^2_{test}$  is small. To explain means the data is interpretable,  $\vartheta_\mu = 0$  for "most"  $\mu$ 's. To predict means to demand  $\chi^2_{train}$  is small.

We want to understand these motifs.

Recall the DNA motif with major and minor grooves.

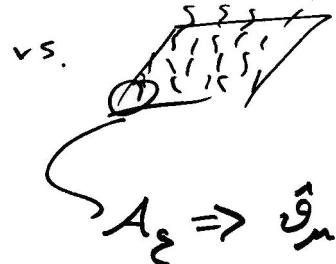
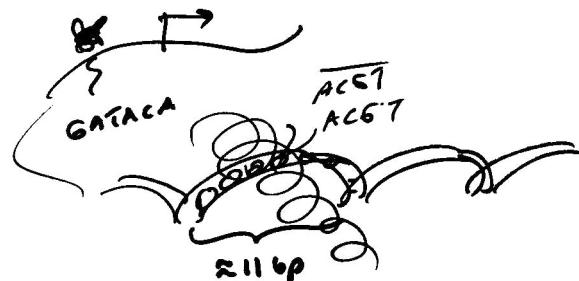
Figure 9.19:  $\mathcal{A}_g \Rightarrow \hat{\vartheta}_\mu$ ,  $\mathcal{A}_g = \sum_\mu \mathcal{N}_g^\mu \vartheta_\mu$ 

Figure 9.20: DNA sketch

We want to solve the regression problem:

$$\mathcal{A}_g = \sum_\mu \mathcal{N}_g^\mu \vartheta_\mu. \quad (9.171)$$

What is  $k$ ? Depends on organism and data. Yeast has about 6,000 genes. Lambda-phage has about 50. Say we're dealing with yeast. How do we discover the motifs in the DNA? What about a 7-mer?

$$k = \frac{4^7}{2} = \frac{2^{14}}{2} = 2^{13} = 2^{10} \cdot 8 \quad (9.172)$$

$$= 1024 \cdot 7 \approx 8000 \quad (9.173)$$

So we have more motifs than genes. This problem is ill-posed. Hence we do not have an invertible matrix to find the unknowns. We can regularize:

$$\min \chi^2 + l^{-2} \sum \vartheta^2 \quad (9.174)$$

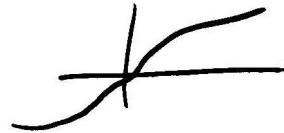


Figure 9.21: linear regression

Then we might need to compute

$$\min \chi^2 + \alpha \int dx (\nabla f)^2 \quad (9.175)$$

In the reduce paper:

$$\mathcal{A}_g = \sum_{\mu} \mathcal{N}_g^{\mu} \vartheta_{\mu} \quad (9.176)$$

$$\chi^2 = \sum_g (\mathcal{A}_g - \sum_{\mu} \mathcal{N}_g^{\mu} \vartheta_{\mu})^2 \quad (9.177)$$

We minimize as follows.

$$\hat{\mathbf{n}}_g = \mathbf{g} = \frac{\mathcal{N} - \langle \mathcal{N} \rangle}{\sigma_g} \quad (9.178)$$

- HW: Show that for robust regression, the motif  $\mu$  chosen is motif such that  $(\mathbf{n}_{\mu}, \mathbf{a})$  is maximized

Now we are trying to explain the data by looking over many motifs to minimize  $\chi^2$ .



Figure 9.22: DNA

Methods:

1. Bayesian information criterion. This is just like polynomial regression. What degree polynomial should one use?

$$p(\mathcal{A}|k) \quad (9.179)$$

$$\hat{\chi}^2 + \frac{k}{2} \ln N \quad (9.180)$$

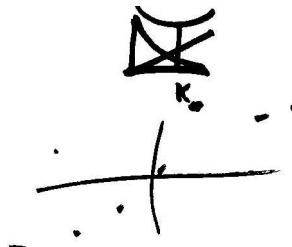


Figure 9.23: BIC data



Figure 9.24: Cross-validation error

2. Cross-validation
3. “p-values” Let’s plot the correlation between the vector of expressions and vector of occurrences of that motif. We want to minimize the complexity of the model.

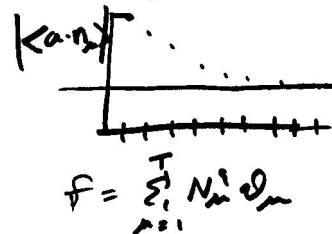


Figure 9.25: p-values plot

Different communities have different ideas of how simple is desired.

Let’s use robust regression which adds nonzero data one by one.

Can also take the vector of expressions and randomize over genes.

$$\mathcal{A}_g = \sum_{\mu} \mathcal{N}_g^{\mu} \vartheta_{\mu} \quad (9.181)$$

In MATLAB pseudocode, the matrices could be:

$$A = A(\text{randperm}(G)) \quad (9.182)$$



Figure 9.26: robust regression p-value plot

$$N = N(\text{randperm}(K), ;) \quad (9.183)$$

$$N = N(:, \text{randperm}(G)) \quad (9.184)$$

$$N = N(\text{randperm}(K); \text{randperm}(G)) \quad (9.185)$$

So the question is what is the null model? This is not clear in biology. Thus one must be careful in evaluating the significance of a quoted p-value.

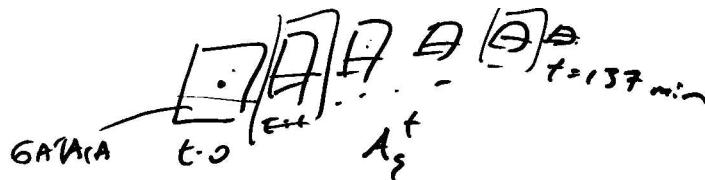


Figure 9.27: microarray data at different times

So remember the three ideas in complexity control and not overfitting data.

### 1. Bayesian information criterion

Recall we needed

$$y = f + \eta \quad (9.186)$$

$$p(\eta) \quad (9.187)$$

$$|\nabla \nabla \xi^2| = ZN^K \quad (9.188)$$

but nature does not supply this.

### 2. Cross-validation

This method has problems too. The matrix looks like:

$$y^i = \sum_{j=1}^4 \phi_j^i \vartheta_j \quad (9.189)$$

### 3. p-valueology

The problem here is the arbitrariness of choosing a threshold p-value.



Figure 9.28: CV matrix

The moral is that there are many ways to make mistakes.

We have discussed mostly linear regression which works well for linear models and less well for nonlinear models.

$$y^i = \sum_{j=1}^4 \phi_j^i \vartheta_j \quad (9.190)$$

BIC:  $p(\mathcal{A}|k)$

In p-valueology we have the genes:

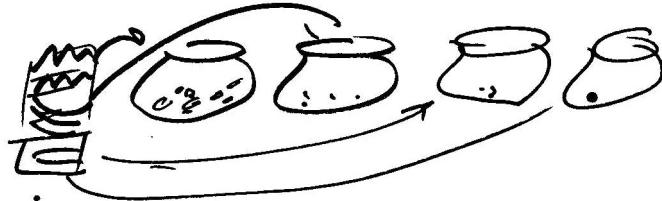


Figure 9.29: genes in buckets

Think about cases when you want to predict

$$\{1, 0\} = y = f(x) = \sum \mathcal{N}^\mu \vartheta_\mu \quad (9.191)$$

$$= f(\mathcal{A}) \quad (9.192)$$

Consider an example. You measure the height of people in a room. Then you classify the genders of the people.

How do you measure the accuracy of your prediction? What is the best value of  $\vartheta$ ?

What if you have  $p(g|h)$ ?

$$y = f + \eta \quad (9.193)$$

$$p(\eta) = \mathcal{G}(\eta; 0; \sigma) \quad (9.194)$$

$$p(h|b) = \mathcal{G}(h; \mu_b; \sigma_b) \quad (9.195)$$

$$p(h|g) = \mathcal{G}(h; \mu_g; \sigma_g) \quad (9.196)$$

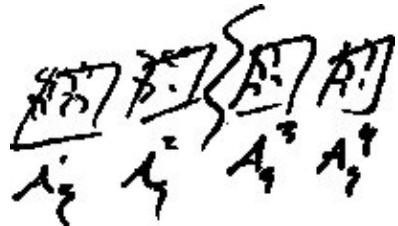


Figure 9.30: microarrays

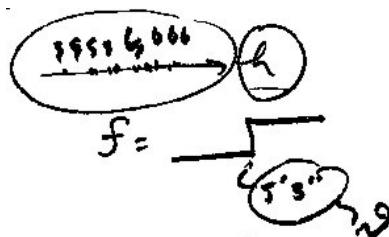
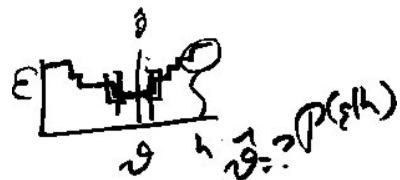


Figure 9.31: gender and height sketch

Figure 9.32: Error versus  $\vartheta$ 

$$\hat{\vartheta}_{ML} = \max_{\vartheta} p(\mathcal{A}|\vartheta) \quad (9.197)$$

$$(\hat{\mu}, \hat{\sigma})_{ML} = \max_{\mu, \sigma} p(h_{boys}|\mu, \sigma) \quad (9.198)$$

- HW 2: Show  $\hat{\mu}_{ML}$  for a Gaussian.

How do you build a classifier? Which variables are involved?

We are confronted with the curse of dimensionality. How do we solve the problem?

Create a function which divides into minus and plus classes.

$$F(\mathbf{x}) \quad (9.199)$$

Use data  $\{x_i, y_i\}$ . This is contrasted with linear regression:

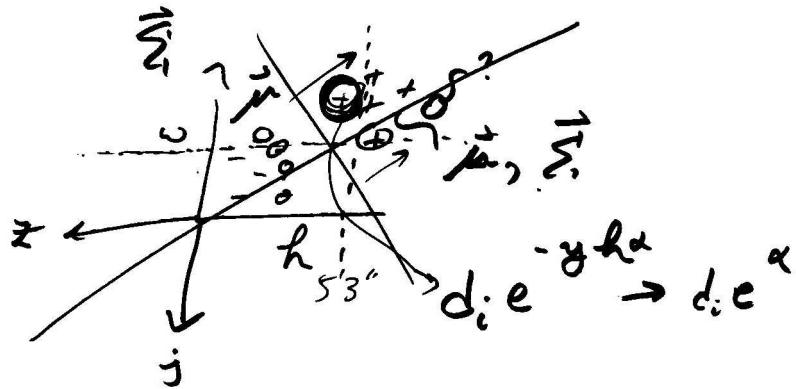


Figure 9.33: Boy/girl data

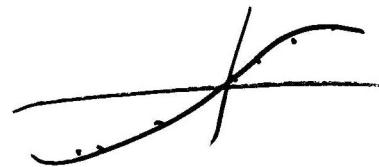


Figure 9.34: Linear regression data

Here is a motivating example.

$$\mathcal{Z} = \langle e^{-yF} \rangle_y = \sum_{y=\pm} e^{-yF} p(y) \quad (9.200)$$

$$\mathcal{Z}(F) = e^{-F} p(+) + e^F p(-) \quad (9.201)$$

$$0 = \mathcal{Z}' = -e^{-F} p(+) + e^F p(-) \quad (9.202)$$

$$e^{2F} = \frac{p(+)}{p(-)} \Rightarrow F = \frac{1}{2} \ln \frac{p(+)}{p(-)} \quad (9.203)$$

$$\mathcal{Z} \approx \sum_i e^{-y_i F(x_i)} = \sum_i 2^{-y_i \sum_{k=1}^T \vartheta_k \phi_k(x_i)} \quad (9.204)$$

$$y = f = \sum_i \vartheta_k \Phi_k(x_i) \quad (9.205)$$

- HW 3: Solve for best  $\alpha$ :  $\sum_i \exp \left( -y_i \left[ \sum_{n=1}^T \vartheta_n \phi_n(x_i) + \alpha h_k(x_i) \right] \right)$ , with
  - $\sum_i d_i \exp (-y_i \alpha h_k(x_i))$
  - $y_i = \{-1, 1\}$
  - $h_k(x_i) = \{-1, 1\}$

This is an algorithm for learning the general classifier from a set of basis functions which are themselves little classifiers. This is a technique called boosting and is very good for high dimensional systems.

Another method is called Support Vector Machines.

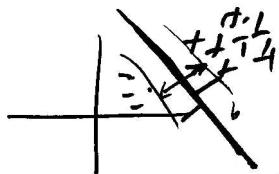


Figure 9.35: Support Vector Machines

Consider the model

$$\chi^2 + \alpha \sum \vartheta^2 \quad (9.206)$$

$$f = \sum \vartheta \Phi \quad (9.207)$$

We want to minimize the error

$$\min \sum \vartheta^2 \quad (9.208)$$

such that

$$y_i f_i > 1 - b \quad \forall i \quad (9.209)$$

then

$$\chi^2 = (y - f)^2 = y^2 + f^2 - 2yf \quad (9.210)$$

## 9.6 Lecture 22 - Final review

### 9.6.1 Schrodinger's cat

Lack of determinism in quantum mechanics. If a photon is emitted some radioactive material decays and kills the cat.

But you don't know what state the photon inhabits. Just as the photon is in a superposition state the cat is in a superposition of alive and dead states.

### 9.6.2 Van Kampen's (sp??) cat

By analogy with the previous example, imagine a virus which injects its DNA into E. Coli. There may be 3 copies of protein. They may be transcriptional regulatory proteins, which, in abundance, prevents the expression of a gene on the viral DNA. Otherwise, viral DNA encodes for another transcriptional factor which blocks appearance of the first transcription factor. So the state of the E. Coli. is best given by a probability distribution  $p(n)$ .

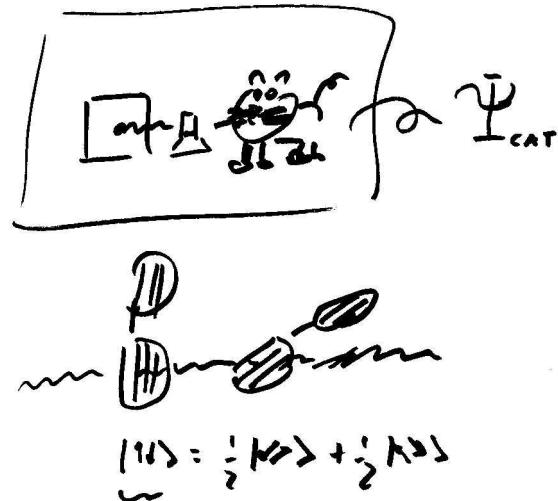


Figure 9.36: Schrodinger's cat

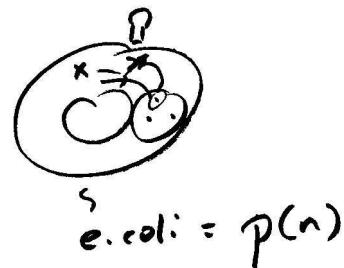


Figure 9.37: E. Coli.

### 9.6.3 Maximum likelihood

$$\hat{\vartheta}_{Ml} = (H^T H)^{-1} H^T \mathbf{y} \chi^2 \quad (9.211)$$

Say you have

$$U(\mathbf{v}) = \frac{1}{2} \mathbf{v}^T M \mathbf{v} \quad \text{cf. } U(x) = \frac{1}{2} kx^2, U' = kx \quad (9.212)$$

$$= \frac{1}{2} \sum_{i,j} v_i M_{ij} v_j \quad (9.213)$$

We have the formula

$$\frac{\partial v_i}{\partial v_k} = \delta_{i,j} = \begin{cases} 0 & i \neq k \\ 1 & i = k \end{cases} \quad (9.214)$$

Now take the derivative

$$\partial_{v_k} U = \frac{1}{2} \sum_{ij} [\delta_{i,k} M_{ij} v_j + v_i M_{ij} \delta_{j,k}] \quad (9.215)$$

$$= \frac{1}{2} \sum_j M_{kj} v_j + \underbrace{\frac{1}{2} \sum_i v_i M_{ik}}_{\frac{1}{2} \sum_{ij} v_i M_{ij} \delta_{j,k}} \quad (9.216)$$

$$= \frac{1}{2} \hat{\mathbf{e}}_k^T M \mathbf{v} + \underbrace{\frac{1}{2} \mathbf{v}^T M \hat{\mathbf{e}}_k}_{\hat{\mathbf{e}}_k^T M^T \mathbf{v}} \quad (9.217)$$

$$= \frac{1}{2} \hat{\mathbf{e}}_k^T (M + M^T) \mathbf{v} \quad (9.218)$$

For example

$$U = \frac{1}{2} kx^2 \quad (9.219)$$

$$U' = kx = \frac{1}{2}(k+k)x. \quad (9.220)$$

A simpler example is

$$U = \mathbf{v}^T \cdot \mathbf{b} = \sum_i v_i b_i \quad (9.221)$$

$$\partial_{v_k} U = \sum_i \delta_{i,k} b_i = b_k = \hat{\mathbf{e}}^T \mathbf{b} \quad (9.222)$$

$$\partial_{v_k} \frac{1}{2} \mathbf{v}^T M \mathbf{v} = \frac{1}{2} \hat{\mathbf{e}}^T (M + M^T) \mathbf{v} \quad (9.223)$$

Now what are the implications for  $\chi^2$ ?

$$\chi^2 = \frac{1}{2\sigma^2} \sum_i^N (y_i - f(x_i))^2 \quad (9.224)$$

$$f(x_i) = \sum_m^K \vartheta_m \Phi_m(x_i) = \sum_m H_{im} \vartheta_m = H\vartheta \quad (9.225)$$

$$\chi^2 = \frac{1}{2\sigma^2} \sum_i^N (y_i - \hat{\mathbf{e}}_i^T H \vartheta)^2 = \frac{1}{2\sigma^2} (\mathbf{y} - H\vartheta)^T \cdot (\mathbf{y} - H\vartheta) \quad (9.226)$$

$$2\sigma^2 \chi^2 = \underbrace{\mathbf{y}^2}_0 - \mathbf{y}^T H \vartheta - \underbrace{\vartheta^T H^T \mathbf{y}}_{\mathbf{y}^T H \vartheta} + \vartheta^T H^T \vartheta \quad (9.227)$$

where we have used  $(H\vartheta)^T = \vartheta^T H^T$ . We find

$$\sigma^2 \chi^2 = C + \frac{1}{2} \vartheta^T M \vartheta - \underbrace{\vartheta^T H^T \mathbf{y}}_{\vartheta^T H^T \mathbf{y}} \quad (9.228)$$

where  $M = H^T H$  has all positive eigenvalues because it is symmetric,  $(H^T H)^T = H^T H$ . Now differentiate

$$\partial_{v_k} \sigma^2 \chi^2 = \frac{1}{2} (M + M^T) \vartheta - H^T \mathbf{y} = 0 \quad (9.229)$$

By symmetry  $\frac{1}{2}(M + M^T) = \frac{1}{2}(H^T H + H^T H) = H^T H$ . Now minimizing yields

$$\hat{\vartheta} = (H^T H)^{-1} H^T \mathbf{y} \quad (9.230)$$

$$S(k) = \chi^2 + \frac{k}{2} \ln N \quad (9.231)$$

$$\int d^k \vartheta e^{-\frac{1}{2} \chi^2} \approx e^{-\frac{1}{2} \hat{\chi}^2} \sqrt{\left| \frac{2\pi}{\nabla \nabla \chi^2} \right|} \quad (9.232)$$

$$\chi = lN \tanh \left( \frac{Fl}{k_B T} \right) \quad (9.233)$$

$$\mathcal{A} = \sum_{\mu} \vartheta_{\mu} \mathcal{N}_g^{\mu} \quad (9.234)$$

#### 9.6.4 New material

Recall we looked at microarray data and did statistics. First we did regression:

$$\mathcal{A}_g = \sum_{\mu} \vartheta_{\mu} \mathcal{N}_g^{\mu} \quad (9.235)$$

$$y = f(\vartheta, x) \quad (9.236)$$

Then we talked about classification.

$$f(\vartheta, x_i) = \sum_t \vartheta_t h_t(x_i) \quad (9.237)$$

$$= \frac{1}{2} \ln \frac{p(+|\mathbf{x})}{p(-|\mathbf{x})} \quad (9.238)$$

We can choose the complexity of the model.

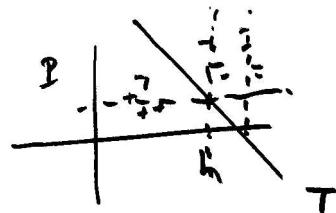


Figure 9.38: plus/minus data



Figure 9.39: classification parameters

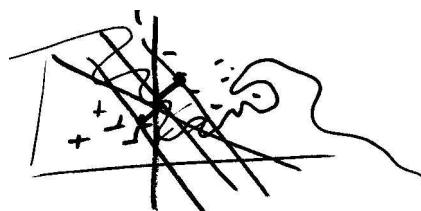


Figure 9.40: Dividing data into two classes

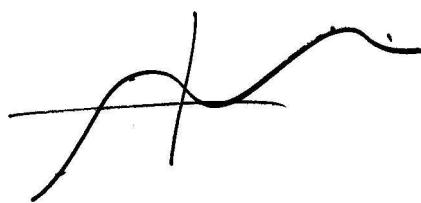


Figure 9.41: nice data fit

Up until now we have used “nice math.”

$$y = f + \eta \quad (9.239)$$

$$p(\eta) = \mathcal{G}(\eta) \quad (9.240)$$

We classify the data using parameters. We prefer a simple classification.

We also want a nice fit with data.

Now let's consider a more difficult case. We don't calculate probability but rather bound how big the probability could be (large deviation theory).

$$\rho(\mathbf{x}_i, y = \pm 1) \quad (9.241)$$

Given a positive variable  $x > 0$  empirical estimate for the mean  $\mu = \langle x \rangle = \int_0^\infty dx x p(x)$ . We can say  $p(x > \rho\mu) < Q$ . Let's see this.

$$p(x > \rho\mu) = \int_{\rho\mu}^\infty dx \rho(x) \equiv \int_0^\infty dx H(x - \rho\mu)\rho(x) \quad (9.242)$$

where  $H$  is the Heaviside function given in the figure.

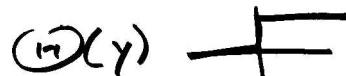


Figure 9.42: Heaviside function

So let's plot this function:

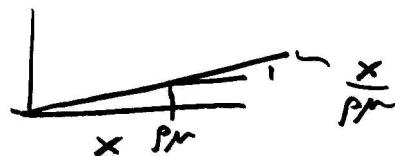


Figure 9.43: probability plot

$$p(x > \rho\mu) = \int_0^\infty dx \rho(x) H(x - \rho\mu) < \int_0^\infty dx \rho(x) \frac{x}{\rho\mu} = \frac{\mu}{\rho\mu} \quad (9.243)$$

$$p(x > \rho\mu) < \frac{1}{\rho} \quad (9.244)$$

If  $\rho < 1$  this bound is trivial. This is one example of the use of bounds.

Recall measuring the heights of boys and girls. We tabulate data as in the figure.

Say you didn't record genders but just the heights as in the figure.

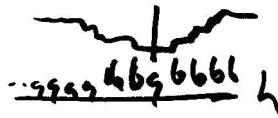


Figure 9.44: boy and girl height data



Figure 9.45: height plots without gender

We have supervised learning such as regression or classification, using  $y = f(x)$ . Or there is unsupervised learning or clustering using  $x_i \rightarrow z(i)$ . Without class labels it is difficult to say whether model is improving.

Say you believe the probabilities obey

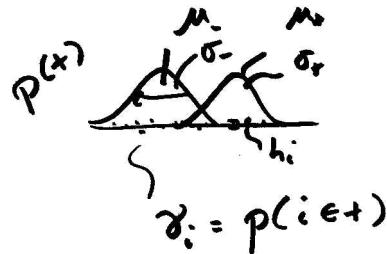


Figure 9.46: two peaked probability

$$p(h) = \sum_{\pm} \underbrace{p(h|\pm)}_{\mathcal{G}(h;\mu_{\pm},\sigma_{\pm})} p(\pm) \quad (9.245)$$

$$\gamma_i = p(i \in +) \quad (9.246)$$

Then you could find

$$(\widehat{\mu}_{\pm}, \widehat{\sigma}_{\pm}) = \arg \max_{\mu_{\pm}, \sigma_{\pm}} p(D|\mu_{\pm}, \sigma_{\pm}) \quad \text{Maximization} \quad (9.247)$$

$$\gamma_i = p(i \in +) \quad \text{Expectation} \quad (9.248)$$

This process can be repeated. In the end you get the mean of both classes and the probability that an individual is in either class. This is a way of doing clustering.

Another approach to clustering involves measuring the heights as in the figure. And clustering nearby groups. The choice here does not minimize any cost function. It is called

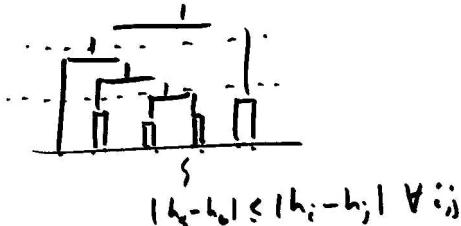


Figure 9.47: agglomerative hierarchical pairwise clustering

agglomerative hierarchical pairwise clustering.

### 9.6.5 Outline of this part of class

- supervised learning
  - (linear) regression
  - classification
    - \* boosting
    - \* SVM

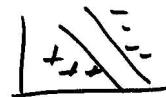


Figure 9.48: classification

The methods for supervised learning are called complexity control. Methods include

1. Cross validation
2. Bayesian information criterion

$$S(k) = \chi^2 + \frac{k}{2} \ln N \quad (9.249)$$

3. p-values

Today we learned

- unsupervised learning/ clustering

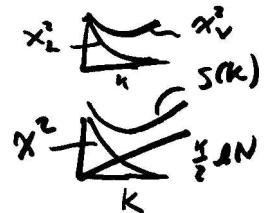


Figure 9.49: Bayesian information criterion

- generative

$$p(x) = \sum_z p(x|z) \underbrace{p(z)}_{\text{learn this}} \quad (9.250)$$

- non-generative e.g. dendrogramming  $d(x_i, x_j)$

These were methods to see whether you are overfitting data.



## Chapter 10

# Classification: modeling from data without data modeling

### 10.1 A story

Consider the optical character recognition (OCR) problem faced by the united states postal service (USPS) in the early 1990's. Armed with a corpus of XXX's of labeled images of the ten digits 0-9, they posed a computational challenge to the community: design an algorithm which can automatically detect the class (or digit) using only optical data.

One could imagine proposing a theory of threes and fours (*e.g.*, “threes are kind of round, with a whole on the left; but fours are kind of pointy”) Rather than methods which began with an *a priori* theory of ‘3’, or of ‘4’, etc, the successful approach has been that of *Machine Learning*, the design of a “machine” (here meaning an algorithm) which learns which are the discriminative rules and features from the labeled data. At this point ML-based OCR works with an error rate of approximately XX %, far below that of human-labeled data.

### 10.2 Data without the model: prototype models

handout: p416-419 of Tishby

### 10.3 Data modeling: classical classification

### 10.4 Feature selection: classical approaches

#### 10.4.1 Discriminability

aka ROC

#### **10.4.2 Nonparametric statistics**

monotonicity

### **10.5 The machine learning approach**

### **10.6 Large margin classification**

#### **10.6.1 SVMs**

Classification of cancer types, for example discrimination of leukemia classes in 2000 [1]

#### **10.6.2 Boosting**

### **10.7 Applications**

#### **10.7.1 Clinical**

#### **10.7.2 Network classification**

#### **10.7.3 Network inference**

# Chapter 11

# Clustering

## 11.1 Basic notions

### 11.1.1 History

In the bad old days....[?]

### 11.1.2 Why would you do this?

annotation

sequence information

lessons learned:  $X, Z, Y (Y_1, Y_2, \dots)$

Imagine someone walking into a classroom full of students, who records the height and gender

## 11.2 Orthodox clustering

handout: Eisen 1998

### 11.2.1 Generative

mixture models

### 11.2.2 Hierarchical

distance functions. the algorithm

**11.2.3 How many clusters?**

as a complexity control problem

as a side-information/cross validation problem

**11.3 Foresighting and context****11.3.1 Unorthodox (supervised) clustering****11.3.2 from side information to relevant information**

## Chapter 12

# Biological Information

### 12.1 What is information?

#### 12.1.1 History: Shannon's postulates

##### Homework:

Prove entropy is monotonically increasing for any solution to the diffusion equation, and that the steady state is necessarily a constant distribution

#### 12.1.2 Information is average dependence

The definition of statistical independence is a relation between the joint distribution  $p(x, y)$ , read as “the probability that  $X = x$  and  $Y = y$ ”, and the marginals  $p(x)$  and  $p(y)$ , namely

$$\text{“}X \text{ and } Y \text{ are independent} \text{”} \iff p(x, y) = p(x)p(y) \quad \forall x, y \quad (12.1)$$

We might then quantify how independent  $x$  and  $y$  are or, in other words, how informative  $x$  and  $y$  are about each other, by the quotient  $p(x, y)/p(x)p(y)$ . Of course this would only be for one value; we'd rather look at this quantity averaged over all values. In fact, since these are all distributions, and therefore positive, we could also quantify this by investigating the average of the logarithm of the quotient. This average quantifies how informative  $x$  and  $y$  are about each other and is called the “mutual information”

$$I[p(x, y)] = \int dx dy p(x, y) \ln p(x, y) / (p(x)p(y)) \quad (12.2)$$

between  $x$  and  $y$ . Note that you might see  $I(X, Y)$  as a shorthand, but it is not a function of  $X$  and  $Y$ ; rather it is a scalar-valued functional of the joint probability distribution (a function)  $p(x, y)$ .

To relate mutual information to the entropy  $-\langle \ln p(x) \rangle$ , consider performing an experiment in which you choose  $x$  and measure  $y$ . If we have performed the experiment many many times, we can infer from the results of the experiment  $p(y|x)$ , the probability

of observing a certain outcome given our choice of the input. Clearly we can not quote the entropy of  $y$  overall, merely the entropy as determined by sampling over some range of  $x$ . We relate this to the mutual information by noting

$$I = \langle \ln p(y|x)/p(x) \rangle \quad (12.3)$$

$$= \int dx p(x) \int dy p(y|x) \ln p(y|x) - \int dx p(x) \ln p(x) \quad (12.4)$$

$$= H[p(x)] - \langle H[p(y|x)] \rangle_x \quad (12.5)$$

emphasizing that the mutual information can be thought of as the average reduction in entropy in one quantity by knowing the value of another. In the simple case where we sample  $x$  evenly,  $p(x) = 1/L$  and  $I[p(y,x)] = \ln L - H[p(y|x)]$ .

What if for every  $x$  there is one and only one value of  $y = f(x)$ ? In the discrete case, we denote  $p(y|x) = \delta_{y,f(x)}$  and

$$I = H[p(x)] + \sum_x p(x) \sum_y p(y|x) \ln p(y|x) \quad (12.6)$$

$$= H[p(x)] + \sum_x p(x) \sum_y \delta_{y,f(x)} \ln \delta_{y,f(x)} \quad (12.7)$$

$$= H[p(x)] \quad (12.8)$$

since  $1 \ln 1 = 0 \ln 0 = 0$ .

Eqn. C.28 is nice because, if I now ask you to calculate the entropy of the  $d$ -dimensional Gaussian distribution, that is

$$-\int d^d x p \ln p = -\langle \ln p \rangle \quad (12.9)$$

that's what the entropy is: the average of the log of the probability.

$$\ln p = -\frac{1}{2}x^T \Sigma^{-2}x - \frac{1}{2} \ln |2\pi\Sigma^2| \quad (12.10)$$

and therefore the entropy  $H_d$  is

$$H_d = \frac{d}{2} \ln |2\pi\Sigma^2| \quad (12.11)$$

note the entropy goes up in higher dimensions and, that, since the determinant of a matrix is the product of the diagonal elements, for uncorrelated degrees of freedom,  $\sigma_{ij} = \sigma_{ii}\delta_{ij}$ .

$$H_d = d/2(1 + \ln 2\pi) + \sum_i^d \ln \sigma_{ii} \quad (12.12)$$

$$H_d/d = \frac{1}{2}(1 + \ln 2\pi) + \langle \ln \sigma_{ii} \rangle_i \quad (12.13)$$

so the entropy is simply the sum of the logarithm of the standard deviations.

Of course, this was all just preamble to the homework problem.

$$I(x_1, x_2) = \langle \ln p / pp \rangle \quad (12.14)$$

such that if  $x_1$  and  $x_2$  are statistically independent, then by definition the joint is the product of the marginals,  $p(x, y) = p(x)p(y)$ , and  $I = 0$ .

Note that we can rewrite this as

$$I = H(xy) - H(x) - H(y) \quad (12.15)$$

since these are all normal distributions, we can use the result we worked out for the d-dimensional Gaussian, and we get

$$I = \frac{1}{2} + \ln 2\pi\sigma_{11}^2 + \frac{1}{2} + \ln 2\pi\sigma_{22}^2 - 2/2 - \frac{1}{2} \ln \sqrt{2\pi|\Sigma|} \quad (12.16)$$

$$= -\frac{1}{2} \ln |\Sigma|^2 / \sigma_{11}^2 \sigma_{22}^2 \quad (12.17)$$

$$= -\frac{1}{2} \ln(\sigma_{11}^2 \sigma_{22}^2 - \sigma_{12}^2) / (\sigma_{11} \sigma_{22}) \quad (12.18)$$

$$= -\frac{1}{2} \ln 1 - \rho^2 \quad (12.19)$$

So we've just shown that, for a Gaussian distribution, the information is monotonically related to the square of the Pearson correlation coefficient. One interesting thing about it is that the PCC enters as the square, which means there's just as much information in talking to someone who always lies to you as there is in talking to someone who always tells you the truth. Watching a weatherman who is right 20% is just as useful as watching one who is right 80% of the time.

Later, you could replace  $y$  with some function  $z$ , e.g.,  $z = -y$  if you like to watch the bad weatherman, but you can't somehow gain information about  $x$  by replacing  $y$  with  $-y$ . At best, you'll have the same amount of information. And, in general, if we were to look at some function  $z(y)$ , or perhaps some stochastic function with a probabilistic assignment  $p(z|y)$ , you can't somehow gain information about  $x$  by transforming  $y$ . In information theory, one expresses this as an inequality

$$\text{if } p(z|y, x) = p(z|y), \text{ then } I(x, z) \leq \min\{I(x, y), I(y, z)\}. \quad (12.20)$$

Consider the game telephone, in which child  $A$  tells child  $C$  something, and child  $Z$  tells child  $y$  the message. The information in the chain is limited by whether

What on earth does this have to do with biology? In systems biology, it may be the case, for example, that  $X$  controls  $Y$  by acting as a clamp. When  $X$  is on  $Y$  is clamped to the 0 state, but when  $X$  is off,  $Y$  is free to fluctuate between -1 and 1. This configuration, in fact, has  $\rho = 0$  but has mutual information (homework). And people do this, for example, to establish one "edge" in a genetic network. Such a network is called a relevance network.

What about this telephone inequality? If you look at the mutual information between  $X$  and  $Z$ , and it turns out that  $X$  and  $Y$  have strong information, and  $X$  and  $Z$  have strong information, but  $Y$  and  $Z$  have some weaker information then you would say well, i can

explain these data simply by having X-Y and X-Z. I don't' need to draw an edge Z-X. And, in fact, this is exactly what is done in a technique called ARACNE by Andrea Califano for analyzing microarray data. At the end, you assign edges if I exceeds a threshold, and delete these edges, if they can be more simply assigned via the telephone inequality<sup>1</sup>

### 12.1.3 Re-parameterization-invariance

A second appeal of  $I$  over  $\rho$  is that we might not know in advance what is the most important thing to measure, and from it which statistics to compute. For example, say I measure the correlation between expressions  $X$  and  $Y$ . But if expressions are always positive, and if they seem to be log-normally distributed, perhaps we should calculate the correlation between  $\ln X$  and  $\ln Y$  instead. Or between  $X^2$  and  $Y^2$ . Which is "best"? Triangle inequality states that no subsequent stochastic transformation can increase the information, but we can also show that for any deterministic, monotonic transformation, the information is in fact invariant.

Consider replacing  $x$  with  $f(x)$  and  $y$  with  $g(y)$  (and the distribution  $p(x, y)$  with the distribution  $q(f, g)$ ). All one needs to know is how probability distributions change on introducing a change of variables. If you believe in calculus, then you must believe

$$1 = \int dx dp(x, y) = \int df dg q(f, g) \quad (12.21)$$

$$= \int dxdy \left( \frac{\partial_x f}{\partial_x g} \frac{\partial_y f}{\partial_y g} \right) q(f(x), g(y)) \quad (12.22)$$

$$= \int dxdy J(f(x), g(y)) q(f(x), g(y)) \quad (12.23)$$

So we must have  $p(x, y) = J(f(x), g(y))q(f(x), g(y))$ . For independent transformations,  $J$  is diagonal, and  $p = f_x g_y q$ , where subscripts indicate differentiation. Similarly, we find  $p(x) = f_x q(f)$  and  $p(y) = g_y q(g)$ . We then have

$$I[p(x, y)] = \int dxdy p(x, y) \ln \frac{p(x, y)}{p(x)p(y)} \quad (12.24)$$

$$= \int df dg q(f, g) \ln \frac{q(f, g)}{q(f)q(g)} \quad (12.25)$$

$$= I[q(f, g)] \quad (12.26)$$

This is not true, for example, of the entropy:

$$H[p(x)] = - \int dx p(x) \ln p(x) \quad (12.27)$$

$$= - \int df q(f) \ln q(f) - \int df q(f) \ln f_x \quad (12.28)$$

---

<sup>1</sup>This is sometimes called the data-processing inequality or erroneously referred to as triangle inequality. "Triangle inequality" more generally, refers to any inequality of the form  $d(a, c) \geq d(a, b) + d(b, c)$ . So the telephone inequality is one triangle inequality, where  $d(a, b) = I[p(a, b)]$  and  $p(c|a, b) = p(c|b)$ .

$$\neq H[q(x)] \quad (12.29)$$

but is true of the *relative entropy*, aka Kullback-Leibler divergence <sup>2</sup>

$$D_{KL}[p_1(x)||p_2(x)] \equiv \int dx p_1(x) \ln p_1(x)/p_2(x), \quad (12.30)$$

$$= \int df q_1(f) \ln q_2(f)/q_1(f), = D_{KL}[q_1(f)||q_2(f)] \quad (12.31)$$

### 12.1.4 $D_{KL}$ : how many questions must I ask?

#### $D_{KL}$ for Gaussian

for  $q = N(x; \alpha\sigma)$  and  $p = N(x; \sigma)$   $D = \int q \ln p/q = \frac{1}{2}(1 - \alpha^2) + \ln \alpha \leq 0$ .

#### $D_{KL}$ for constants

Note  $D_{KL}[p(x)||1/L] =$

#### $D_{KL}$ for marginals vs joints

And finally Note  $I[p(y, x)] \equiv D_{KL}[p(x, y)||p(x)p(y)]$

$$I[p(x, y)] = D_{AL}p(x)p(y)||p(x, y) \quad (12.32)$$

meaning, in short, *the mutual information between x and y is the number of binary questions one must ask to know that x and y are not independent.*

## 12.2 Density estimation

### 12.2.1 Maximum entropy distribution

max S implies exponential family.

I must take the derivative of this thing with respect to the probability  $p_i$  at some particular location  $i$  This yields

$$\ln p_i - p_i/p_i + \lambda_1 \dots \quad (12.33)$$

max H given  $\langle U_i \rangle$  implies Boltzmann, with  $T = T(\langle U \rangle)$  or  $\langle U \rangle = f(T)$ . max H given mean+variance implies Gaussian P(sum of Gaussian-distributed observations)= $\chi^2$  max H given mean implies exponential P(sum of exponential-distributed observations)=Poisson

---

<sup>2</sup>Erroneously sometimes referred to as the Kullback-Leiber distance – note that it does not obey the triangle inequality  $d(a, b) \geq d(a, c) + d(c, b)$  and is therefore not a distance

### 12.2.2 Complexity control

Data are abundant in neuroscience, which is why neuroscientists have a long tradition of exploiting information theoretic quantities.

(12.34)

So you can see why this is appealing. it involves no parameters, no assumptions about the underlying distributions.

#### Question:

what is a disadvantage

A:

it's hard to estimate it That is the big problem with Nature. Nature never hands you distributions, Nature only hands you observations. So there is an intermediate act of going from observation to distributions. Never believe someone who says "I measured X and Y, and the mutual information is .34 bits." this sentence is meaningless without telling you how they went from observation to distribution. If I hand you  $\rho$  there's no ambiguity. The average means average over observation – That is a "statistic" in addition to being an expectation. It's a statistic in the sense that I took a whole bunch of observations and, instead of telling you of expression for gene X and expression for gene Y and, instead of telling you all those numbers i just told you this statistic, as a summary of the data.

but if i knew the distribution I could tell you what i expected this value to be, but...

so one of these is a 'forward' observation (i hand you the distribution and ask you what you expect — a question with a unique answer) and the other is a 'reverse' (i hand you observations and ask you to find a good summary — a question with, potentially, multiple answers).

If i knew the distribution, it *might* turn out that the statistic is a useful observation to look at ; but if i didn't know the distribution might turn out that the statistics is an uninformative thing to look at, like RBI, which turns out not to tell you o much about the value of a single player.

So before I

### 12.2.3 Stat mech, revisited

#### Boltzmann as maximum entropy

#### Correlation vs information

$$\rho = \langle (x - \bar{x})(y - \bar{y}) \rangle / \sigma_x \sigma_y \quad (12.35)$$

(or sum)

**Homework:**

calculate the correlation function and the mutual information for the following cases

- $p(y = x) = z + \epsilon; p(y \neq x) = z - \epsilon$  where  $\{x, y\} \in -N, -(N-1), \dots, (N-1), N$ .
- $p(-1, 0) = \epsilon, p(1, \pm 1) = \epsilon/2$ .
- $p(x, y) = (1/Z) \exp(-\frac{1}{2}(x, y)\Sigma^{-2}(x, y)^T)$  for  $M$  an arbitrary symmetric, positive-definite matrix

## 12.3 The information bottleneck

Maximizing entropy, as we saw, seems to give you something for (almost) nothing. One of the applications in which this is particularly true is the application of information theoretic ideas, specifically “rate-distortion theory,” to supervised clustering.

we talked about clustering in the context of gene clustering, which was one of the earlier applications of clustering in modern “data driven” biology.

the result was, say, you would kick yeast in 20 different ways, and you would result in these huge tables of numbers. where the number is the expression of a gene

the first thing they did was to group genes together and for that they needed s clustering.

as a toy, we discussed how you would cluster heights. So imagine you walk in to a room and measure everyone’s heights. If you have some prior believe that the room is divided into boys and girls, you might group the data into two Gaussians.

I could improve on this by keeping some of the information that’s relevant to me, in this case gender, and attempting to define some simple function to model or represent  $X$ . E.g., i am looking for the simplest description of the data which preserves the information that’s relevant to me.

So say that you really believe d that boys and girls were drawn from some normal distribution. Then you might say well, i could maximize the probability of all the data. which is the sum over being girls+boys  $p(h|\pm)p(\pm)$ , choose  $p(h|\pm)$  to be a normal distribution with variances appropriate to each.  $p\pm$  is proportional...

This is an example of a parameterized distribution.

and you can do this, in fact, without the labels. You can iteratively refine who is assigned to which class, and reassign the parameters to maximize the likelihood. You will end up with ...

So this is an approach to clustering which is unsupervised. You threw away the labels but gained a very simple description. Now you can call your grandma and, instead of telling grandma all the observations you made, you can just say 30 percent were girls. Girls are about  $5' \pm .3$  and boys are  $5'6' \pm 6$ . You’ve taken all these data  $X$  and reduced them to set of parameters  $Z$ . You did this, though, without using the labels  $Y$ .

In fact, with a generative model, we could generate new data consistent with the distribution (we know  $p(h, \pm)$ ).

## **12.4 What is modeling?**

## **Part IV**

## **Coda**



## Chapter 13

# Summary

### Parts list

#### Cells

are  $1\mu$  in size, which is about the wavelength of light.

#### Cells have holes.

biology exploits things going in and out of holes to regulate many processes, including ATP production, rotary flagellar motors, etc.

#### Inside the cells are structural filaments

like microtubules and actin. These are used as the substrates for motor proteins, for supplying cell rigidity, and for shuttling chromosomes around during cell division via their polymerization and depolymerization.

1. these are *polymers*, meaning they are made out of *monomers*. In this case the monomers are globular proteins. Proteins, unlike vitamins or lipids, are made by the organism.
2. Other proteins act as motors, chewing ATP to change chemistry into physics.
3. Other proteins are transcription factors. Transcription factors bind to motifs in DNA near coding regions
4. Other proteins are light bulbs (like “GFP”: green fluorescent protein). These are discussed in class more for their biotechnological impact than for their biological impact.

## Proteins

These structural filaments are actually polymers, long chains of monomers, where each monomer is itself a distinct protein. Proteins are created via ‘translation’: mRNA is translated into a new language (in the 20-letter amino acid alphabet instead of the 4-letter nucleotide alphabet).

## Nucleic acids

You might ask where the mRNA came from. mRNA is a type of nucleic acid, a second family of polymers (in addition to polymers of proteins). These are not made in the same way protein is made. They include DNA and RNA.

DNA and RNA have their own “parts list,” an informative parts list

1. coding region is transcribed into mRNA
2. motifs are binding sites for the transcription factors

## 13.1 Physics at the scale of cell

### T

$$\kappa_B T = 4pnNM$$

## Diffusion

$$D = T/6\pi\eta a$$

## The Grant Unified Theory of Probability (GUTOP)

probability:  $x_t + 1 = xt + \delta$  leads to binomial, but also to Gaussian

## Diffusion with drift

$$p_t = -j_x = -dx(-Dr' + rv)$$

Special cases:

1.  $J = 0 \rightarrow$  Boltzmann
  - (a)  $U \propto \pm 1$
  - (b)  $U = x$
  - (c)  $U = x^2$
  - (d)  $U = \ln x$ : condensation
  - (e)  $U = e^{-r}/r$ : screening
2.  $J \neq 0 \rightarrow$  Kramers escape+motor proteins

## 13.2 Systems biology

### Chemical kinetics

1. MMK
2. hill K

### Transcriptional and translational dynamics

(Griffith models)

### Dynamical systems

1. 1D:  $f' > 0$
2. 2D:  $\tau - \Delta$  plane

### Stochastics

1.  $x \rightarrow 0$  and  $0 \rightarrow x \Rightarrow p_n(t) = -rnp_n + r(n+1)p_{n+1} - \alpha p_n + \alpha p_{n-1}$
2. GUTOP

probability:  $x_{t+1} = x_t + \delta$  leads to binomial, but also to Gaussian also leads to Poisson, binomial also leads to Poisson

### Networks

1.  $p(k) \propto k^{-\nu}$ : “scale free”
2.  $p(k) \propto z^k/k!$ : “Poisson”

## 13.3 Modeling

### Complexity control

1. CV
2. p-values
3. BIC

**Classifiers****Clustering**

1. orthodox clustering
2. unsupervised clustering
3. IB

$$\int d^d x \frac{e^{-x^T \Sigma^{-2} x}}{\sqrt{|2\pi\sigma^2|}} = 1$$

## Chapter 14

# Final thoughts: What is a model?

In this class, a **model** is a simplified description of the world which preserves some information deemed relevant.

A **good** model is sufficiently complex to preserve relevant information (e.g., agree well with data), but simple to generalize (doesn't overfit) and to be interpretable.

Establishing this balance is usually called "model selection." A more descriptive term might be "complexity control"

**Interpretability**

**Generalization**

**Simplicity**



## Appendix A

# Biology as told by a theoretical physicist

### A.1 What is biophysics?

In the abstract, I mention the revolution of biological physics, and discuss things that have happened in the last 10 years. So you would think biophysics is very very fresh. Very au courant. But in fact biophysics is a very old word (introduced in 1892 for aetiology, used in 1913 for artificial life, and in 1926 for lipid biophysics). And, in fact, people have been speculating for a very long time what benefits might be enjoyed when physics, which has done such a very good job of explaining some things, is applied to biology, which is very important to us, being living (and therefore dying) things. Early excursions of physicists into biology include Delbrück's phage group, Schrodinger's book "what is life," which itself motivated Watson to study biology. In fact, biophysicists sometimes point to the importance of DNA in biology and, owing to the physical training of its early pioneers, claim "physicists invented molecular biology." But you can decide for yourself if this statement says more about biology or more about physicists. (Of course, if you count Hooke as a physicist, then physicists invented cellular biology, too, but Hooke would have called himself a natural philosopher.)

The question then is: What took them so long?

Everyone on the street knows what biology is. Not everyone on the street really knows what physics is, although they probably have some keywords associated with it (like "bombs – they kill people" or "colliders – they are expensive and release radiation" or "string theory – that's so hot right now. String theory"). Any reader who has completed the core curriculum of an undergraduate major in physics will have her or his own view of what physics is. For most students, physics is: electricity and magnetism (EM), statistical mechanics (SM), quantum mechanics (QM), and classical mechanics (CM). These are the four pillars of physics. Quantum Mechanics describes things smaller than an atom (.1 nanometers). Classical mechanics describes planets, billiard balls, and inertial dynamics – things either standing still or moving in a vacuum. Statistical Mechanics is about things at

equilibrium. Let's talk about each of these in turn, both to set the historical context and the physical context of biophysics. What about molecules, proteins, nucleic acids, cells, genes, and generally things at a scale between 1 nanometer and 1 micron? This is much smaller than we are. In fact it is 6 to 9 “orders of magnitude” smaller than we are (don't worry if you don't know yet what that means). In terms of physics, this world is very different from the things in the four pillars. (i) Nothing moves like a billiard ball;  $F$  doesn't equal  $ma$ . (ii) Nothing is in equilibrium, or else it would be dead, and (iii) things are bigger than the quantum scale, for the most part. (Unless it's biochemistry. Also, just because there's no quantum doesn't mean there's no uncertainty. We'll go there later). This limits the utility of one's training in CM, SM, and QM, respectively.

In fact there is plenty of EM, but it's very unlike the EM in a vacuum. It's actually surprisingly complicated, and involves some SM as well. Some of this we'll revisit in Sec. 2.4, “Electrostatic surprises in vivo.”

The physics in biophysics is a particular blend of statistical dynamics, EM, and Aristotelian mechanics. It was built by scientists who understood the four pillars well – well enough to stitch their own web using these pillars as supports. You have to understand a science very deeply to build and solve new models out of it, which must be done to describe a new application with it.

So, what is this world that is unlike the world of billiard balls and wave functions and ideal gases? Let's meet the relevant players and try to develop a mental picture of physics at the scale of the cell. En route we will see in what ways we can and can not apply physics as we know it.

## A.2 The “parts list:” a physicist cartoons biology



Figure A.1: parts list, as drawn by a theoretical physicist

It's important to know the nouns, so we can start speaking the language. Also it's useful to think about what these biological entities physically are. We'll enumerate the parts

list down to the scale of nucleic acids, which are only a bit wider than atoms themselves. This is the *physical* parts list – we will get to an *informatic* parts list later in this chapter, which will be central in Parts II and III.<sup>1</sup> (see [5] ch 2, “What’s inside Cells”, esp. Fig 2.1)

PCN calls this list the *Dramatis personae*

### A.2.1 Cells

In the Feynman lectures on Physics, Feynman writes about how much one could learn about the world, at scales larger than quantum mechanics, from the datum that the world is composed of atoms. Similarly, there is much one can learn about biology at scales larger than the wavelength of light from the fact that we and most of the organisms we interact with are made of cells.

That is,

“cell:biology::atom:physics”

- Cells are defined by walls.
- How to walls work? where do they come from? (you eat them)
- “lipids” and “hydrophobicity”: cooking things to make them ordered
- Digression: how big is light? So what? (How do we know these things?)
- what is inside a cell? (goodsell’s fig 5.2)
- vocabulary: pro and eu-karyotes

### A.2.2 proteins

- where do they come from?
- what do they do? (move; control who goes where; make DNA; proofread)
- really? they do that? how? ATP.
- wait – but where does ATP come from?

### A.2.3 small molecules

- ATP & ADP
- the mitochondria story
- are there other ATPs? Yes GTP, an NTP
- wait – is that the same “adenine”? yes.

---

<sup>1</sup>“Parts list” appears in quotes because this section is heavily inspired by Chapter 2 of Phil Nelson’s excellent book on biological physics [5]

#### A.2.4 polymers

- synthesized (e.g., nucleic acids)
- auto-generated (e.g., MTs)

#### A.2.5 nucleic acids

how big is DNA? how big is an atom?

Major missing point: how does ATP get made?

#### A.2.6 Final ideas

##### How do things stay together?

(i) self-assembly (ii) active assembly and transport (iii) specific interactions. How strong is self-assembly? How active is the transport? How do things get specificity? See next chapter.

##### Where do these things come from?

there are some things you make, and other things you eat.

###### Things you make

proteins

structural filaments

actin/MT

monomers → hydrophobicity → polymers

motor proteins

they walk along tracks made by the structural filaments

used to make ATP and to turn ATP into motion

polymerases, which make nucleic acids, e.g., transcription

###### Things you eat

small molecules/acids

ATP

H<sub>2</sub>O

DNA/RNA

**Homework:**

1. PCN 2.1: learn the Greek alphabet. Certain Greek letters have special meaning in this class. For example,  $\mu$  often means “micron.” I’ll try to use  $\eta$  for viscosity but I might slip up and use  $\mu$ .
2. PCN 2.2-2.5 (choose at least one, and tell me which one you chose and which structure you looked at. What, if anything, did you learn about it?)
3. quantitative estimates/scales of things: how many cells make up you?

**A.3 “The machinery of life:” a biologist draws biology**

2

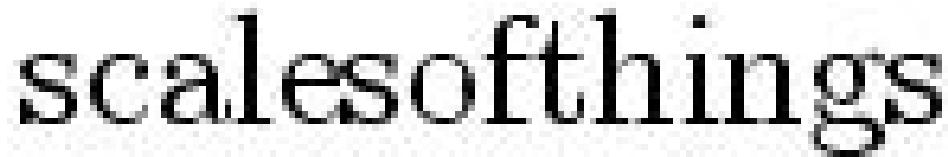


Figure A.2: scales of things

---

<sup>2</sup>The quotes refer to the excellent book of the same title by David Goodsell, from which several of these pictures are used with generous permission of the author [?]

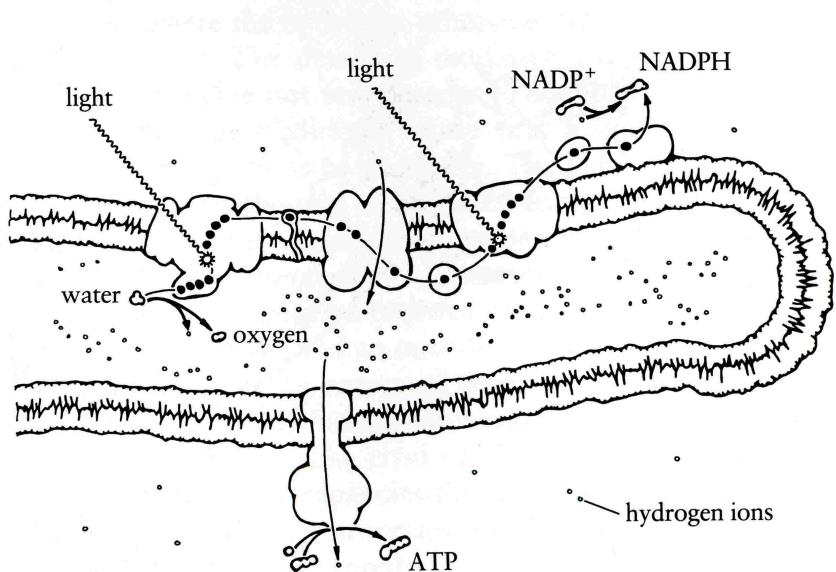
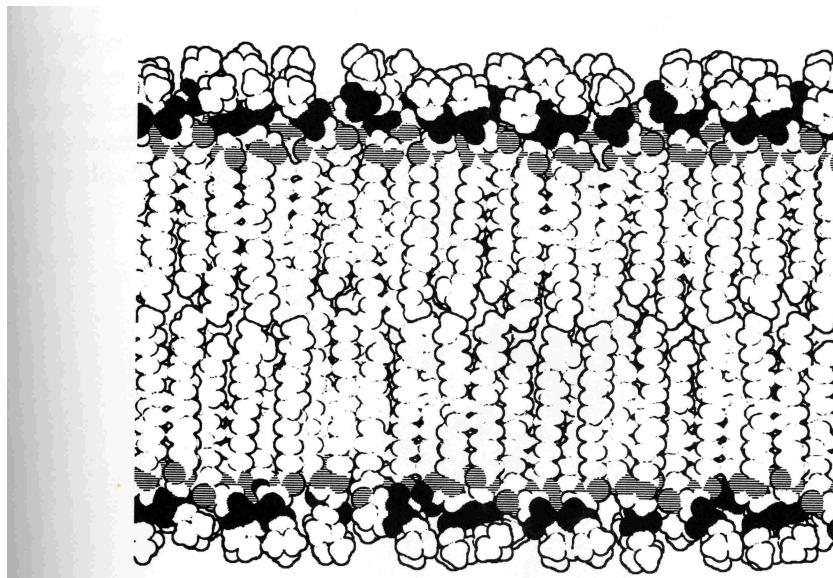


Figure 3.4 Photosynthesis

Figure A.3: ATP, the central energy storage molecule, being born. Note that it is not made the way protein is made (transcription and translation), nor the way “vesicles” are made (self-assembly via hydrophobic forces)



**Figure 2.6 Lipid Bilayer**

Many individual lipid molecules aggregate side by side to form a dynamic lipid bilayer, shown here in cross section. (10,000,000  $\times$ )

Figure A.4: A typical lipid bilayer, made via hydrophobic self-assembly, like oil and water.

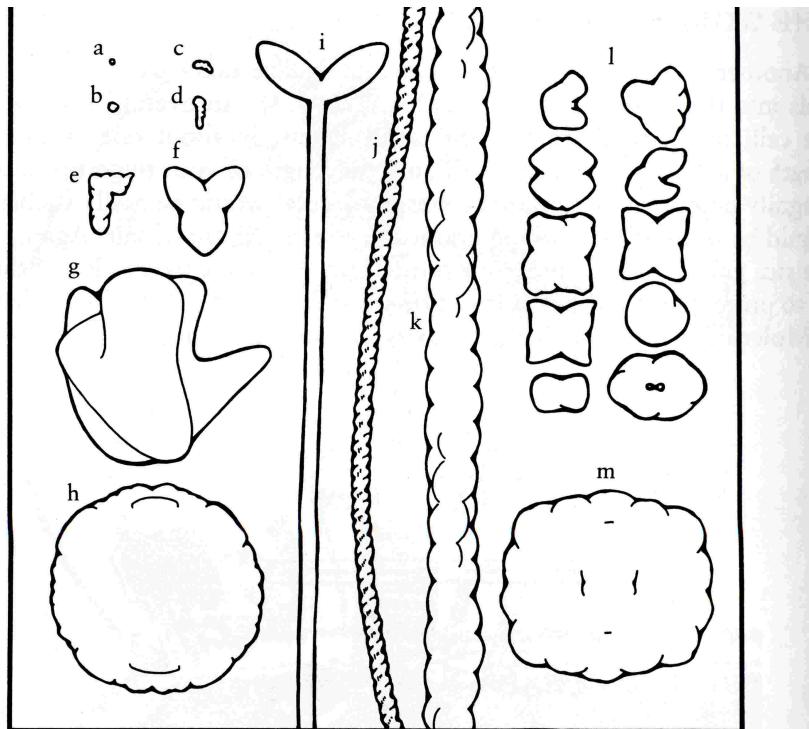


Figure 1.4 One Million Times Magnification

- a. Carbon atom.
- b. Glucose.
- c. Adenosine triphosphate (ATP).
- d. Chlorophyll.
- e. Transfer RNA.
- f. Antibody.
- g. Ribosome.
- h. Poliovirus.
- i. Myosin.
- j. Deoxyribonucleic acid (DNA).
- k. Actin.
- l. The ten enzymes of glycolysis.
- m. Pyruvate dehydrogenase complex.

Figure A.5: scales of small things

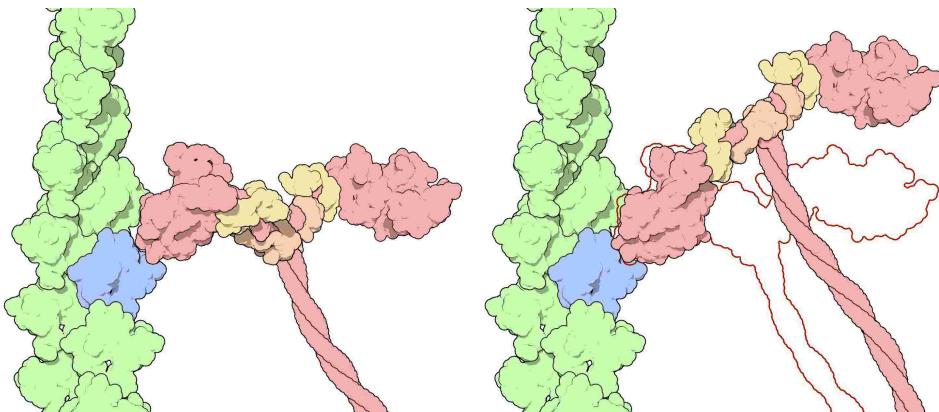


Figure A.6: mechanochemical coupling

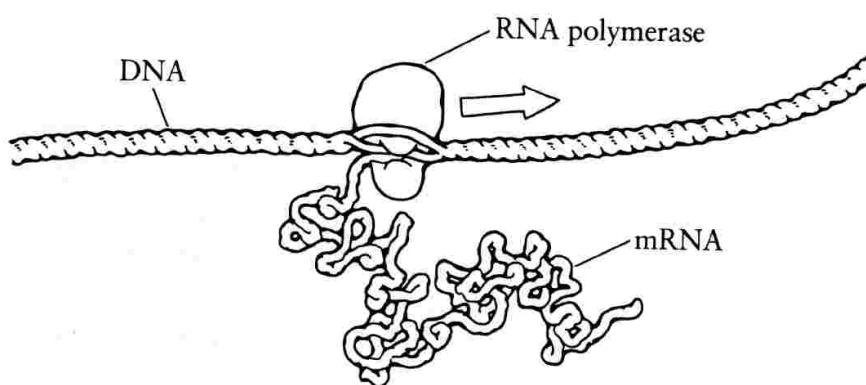


Figure A.7: mRNA being born

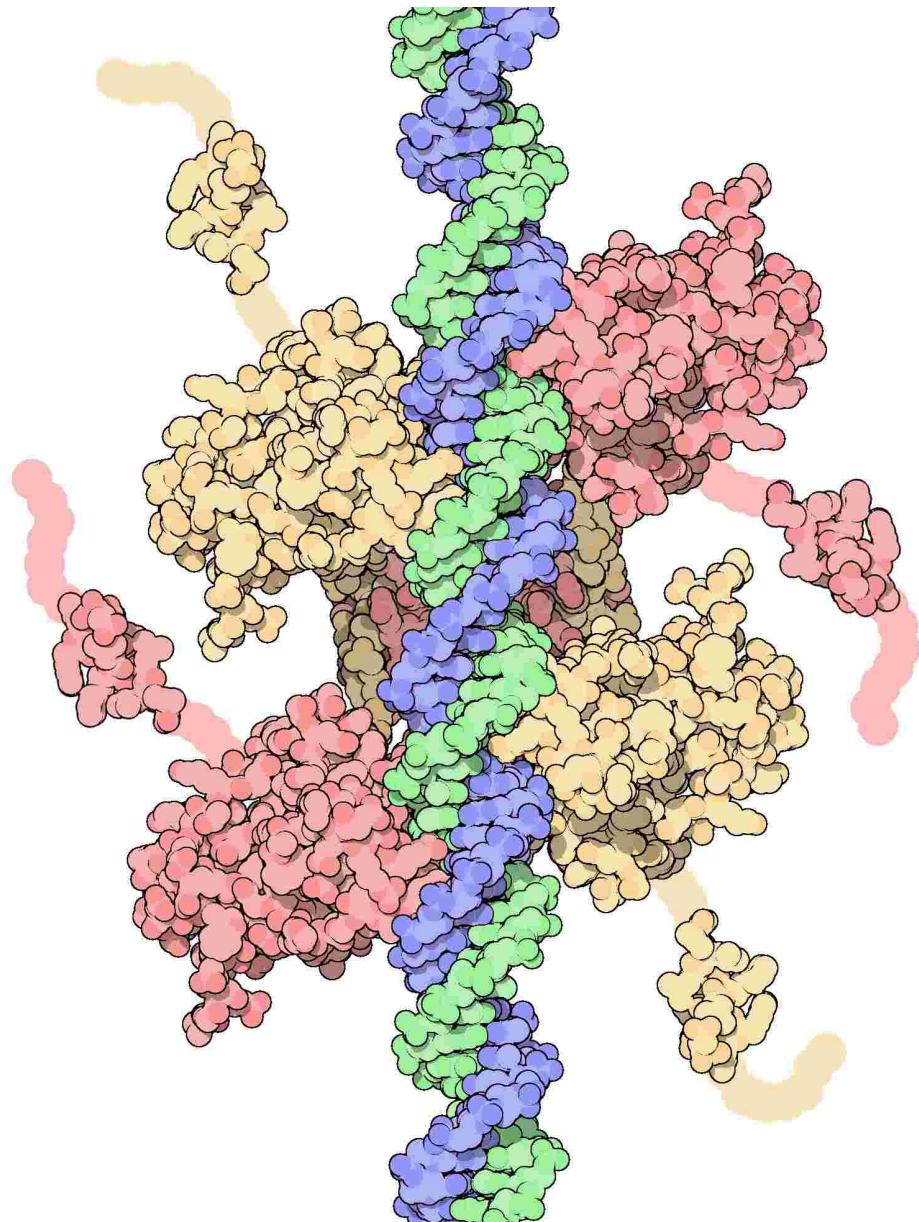


Figure A.8: A transcription factor binding to DNA. This interaction, between a protein and a nucleic acid, is the fundamental interaction of transcriptional regulation — genes turning “on” and “off.” There will be plenty more to say about this in Parts II and III.

245 A.3. "THE MACHINERY OF LIFE:" A BIOLOGIST DRAWS BIOLOGY

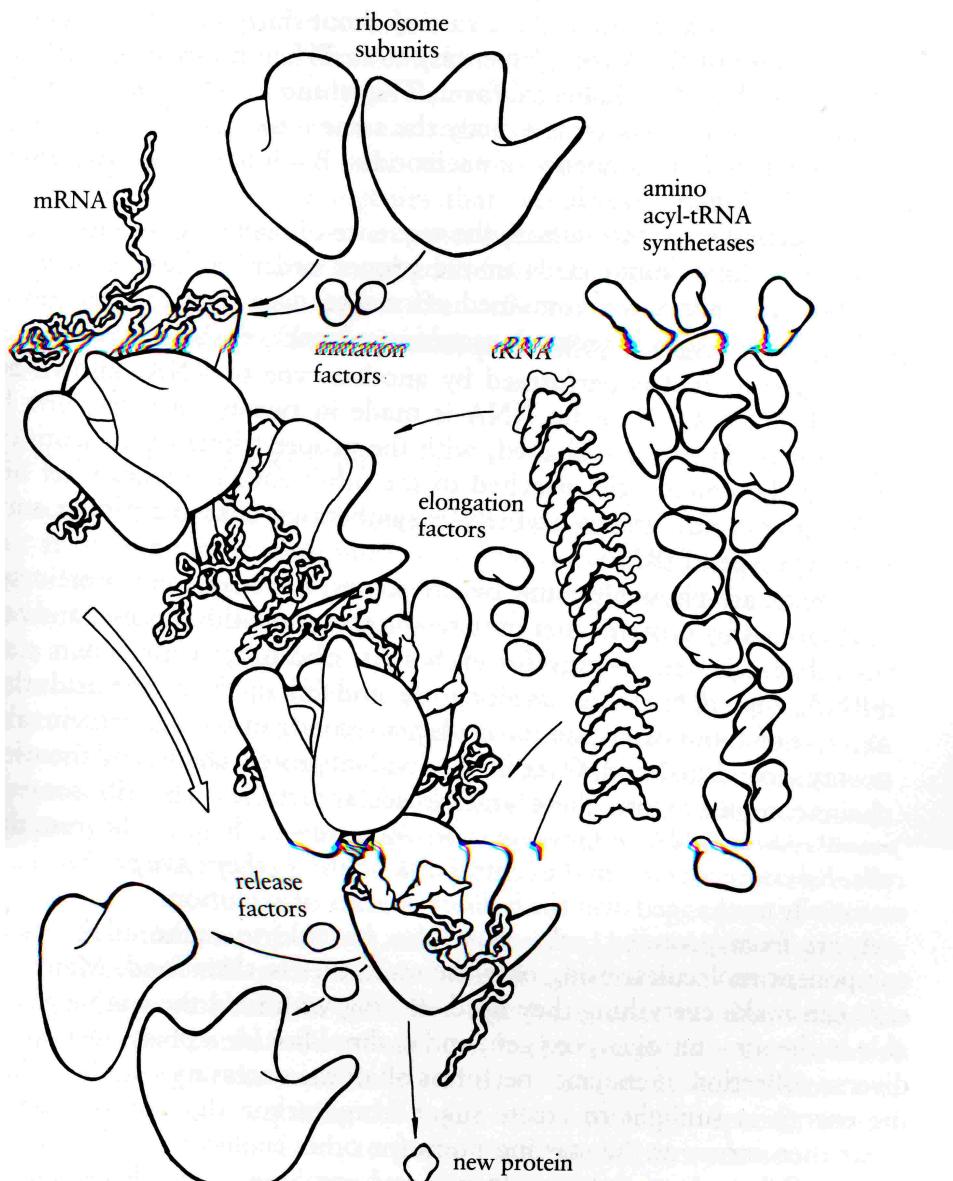


Figure A.9: Protein synthesis

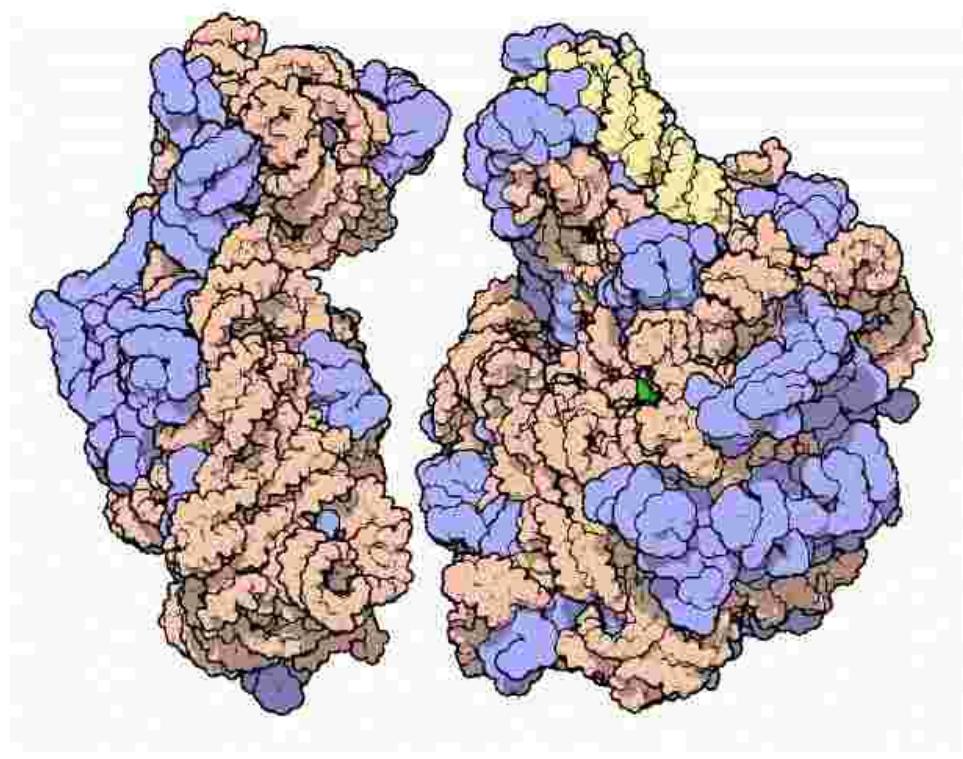


Figure A.10: The ribosome, a machine made of nucleic acids and proteins for translating mRNA into protein.

## Appendix B

# Dimensions of things

### B.1 Dimensional analysis

Almost any mathematical model of the world will be expressed in dimensionful quantities. Examples of these are velocity or height, which can be measured in *units* appropriate to the dimension. For example,

height has **dimensions** of length and can be expressed in **units** of, for example, feet, yards, meters, smoots, or light-years.

You've probably heard at some point the definitional relation "distance equals rate times time." Let's write that out as an equation:

$$d = r\tau \tag{B.1}$$

and ask if it makes *dimensionful sense*. A rate has dimensions of length over time. The notation for this is with brackets []:

$$[r] = l/t \tag{B.2}$$

$$[d] = l \tag{B.3}$$

$$[\tau] = t \tag{B.4}$$

making it pretty clear what dimensionful sense means here: when multiplying, one multiplies the dimensions

$$[d] = [r\tau] = [r][\tau] \tag{B.5}$$

$$l = (l/t)t \tag{B.6}$$

and it all works out OK.

What about addition? Walk 10 feet. Then drive 5 miles more. It better be the case that the total distance traveled has the same dimensions; different *units* for summands is just fine. Avoiding adding, say *10feet* and *20minutes* is what we mean by comparing apples and oranges.

It's important to keep these straight. This trivial observation is our starting point for the discussion of dimensional analysis.

Usually only three dimensions are necessary to describe everything in a physical problem, which is where many of you have seen mathematics applied before. Pathological cases abound, for example, in electrodynamics, but let's just stick with the three most common: length, time, and energy.<sup>1</sup> Other problems might have other dimensions (money, apples, oranges, etc), but let's work through this specific example to illustrate.

Let's return to Eqn. B.17. The degree of freedom  $x$  measures a distance – a length. The coordinate  $t$  has dimensions of time. What about the parameters  $g, \kappa$ ?

Looking at the LHS<sup>2</sup> of Eqn B.17, we see it has dimensions of

$$[\ddot{x}] = \left[ \frac{d^2x}{dt^2} \right] = \frac{[x]}{[t^2]} = l/t^2. \quad (\text{B.8})$$

What about the definition  $h = (1/2)\kappa x^2$ ? A little algebra<sup>3</sup> shows the dimensions of the curvature  $\kappa$  are

$$[h] = [(1/2)][\kappa][x^2] \quad (\text{B.9})$$

$$l = [\kappa]l^2 \quad (\text{B.10})$$

$$[\kappa] = 1/l \quad (\text{B.11})$$

just like in high school, when someone told you curvature can be expressed as

$$\kappa = \frac{y''}{\sqrt{1+y'^2}} \quad (\text{B.12})$$

$$[\kappa] = [y''] = l/l^2 = 1/l. \quad (\text{B.13})$$

Several important things just happened there –

1. Our first dimensionless number 1/2. Numbers have no dimensions.
2. Derivatives have dimensions! For example,  $y'' \equiv \partial_x^2 y$  has dimensions of inverse length
3. If you know the dimensions of one summand, you know the dimensions of all summands. For example, how do we know  $y'^2$  is dimensionless?

---

<sup>1</sup>Most people use length, time, and mass, but that's really a bias towards classical dynamics (and inertial dynamics in particular). What is a “mass” anyway? Mass is a dimension in itself, but in some cases it's also useful to think of it in terms of Energy. In case you forget their relation, you can remember “ $E = mc^2$ ”: mass is an energy per squared velocity. The notation for this is the brackets  $[\dots]$ :

$$[m] = Et^2/l^2. \quad (\text{B.7})$$

<sup>2</sup>“Left-Hand Side”

<sup>3</sup>In fact most of “dimensional analysis” is just algebra.

Finally let's look at the parameter  $g$ :

$$\ddot{x} = -g\kappa x \quad (\text{B.14})$$

$$[\ddot{x}] = -[g\kappa x] \quad (\text{B.15})$$

$$l/t^2 = [g](1/l)l \quad (\text{B.16})$$

So  $g$  is an acceleration, with dimensions of  $l/t^2$ . It turns out to be about 9.80665 *meters/s<sup>2</sup>*. What happened to the minus sign? Since  $(-1)$  has no dimensions, it does not enter into dimensional analysis.

Note that we can combine  $g\kappa$  into a new parameter with dimensions of  $1/t^2$  alone. This is a crucial point and motivates dimensional estimation and Buckingham's “ $\Pi$ ” theorem.

## B.2 Dimensional estimation

### Example: ball in a bowl

We'll begin with the basic equation of motion of classical mechanics:  $F = ma$ . If the bowl has a height given as  $h(x) = (1/2)\kappa x^2$ , then  $F = ma$  demands

$$m\ddot{x} = -mg\kappa x \quad (\text{B.17})$$

where “ $=$ ” =  $\partial_t^2 = \frac{d^2}{dt^2}$ . Don't panic if you've never taken a physics class or don't know what anything in this equation means.

Since we introduced Eqn. B.17 as a model for a ball in a bowl, you will not be surprised to hear that the solution describes oscillations. Now imagine that I asked you for the period of the oscillation as a function of the parameters in the problem. Could you make a guess? Remember that the period has dimensions, so whatever the answer is, it must be related to some combination of parameters that has the correct dimensions.

In this case, there is only one combination of parameters with the right dimensions:  $(g\kappa)^{-1/2}$ . So whatever the answer is, it must be some dimensionless number, like 1/2 or 137 or  $e^\pi$ , times this dimensionful quantity.



## Appendix C

# A brief review of all of undergraduate mathematics

### C.1 Pre-calculus

It's a good idea, in this course, to be on good terms with logarithms. Specifically, it'd be good to understand the following facts:

1.  $\ln(ab) = \ln a + \ln b$
2.  $e^0 = 1$
3.  $a \ln b = \ln b^a$
4.  $\ln 1 = 0$

#### **Homework:**

: prove line 4 using lines 2 and 3.

### C.2 Differential calculus

#### C.2.1 Differentiation and derivatives

Calculus is a fantastically beautiful idea – a real triumph of civilization, answering questions at least as old as those of the Greek natural philosophers (e.g., Zeno's paradox<sup>1</sup> of 450 B.C.E.)

---

<sup>1</sup>Zeno of Elea (490 BCE-425 BCE) — not to be confused with Zeno of Sidon (now Saida in Lebanon) (150 BCE-70 BCE).

If a body moves from A to B then before it reaches B it passes through the mid-point, say B<sub>1</sub> of AB. Now to move to B<sub>1</sub> it must first reach the mid-point B<sub>2</sub> of AB<sub>1</sub>. Continue this argument to see that A must move through an infinite number of distances and so cannot move.

The essential idea was that an infinite number of infinitesimally small quantities could sum to a finite amount, disproving Zeno's paradox.

Progress really heated up with Kepler's 'method of indivisibles'<sup>2</sup>, which showed  $\int_0^a dx x^n = a^{n+1}/(n+1)$  and Newton's method of fluxions, which showed  $d/dx(x^n) = nx^{n-1}$ .

### C.2.2 Taylor expansion, the most powerful and general tool in over 300 years of Calculus™

What Newton really noted is that[4],

In the time in which  $x$  by flowing becomes  $x + \delta$ , the quantity  $x^n$  becomes  $(x + \delta)^n$  i.e. by the method of infinite series,

$$(x + \delta)^n = x^n + n\delta x^{n-1} + (n^2 - n)/2\delta^2 x^{n-2} + \dots \quad (\text{C.1})$$

Of course this is just a special case of a more general statement about how an arbitrary function  $f(x + \delta)$  is related to  $f(x)$ :

$$f(x + \delta) = f(x) + \delta f'(x) + \frac{1}{2}\delta^2 f''(x) + \dots \quad (\text{C.2})$$

an expansion which Lagrange called the basic principle of the differential calculus, and which we now call Taylor expansion<sup>3</sup>.

### C.2.3 Application

Taylor expansion, the most powerful and general tool in over 300 years of Calculus™ is at the root of, among other things, much of approximation and asymptotics, including both numerical and analytical techniques for approximating functions.

---

<sup>2</sup>Johannes Kepler (1571-1630) is the same Kepler of Kepler's three laws of planetary motion. "He also did important work in optics (1604, 1611), discovered two new regular polyhedra (1619), gave the first mathematical treatment of close packing of equal spheres (leading to an explanation of the shape of the cells of a honeycomb, 1611), gave the first proof of how logarithms worked (1624), and devised a method of finding the volumes of solids of revolution that (with hindsight!) can be seen as contributing to the development of calculus (1615, 1616). Moreover, he calculated the most exact astronomical tables hitherto known, whose continued accuracy did much to establish the truth of heliocentric astronomy." [2]

<sup>3</sup>Brook Taylor (1685-1731) claimed it was a comment made in Child's Coffeehouse that led to his 1715 paper introducing the above expansion. Of course the same idea had already been noted by Newton, Leibniz, and other luminaries. Taylor did invent integration by parts, though ( cf. Sec C.3.2).

The simplest of these is obtained as in Newton's example, when we wish to study a function near some particularly interesting value.

As an example, a student once wrote

I don't understand what you mean by  $y \ll t$  in terms of applying it to a Taylor expansion. Do you make the  $y$ 's approach 1? Do I only Taylor expand the  $\ln(\dots)$  parts? If so, what do I do with the  $y$ 's that are not inside a function?

Taylor expansion is a trick for evaluating a function nearby to a value where evaluating the function is easier, or for turning nonlinear functions into linear functions.

If someone says  $y \ll t$ , chances are that what they want you to do is to identify everywhere that  $y + t$  occurs in an expression and Taylor expand.

In general, we can write  $\epsilon = y/t$

$$f(t, y) = f(t, \epsilon \cdot t) = g(t, \epsilon) \approx g(t, 0) + \epsilon g'(t, 0) \quad (\text{C.3})$$

where the prime  $g'$  refers to differentiating by the second argument.

More specifically, let's consider

$$\ln(a + b) \quad (\text{C.4})$$

when  $b \ll a$ . We can rewrite this as (defining  $\epsilon \equiv b/a$ )

$$\ln a + \epsilon \cdot a = \ln a(1 + \epsilon) = \ln(1 + \epsilon) + \ln a \approx \epsilon + \ln a + \mathcal{O}(\epsilon^2). \quad (\text{C.5})$$

Implicitly, we had

$$f(a, b) \equiv \ln(a + b) \quad (\text{C.6})$$

$$g(a, \epsilon) \equiv \ln a(1 + \epsilon) \quad (\text{C.7})$$

And used the fact that

$$\partial_x \ln x = 1/x \rightarrow \ln(1 + x) \approx \ln 1 + x(1/1) + \dots \quad (\text{C.8})$$

### C.3 Integral calculus: the opposite of differential calculus

The way one is usually introduced to Calculus is via integrals, which are first presented to you as sums of an infinite number of infinitely small things, or the limit of sums of a function evaluated at many many locations  $x$ , separated by little displacements  $\delta x$ , in the limit as  $\delta x \rightarrow 0$ .

That is, we evaluate  $f(a)$ ,  $f(a + \delta x)$ ,  $f(a + 2\delta x)$ , ... and sum them up. Sometimes this is also described as finding the area under the graph of  $f(x)$  as  $x$  runs from  $a$  to  $b$ .

So we define

$$\int_a^b dx f(x) = \lim_{\delta x \rightarrow 0} \delta x \sum_{j=0}^{j=(b-a)/(\delta x)} f(a + j \cdot \delta x). \quad (\text{C.9})$$

But a more useful way is to think of integration and differentiation as simply two sides of the same coin. This is particularly useful for relations such as conservation laws – these are differential relations, but to understand why they are important, you integrate them, and think about the consequences for the integrated version of the equation.

We could define integrals not by sums but by derivatives; for example

$$\int_a^b dx f(x) \equiv F(b, a) \quad (\text{C.10})$$

$$\text{such that } \partial_b F(b, a) = f(b) \quad (\text{C.11})$$

$$\text{and } \partial_a F(b, a) = -f(a) \quad (\text{C.12})$$

Now the only trick is, if someone hands you not  $\int d(F(x))$  but  $\int f(x)dx$ , figuring out  $F(x)$  such that  $dF/dx = f$ . Most of high school calculus is about this problem, but there's much more to integrals than that. As a special case, consider what happens when we write the product rule (for functions  $v(x), u(x)$ ) in its differential, and then its integral, form:

$$\frac{d}{dx}(uv) = v\frac{du}{dx} + u\frac{dv}{dx} \quad (\text{C.13})$$

$$\int_a^b d(uv) = u(b)v(b) - u(a)v(a) = \int_a^b vdu + \int_a^b udv, \quad (\text{C.14})$$

$$\int_a^b udv = uv|_a^b - \int_a^b vdu \quad (\text{C.15})$$

making clear that integration by parts and the product rule are also two sides of one coin. As a special case, consider  $u = 1$ ; in this case

$$\int_a^b dv = v|_a^b \quad (\text{C.16})$$

that is, the fundamental theorem of calculus.

### C.3.1 Simple examples:

### C.3.2 The best trick in calculus: integration by differentiation

There are many tricks that you will learn in calculus. The best tricks, though, involve turning an integral problem into a differentiation problem

### Integration by parts

**Integration by differentiation (a.k.a., ‘Moment generating functions’ or ‘partition functions’)**

### A duality: integration is differentiation

You might consider this fact trivial or mystifying. The best approach is to consider it mystifying until you realize it’s trivial. Then you will have attained enlightenment.

### C.3.3 The Gaussian integral, the only integral you really need to know<sup>TM</sup>

There’s only one integral you need to know.

#### Question:

what is that integral?

**A:**

the Gaussian integral

It’s rare that anyone calculates in a textbook – analytically – an integral more complicated than the Gaussian integral. If we face something more complicated, we generally approximate it by a Gaussian integral in the end, anyway (*cf.* . Sec. C.3.3)

#### Definition

Here’s a convenient form in which to remember the Gaussian integral

$$1 = \int_{-\infty}^{+\infty} dx \frac{e^{-\frac{1}{2}(x/\sigma)^2}}{\sqrt{2\pi\sigma^2}} \quad (\text{C.17})$$

It’s a convenient form to remember the Gaussian integral because it looks like a probability normalization condition.

$$p(x) = etc \quad (\text{C.18})$$

looks like a probability density for  $x$ . Note that, in its current form, the left hand side has no dependence on sigma.

#### Proof

Now, if you wanted to solve the integral, you would just integrate this, but of course nobody knows how to solve the Gaussian integral. You don’t ever *solve* the Gaussian integral. What you do is solve for the *square* of the Gaussian integral. So I will now square both sides of this equation.

$$1 = \int dx_1 dx_2 e^{-\frac{1}{2}x_1^2/\sigma_{11}^2 - \frac{1}{2}x_2^2/\sigma_{22}^2} \frac{1}{\sqrt{(2\pi)(2\pi)\sigma_{11}^2\sigma_{22}^2}}. \quad (\text{C.19})$$

To solve this directly, first rescale as  $u = x_1/\sigma_{11}$ ,  $v = x_2/\sigma_{22}$ , and change to polar coordinates  $(u, v) = (r \cos \theta, r \sin \theta)$ :

$$1 = \frac{\sigma_{11}\sigma_{22}}{2\pi\sigma_{11}\sigma_{22}} \int dudv e^{-\frac{1}{2}u^2+v^2} \quad (\text{C.20})$$

$$= \frac{1}{2\pi} \oint_0^{2\pi} d\theta \int_0^\infty r dr e^{-\frac{1}{2}r^2} = - \int_1^0 d(e^{-\frac{1}{2}r^2}) = -(0-1) = 1 \quad (\text{C.21})$$

Note also that this can be rewritten as the two-dimensional Gaussian integral

$$1 = \int dx_1 dx_2 e^{-\frac{1}{2}(x_1 x_2) \Sigma^{-2} (x_1 x_2)^T} \frac{1}{\sqrt{|2\pi\Sigma^2|}} \quad (\text{C.22})$$

where

$$\Sigma^{-2} \equiv \begin{pmatrix} \sigma_{11}^{-2} & 0 \\ 0 & \sigma_{22}^{-2} \end{pmatrix} \quad (\text{C.23})$$

All I did was square the integral, which is how people solve the integral in the first place, and then I rewrote it as a matrix.

### The n-dimensional case

Now, again, the left hand side doesn't have any dependence on these  $\sigma$ 's, and, in fact, I could have chosen these  $x_1$  and  $x_2$  awkwardly; specifically, I could have chosen them to be non-diagonal directions as defined by the matrix  $\Sigma$ . I could have performed some rotation in the space  $x_1, x_2$ .  $\Sigma$  is then the covariance matrix, with possibly non-diagonal elements. Nonetheless, the answer is still 1, regardless of what rotation of the dummy variables  $x_1, x_2$  was performed. In general, we write

$$\Sigma \equiv \begin{pmatrix} \sigma_{11}^{-2} & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^{-2} \end{pmatrix} \quad (\text{C.24})$$

where in the diagonal case,  $\sigma_{11} = \sigma_1$ ,  $\sigma_{22} = \sigma_2$ . When there are off-diagonal terms, the dimensionless  $\rho \equiv \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$  is called the Pearson correlation coefficient.

### The Gaussian as an n-dimensional generating function

Now let's recall a very useful fact. When we first introduced the Gaussian integral, we showed that if you put a term  $\alpha$  up in the exponent, and then differentiated with respect to it, you could generate moments easily. So let's define

$$f(\alpha) \equiv \int d^d x e^{-\frac{1}{2}\alpha x^T \Sigma^{-2} x} \sqrt{|2\pi\Sigma^2|} = \sqrt{\frac{|\alpha 2\pi\Sigma^2|}{|2\pi\Sigma^2|}} = \alpha^{d/2}. \quad (\text{C.25})$$

If that's not clear, recall that

$$|\alpha| = \alpha \left| \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \right| = \alpha^2: \quad (\text{C.26})$$

So, differentiating every term in Eqn. C.25, and evaluate at  $\alpha = 1$ , we have

$$f' = \int d^d x \left( -\frac{1}{2} x^T \Sigma^{-2} x \right) e^{-\frac{1}{2} \alpha x^T \Sigma^{-2} x} \sqrt{|2\pi\Sigma^2|} = (-d/2) \alpha^{-d/2-1} \quad (\text{C.27})$$

$$f'(1) = \left\langle -\frac{1}{2} x^T \Sigma^{-2} x \right\rangle = (-d/2). \quad (\text{C.28})$$

In physics, this calculation is called the equipartition theorem.

### The Gaussian as an approximation to any integral

Since the Gaussian integral is the only integral we really know how to solve, what do we do if someone hands us something else? The answer is we turn it into a Gaussian integral.

First consider integrating  $f(x)$ , with  $f(x)$  a positive, dimensionless function, over all  $x$ . We can define  $g(x) = \log f(x)$  to attempt to make it look like a Gaussian integral, and then approximate by Taylor expansion, the most powerful and general tool in over 300 years of Calculus™

$$g(x) \approx g(x_*) + (x - x_*)g'(x_*) + \frac{1}{2}(x - x_*)^2 g''(x_*) \quad (\text{C.29})$$

A great choice for  $x_*$  would be such that  $g'(x_*) = 0$ . For the moment consider the case when there is a unique  $x_*$  such that this holds and that  $g''(x_*) < 0$ . If all of these ( $x \in R$ ;  $f(x) > 0$ ;  $\exists! x_*$  s.t.  $g'(x_*) = 0$ ;  $g''(x_*) < 0$ ), then we can approximate  $\int dx f(x)$  as (introducing a change of variables  $\eta \equiv x - x_*$ )

$$I(x) \equiv \int_{-\infty}^{\infty} dx f(x) \approx \int_{\infty}^{\infty} dx e^{g(x_*) + (x - x_*)g'(x_*) + \frac{1}{2}(x - x_*)^2 g''(x_*)} \quad (\text{C.30})$$

$$= e^{g(x_*)} \int d\eta e^{\frac{1}{2}\eta^2 g''(x_*)} = e^{g(x_*)} \sqrt{-2\pi/g''(x_*)} \quad (\text{C.31})$$

because it's a Gaussian integral, the only integral you really need to know™.

### Caveats and generalizations

Like any approximation, it's important to quantify how bad the approximation is. This approximation is safest when  $f(x)$  is sharply peaked – that is when the curvature  $\kappa \equiv \sqrt{-g''(x_*)}$  is large. In this case one often relaxes several of the assumptions including:

- We might use this even for integrals which do not run from (or to) some finite number  $a$  (rather than over all  $x$ ), since the errors will be bounded by functions of  $\text{erf}(a\kappa)$  which will be small
- We might use this even for dimensionful functions by defining  $g(x) \equiv \log f(x)/f_0$  for some characteristic scale  $f_0$  with the same dimensions as  $f(x)$  (e.g.,  $f_0 = f(x_*)$ ).
- We might consider the case where there are multiple  $x_*$ , in which case the integral is approximated by the sum over these contributions

- We might consider the vector-valued integral
- We might consider a path integral
- We might consider a complex-valued contour integral, in which we should deform the contour to run through  $z_*$  such that  $f'(z_*) = 0$  and to run along the direction of steepest descent (the direction which maximizes  $|f''(z_*)|$ )

Since  $g = \ln f$ ,  $g'(x_*) = f'(x_*)/f(x_*) = 0$ , and  $g'' = (f''f - f'^2)/f^2 = f''/f$ , we can rewrite the most uncontrolled statement of the approximation as, in the vector case,

$$\int d\mathbf{x} f(\mathbf{x}) \approx f(\mathbf{x}_*) / \sqrt{\left| \frac{f''(\mathbf{x}_*)}{2\pi f(\mathbf{x}_*)} \right|} \quad (\text{C.32})$$

where  $f''(\mathbf{x}_*)$  is the Hessian<sup>4</sup> matrix  $[f''(\mathbf{x}_*)]_{ij} \equiv \partial_{x_i} \partial_{x_j} f(\mathbf{x}_*)$  and the bars  $|\cdots|$  indicate the determinant.

## C.4 $\dot{x} = \alpha x$ : The only ODE you really need to know<sup>TM</sup>

This won't take long, but it's important:

$$\dot{x} = \alpha x \quad (\text{C.33})$$

$$\frac{dx}{dt} = \alpha x \quad (\text{C.34})$$

$$\frac{dx}{x} = \alpha dt \quad (\text{C.35})$$

$$\ln x = \alpha t + C \quad (\text{C.36})$$

$$x = x_0 e^{\alpha t}. \quad (\text{C.37})$$

That's it. You now know The only ODE you really need to know<sup>TM</sup> for this course.

## C.5 Solving the Diffusion Equation

The diffusion equation is the first PDE most people solve, and gives a clear and simple introduction to PDEs more generally. One of the many ways to solve this PDE is using Fourier techniques, which can be briefly stated as:

- represent the function in “Fourier space”;
- note that the PDE becomes an ODE in this space;
- solve the ODE;
- and come back from Fourier space to real space.

---

<sup>4</sup>Ludwig Otto Hesse (1811-1874) introduced this determinant in a paper in 1842 during an investigation of cubic and quadratic curves.

### C.5.1 fourier space

Going to fourier space means representing the function in terms of its fourier modes or, for a function defined on the real axis, the fourier transform. The transform relations are:

$$f = \int \frac{dk}{2\pi} e^{-ikx} \hat{f}(k) \quad (\text{C.38})$$

$$\hat{f}(k) = \int dx e^{ikx} f(x). \quad (\text{C.39})$$

Qualitatively, we are representing a function as a linear combination of wiggling modes, each wiggling with shorter wavelength as  $k$  increases, and  $\hat{f}(k)$  is the strength of each wiggling. The glory of this representation is that  $e^{\alpha x}$  is an eigenfunction of differentiation with eigenvalue  $\alpha$ , that is

$$\partial_x^m e^{-ikx} = (-ik)^m e^{-ikx} \partial_x^m f(x) = \int \frac{dk}{2\pi} (-ik)^m \hat{f}(k). \quad (\text{C.40})$$

For a partial differneital eqution, where the function is of two variables, we have, for example

$$\dot{f} = Df'' \int \frac{dk}{2\pi} e^{-ikx} \partial_t \hat{f}(k, t) = D \int \frac{dk}{2\pi} (-ik)^2 e^{-ikx} \hat{f}(k, t) \int \frac{dk}{2\pi} e^{-ikx} (\partial_t \hat{f}(k, t) + Dk^2 \hat{f}) = 0 \quad (\text{C.41})$$

which is solved by  $\hat{f}$  satisfying the ODE

$$\partial_t \hat{f} = -Dk^2 \hat{f}. \quad (\text{C.42})$$

### C.5.2 ODE

This is an eigenfunction equation, with eigenvalue  $-Dk^2$ , which we know (see Sec ??) is solved by

$$\hat{f}(k, t) = \hat{f}(k, 0) e^{-Dk^2 t} \quad (\text{C.43})$$

### C.5.3 solution

To solve this, we need to know the initial data  $\hat{f}(k, 0)$ . For the special case  $f(x, t) = \delta(x)$ , e.g., for which  $f(x, t)$  is a probability distribution localized at the origin at time  $t = 0$ , we have

$$\hat{f}(k, 0) \equiv \int dx e^{ikx} f(x, 0) = \int dx e^{ikx} \delta(x) = 1; \quad (\text{C.44})$$

the fourier transofmr of a delta function is flat.

### C.5.4 come back

We can now invert for the solution in “real space”:

$$f(x, t) = \int \frac{dk}{2\pi} e^{-ikx} \hat{f}(k, t) = \int \frac{dk}{2\pi} e^{-ikx} e^{-Dk^2 t}. \quad (\text{C.45})$$

fortunately this is a gaussian integral, the only intergral you really need to know, for which (see Sec ??)

$$g(k) = Dk^2 t + ikx \quad (\text{C.46})$$

$$g'(k) = 2Dkt + ix \quad (\text{C.47})$$

$$g''(k) = 2Dt \quad (\text{C.48})$$

$$k_* = -ix/(2Dt) g_* = Dt(-x^2/4)/(Dt)^2 + ix(-ix)/(2Dt) = x^2/(4Dt) \quad (\text{C.49})$$

$$g''_* = 2Dt \equiv \quad (\text{C.50})$$

and, defining  $\sigma^2 \equiv 2Dt$ ,

$$f(x, t) = \frac{1}{2\pi} e^{-g_*} / \sqrt{2\pi/g''_*} = \sqrt{2\pi\sigma^2} e^{-\frac{1}{2}(x/\sigma)^2} \quad (\text{C.51})$$

in accord with the intuition we gained by studying  $\langle x_t^2 \rangle = 2Dt$ .

## C.6 Linear algebra

Linear algebra is a beautiful subject. Unfortunately I’m not going to discuss it in any detail here. I urge you though, if you haven’t taken a linear algebra class, at least to flip through a good book on the subject; perhaps one that doesn’t start with row reduction and solving systems of equations, because linear algebra isn’t really about that at all.

In any event, here we’re only going to learn enough linear algebra to change the way you think about dynamical systems.

In Sec. ??, we learned that in fact there is only one ODE you really need to know:  $\dot{x} = \alpha x$ . This seems hard to believe, for example, if you’ve ever taken any physics, and you’ve seen  $F = ma$ , for example the equation of motion for a ball in a bowl of molasses.

$$ma = m\ddot{x} = -b\dot{x} - kx. \quad (\text{C.52})$$

In fact, this is really a special case of Eqn. C.33.

### C.6.1 Dynamical systems and linear algebra

Here is a key fact

Any single  $n - th$  order differential equation can be re-expressed as  $n$  1st order differential equations

as an example of a 2-d dynamical system, look at a ball rolling at the bottom of bowl filled with a viscous fluid. Newton's 2nd law gives the second order ODE

$$m\ddot{x} = -kx - b\dot{x}. \quad (\text{C.53})$$

letting  $v \equiv \dot{x}$ , we can transform this into a set of coupled first order ODEs:

$$\dot{x} = v \quad (\text{C.54})$$

$$\dot{v} = -\frac{k}{m}x - \frac{b}{m}v \quad (\text{C.55})$$

or, in matrix form

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{b}{m} \end{pmatrix} \begin{pmatrix} x \\ v \end{pmatrix}. \quad (\text{C.56})$$

As an example, let's take the ball in the bowl. We can define the velocity  $\dot{x} = v$  and rewrite the one equation as two first order equations:

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= -b/mv - k/mx \end{aligned} \quad (\text{C.57})$$

or, defining the vector  $\mathbf{u} \equiv (x, v)$ ,

$$\dot{\mathbf{u}} = \begin{pmatrix} v \\ -b/mv - k/mx \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -b/m & -k/m \end{pmatrix} \begin{pmatrix} v \\ x \end{pmatrix} = M\mathbf{u} \quad (\text{C.58})$$

where the matrix  $M$  contains only constants

$$M \equiv \begin{pmatrix} 0 & 1 \\ -b/m & -k/m \end{pmatrix} \quad (\text{C.59})$$

and the notation called "matrix multiplication" is just a shorthand for the definitional relation (here subscript indicates component number, *i.e.*,  $u_1 \equiv x$ ,  $u_2 \equiv v$ )

$$\dot{u}_j = \sum_{i=1}^{i=2} M_{ij} u_i. \quad (\text{C.60})$$

It will turn out that all we need to know about the behavior of the ball in the bowl is contained in two numbers which we can calculate from this matrix. To make that more clear, let's step back and consider the simplest 2D dynamical system in the entire world.

## C.6.2 the simplest 2D dynamical system in the entire world

Consider two copies of Eqn. ??, the only ordinary differential equation you really need to know<sup>TM</sup>.

$$\dot{x} = \alpha x \quad (\text{C.61})$$

$$\dot{y} = \beta y \quad (\text{C.62})$$

restricted to real  $\alpha, \beta$ . We can solve it completely of course (*cf.* Sec. C.4):

$$x = x_0 e^{\alpha t} \quad (C.63)$$

$$y = y_0 e^{\beta t} \quad (C.64)$$

how would you characterize the stability of this (linear) dynamical system? Clearly all points flow towards the origin  $(0, 0)$  if  $a < 0$  and  $b < 0$ . If either is positive, the origin is unstable. We could address this more succinctly, and without treating  $a$  and  $b$  separately, by simply asking if the product of the two is positive (in which case they are of the same sign), and then asking if the sum of the two is negative (which, if the product is positive, means that they are both negative). To summarize

$$\tau \equiv \alpha + \beta \quad (C.65)$$

$$\Delta \equiv \alpha\beta \quad (C.66)$$

$$\text{stability} \iff \tau < 0, \Delta > 0 \quad (C.67)$$

which makes things quite simple.

Now that we know everything there is to know, qualitatively, about the behavior of this dynamical system, let's rewrite it as a single, 1d dynamical system:

$$(C.68)$$

### Return to the bowl

Sadly the bowl is not that simple. There is an off-diagonal term in the matrix which couples  $x$  and  $v$ . Another way of saying this is that the axes  $(x, v)$  are not a convenient basis for dealing with this dynamic, as the earlier basis  $(x, y)$  was. It would be great, wouldn't it, if there were some directions in  $(x, v)$  space such that the action of the matrix were as trivial as it is for just  $x$  and  $y$  above, namely multiplication by a constant.

Let's see if we can construct such magic directions, and see what the magic constants would be. That is, we will assert that there are magic directions  $\mathbf{u}$  such that  $M\mathbf{u} = \lambda\mathbf{u}$ , and we will strive to see what that says about the  $\lambda$ , since it is the behavior of the constants which tells you all you need to know about stability.

### Construction of magic directions

Any magic direction  $\mathbf{u}$  would have to obey

$$M\mathbf{u} = \lambda\mathbf{u} \quad (C.69)$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mathbf{u} = \lambda\mathbf{u} \quad (C.70)$$

$$au_1 + bu_2 = \lambda u_1 \quad \text{and} \quad cu_1 + du_2 = \lambda u_2 \quad (C.71)$$

Ultimately it's the  $\lambda$  we care about more than the  $u_{1,2}$ . We can multiply the first equation by  $d$  to solve for  $du_1$ , or multiply the second by  $b$  to solve for  $bu_2$ . The latter gives  $bu_2 = u_1(\lambda - a)$  which, substituted into the top equation, gives FIXXXXXXXXXX

$$au_1 + u_1(\lambda - 1) = \lambda u_1 \quad (C.72)$$

$$u_1 = 0 \text{ or } \lambda^2 - \tau\lambda + \Delta = 0 \quad (C.73)$$

where  $\tau = a+d$  is called the *trace* of the matrix and  $\Delta = ab - cd$  is called the *determinant*. It would seem, then, that all we need to know about the stability of any two dimensional linear dynamical system is contained in these two numbers.

### Homework:

Now try it the other way (multiply the top equation by  $d$  and solve for  $du_1$ ). If the answers differ, you have found a bug in linear algebra.

In fact, then, there are two magic directions, which, as long as they do not point in the same direction, means that we can describe any point in  $R^2$  as a combination of these two magic directions.

$$(\lambda - \tau/2)^2 + \Delta - \tau^2/4 = 0 \quad (\text{C.74})$$

$$\lambda_{\pm} \equiv \tau/2 \pm \sqrt{\tau^2/4 - \Delta} \quad (\text{C.75})$$

We'd like to have the same stability criteria as we did for  $\alpha$  and  $\beta$  in section C.6.2, which were (“ $\alpha + \beta < 0 + \alpha\beta > 0$ ”  $\Rightarrow$  “stability”). Here, we wish to show that the *real* part of  $\lambda$  is always negative.

If  $\Delta > \tau^2/4$ , the real part (for either) is  $\tau/2$ , so we have (“ $\Delta > \tau^2/4$ ” + “ $\tau < 0$ ”  $\Rightarrow$  “stability”).

If  $\Delta < \tau^2/4$ , then each magic value is real and the criteria are the same as they were for  $\alpha, \beta$ :

$$0 < \lambda_+ + \lambda_- = \tau^2/4 - \tau^2/4 + \Delta = \Delta \quad (\text{C.76})$$

$$0 > \lambda_+ - \lambda_- = \tau \quad (\text{C.77})$$

Summarizing, stability  $\iff \Delta > 0, \tau < 0$ , which is the same criterion as for  $\alpha, \beta$ , which we see is the special case restricted to  $b = c = 0$ .

### C.6.3 Complex magic values?

How can these magic values be complex? After all, they indicate a growth rate, no? To make this more clear, return to the ball in the bowl. The trace and determinant are

$$\tau = -b/m \quad (\text{C.78})$$

$$\Delta = k/m \quad (\text{C.79})$$

so, for all parameters positive, the origin is stable, as we expect. What if we changed the sign of the drag constant  $b$ ? Physically this would mean, since  $dE/dt = d(\frac{1}{2}kx^2 + \frac{1}{2}mx^2)/dt = -b\dot{x}^2$ , that the energy of the system *increases* instead of decreases with time, and the origin becomes unstable. What about when  $b$  is positive but very small? When  $4\Delta > \tau^2$ , the eigenvalues are imaginary, including in the special case  $b = 0$ , which is a case we know well:  $\ddot{x} = -(k/m)x \equiv -\omega^2x$ , the solution to which is simple harmonic motion  $x = x_0 \sin(\omega t + \phi)$ <sup>5</sup>. Therefore, the existence of an imaginary part to the eigenvalues implies oscillation near the fixed point.

---

<sup>5</sup>In fact, this result, along with the existence and uniqueness of solutions to  $\ddot{x} = -x$ , gives one way of proving Euler's relation  $e^{it} = \cos t + i \sin t$ .

APPENDIX C. A BRIEF REVIEW OF ALL OF UNDERGRADUATE  
MATHEMATICS

264

## Appendix D

# Probability

**Question:**

Write  $\langle \eta \rangle$  in terms of a probability

**A:**

$$\langle \eta \rangle = \sum_s p_s \eta_s = (1/2)(-1) + (1/2)(1) = 0$$

**Question:**

What properties must a function  $f(x, y, z)$  satisfy to be a probability  $p(x, y, z)$  in  $x, y, z$ ?

**A:**

Two properties:  $f(x, y, z) \geq 0$  and  $\int dx dy dz f(x, y, z) = 1$  (where the integral ranges over all possible values of  $(x, y, z)$ ).

**Question:**

Probability frequently uses the stunning abuse of notation  $p(x, y, \dots)$  with the understanding that  $p(x)$  is not necessarily the same function  $p(y)$ ; that is, the dependent variable is used to indicate which function is being discussed. Respecting this choice, how is  $p(x)$  related to  $p(x, y, z)$ ?

**A:**

$$p(x) = \int dy dz p(x, y, z).$$

**Question:**

How is  $p(y)$  related to  $p(x, y)$  and  $p(x|y)$ ?

**A:**

$$p(x, y) = p(x|y)p(y). \text{ Note that by symmetry, } p(x|y)p(y) = p(y|x)p(x).$$

## D.1 normalization

## D.2 expectation

Consider a game in which a coin is flipped and we exchange money based on the payoff. In the simplest of all possible cases, I pay you a dollar if the coin lands “heads”, you pay me a dollar if it lands “tails.” How much money do you expect to get? There are two possible future states, so

$$\langle m \rangle = \sum_s m_s p_s = (+1)(1/2) + (-1)(1/2) = 0. \quad (\text{D.1})$$

This is how we mathematize expectations: we calculate the value of an outcome given a particular state, multiply by the probability of the state, and sum over states.

## D.3 marginarization

## D.4 bayes rule

## D.5 probabilities and proabability distributions

In Eqn. D.1, we dealt with the probabilities  $p_{\text{heads}} = 1/2 = p_{\text{tails}}$ . These probabilities are dimensionless. Note that this is contrasted with a probability distribution: a probability defined over a continuous variable. The dimensions of the probability distribution are necessarily 1/ the dimensions of its coordinates, since

$$\int dx dy dz \rho(x, y, z) = 1 \rightarrow [\rho(x, y, z)] = 1 / ([x][y][z]). \quad (\text{D.2})$$

The expectation rule then becomes

$$\langle F(x) \rangle = \int dx F(x) \rho(x). \quad (\text{D.3})$$

## Appendix E

# Elastica

What is the bend energy of an elastic rod (a.k.a *elastica*)? Clearly it should depend on the curvature  $\kappa$ . It should be the same for positive and negative curvature so it should really depend on  $\kappa^2$ . But curvature is a function of position (or “arclength”)  $s$  so really it should depend on the integrated, squared curvature over the length  $L$  of the elastica:

$$E_{\text{bend}} \propto \int_0^L ds \kappa(s)^2 \quad (\text{E.1})$$

We represent the constant of proportionality by  $A$ :

$$E_{\text{bend}} = \frac{1}{2} A \int_0^L ds \kappa(s)^2 \quad (\text{E.2})$$

What are the dimensions of  $A$ ? In order to answer this we need to remember what  $\kappa$  is; in high school you learned that the curvature of a function  $y(x)$  is  $y''/\sqrt{1+y'^2}$ <sup>3</sup>. So it has dimensions of  $1/\ell$ .

So the dimensions of the above equation are

$$E = [A]\ell\ell^{-2} \quad (\text{E.3})$$

and  $[A]$  must have dimensions of  $E\ell$ . Right away, we see that  $A$  determines a characteristic length scale at room temperature, which we call the *persistence length*  $\ell_P$ :

$$\ell_P \equiv A/(k_B T). \quad (\text{E.4})$$

If you were raised as an engineer you might remember that the relationship between moments (torques) and curvatures is given by  $m = YI\kappa$ . In fact,  $A$  is simply this combination of  $Y$  (Young’s modulus, with dimensions of pressure  $E/\ell^3$ ) and  $I$  is the moment about a particular axis, with dimensions of  $\ell^4$ :

$$A \equiv YI. \quad (\text{E.5})$$

Note that  $I$  scales with the radius to the 4th power. For a perfect cylinder,  $I = \pi r^4/2$



## Appendix F

# Stirling's approximation

The factorial function comes up very often in probability and statistics, but it's a funny function. For example, we really can define it algorithmically:

$$\gamma(x) = x\gamma(x-1). \quad (\text{F.1})$$

$$\gamma(0) \equiv 1 \quad (\text{F.2})$$

from which  $\gamma(1) = 1, \gamma(2) = 2, \gamma(3) = 6, \text{etc}$ . There have been many notations for this function over the centuries, but the most common at present is  $\gamma(x) = x!$ . A second way of writing this is  $\gamma(x) \equiv \prod_{j=1}^{j=x} j$ . This makes it more clear that  $\gamma$  is defined, at present, only for integer arguments. .

### F.1 Approximate approximation: $n! \approx N^N e^{-N}$

To motivate the approximation, consider that

$$\ln N! = \ln N \cdot (N-1) \cdots 1 \cdot 1 = \sum_{j=1}^{j=N} \ln j \quad (\text{F.3})$$

An uncontrolled approximation would be to approximate the sum by an integral:

$$\ln N! \approx \int_{j=1}^{j=N} dj \ln j = j \ln j - j|_{j=1}^{j=N} = N \ln N - N + 1 \quad (\text{F.4})$$

$$N! \approx N^N e^{1-N} \quad (\text{F.5})$$

### F.2 Next order: $n! \approx N^N e^{-N} \sqrt{2\pi N}$

We can derive a more accurate (and commonly-used) expansion by noting that “factorial” can actually be expressed analytically as an integral expression; not just via an algorithm.

In a letter from January 8, 1730 to Christian Goldbach, the Swiss mathematician Leonhard Euler (1707-1783) proposed the following definition

$$\Gamma(z+1) \equiv \int_0^\infty dt t^z e^{-t} \quad (\text{F.6})$$

obeys  $\Gamma(n+1) = n!$ . To prove this freaky fact, note that  $\Gamma(z)$  can be integrated by parts (cf. Sec. C.3.2) with  $u = t^z$ ,  $v = -e^{-t}$  (and thus  $dv = -e^{-t}$ ):

$$\begin{aligned} \Gamma(z) &= -t^z e^{-t} \Big|_0^\infty - \int_0^\infty (-e^{-t}) z t^{z-1} \\ &= z \Gamma(z-1) \end{aligned} \quad (\text{F.7})$$

and, since  $\Gamma(0+1) = \int dt e^{-t} = 1 = (0)!$  we see that  $\Gamma(z)$  is the same function we defined algorithmically in Sec F.

Our minds sufficiently blown, we now approximate the integral by pretending it's a Gaussian, i.e, defining  $t_*$  such that

$$f(t, z) = t^z e^{-t} \quad (\text{F.9})$$

$$f' = t^{z-1} e^{-t} (z-t) \quad (\text{F.10})$$

$$f'' = t^{z-2} e^{-t} (z^2 - z - 2zt + t^2) \quad (\text{F.11})$$

$$f'(t_*) \equiv 0 \Rightarrow t_* = z \quad (\text{F.12})$$

$$f''(t_*) = -z^{z-1} e^{-z} \quad (\text{F.13})$$

$$f(t_*) = z^z e^{-z} \quad (\text{F.14})$$

$$f(t_*)/f''(t_*) = -z \quad (\text{F.15})$$

Approximating via the trick we learned in Sec. C.3.3,

$$\Gamma(z+1) = \int_0^\infty dt f(t, z) \quad (\text{F.16})$$

$$\approx f(t_*, z) / \sqrt{\left| \frac{f''(t_*, z)}{2\pi f(t_*, z)} \right|} \quad (\text{F.17})$$

$$= z^z e^{-z} \sqrt{2\pi z}. \quad (\text{F.18})$$

Easy as pie.

### F.3 Application to the binomial

#### F.3.1 The binomial is nearly a Gaussian $q = 1/2$ , $j \approx N/2$

Consider for the moment  $a!/(b!c!)$ , approximated by Stirling:

$$\frac{a!}{b!c!} \approx \frac{a^a e^{-a} \sqrt{2\pi a}}{(b^b e^{-b} \sqrt{2\pi b})(c^c e^{-c} \sqrt{2\pi c})} = \left( \frac{a^a}{b^b c^c} \right) \frac{1}{\sqrt{2\pi}} \left( \frac{a}{bc} \right)^{1/2} e^{c+b-a}. \quad (\text{F.19})$$

In the special case where  $a = b + c$ , the final term vanishes. Consider the first (dominant) term in the expansion, rewriting  $b = a/2(1 + \delta)$ ,  $c = a/2(1 - \delta)$ :

$$\begin{aligned}\ln\left(\frac{a^a}{b^b c^c}\right) &= a \ln a - b \ln b - c \ln c \\ &= (a/2) \ln a^2 - (a/2)(1 + \delta) \ln(a/2)(1 + \delta) - (a/2)(1 - \delta) \ln(a/2)(1 - \delta) \\ &= (a/2) (\ln a^2 - \ln(a^2/4)(1 - \delta^2) - \delta \ln(1 + \delta) + \delta \ln(1 - \delta)) \\ &= a \ln 2 - (a/2) (\ln(1 - \delta^2) + \delta \ln(1 + \delta) - \delta \ln(1 - \delta)).\end{aligned}$$

Recalling the first term in the Taylor expansion  $\ln 1 + x \approx x$ , we have

$$\ln\left(\frac{a^a}{b^b c^c}\right) \approx a \ln 2 - (a/2) (-\delta^2 + \delta^2 + \delta^2) = a \ln 2 - (a\delta^2/2) \quad (\text{F.20})$$

$$\left(\frac{a^a}{b^b c^c}\right) \approx 2^a e^{-\frac{1}{2}a\delta^2} \quad (\text{F.21})$$

The final term is approximated

$$\sqrt{a/(bc)} = \sqrt{\frac{4}{a(1 - \delta^2)}} \approx 2a^{-1/2} \quad (\text{F.22})$$

So that we have the approximate relation

$$\binom{a}{b} \approx \sqrt{\frac{2}{\pi a}} e^{-\frac{1}{2}a\delta^2} 2^a \quad (\text{F.23})$$

**F.3.2 The binomial is nearly Poisson:**  $q \ll 1$ ,  $qN \sim 1$   $j \ll N$ 

Now consider the case of “rare events”, when  $q \ll 1$ . Defining  $qN \equiv \lambda$ , we have

$$\binom{N}{j} q^j (1-q)^{N-j} = \frac{N^N e^{-N} \sqrt{2\pi N}}{j! (N-j)^{(N-j)} e^{-N-j} \sqrt{2\pi(N-j)}} (\lambda/N)^j (1-\lambda/N)^{(N-j)} \quad (\text{F.24})$$

$$= \frac{1}{j!} \frac{N^N}{N^{(N-j)}} \frac{e^{-N}}{e^{-N-j}} (1-j/N)^{-N+j-1/2} \lambda^j N^{-j} e^{-\lambda} (1-\lambda/N)^{-j} \quad (\text{F.25})$$

$$= \frac{\lambda^j}{j!} e^{-\lambda} (1-j/N)^{j-1/2} (1-\lambda/N)^{-j} \\ \approx \frac{\lambda^j}{j!} e^{-\lambda} \quad (\text{F.26})$$

# Bibliography

- [1]
- [2]
- [3] the machinery of life.
- [4] Tractatus de quadratura curvarum, 1693. published in 1704 as an appendix to Newton's Optiks.
- [5] Biological Physics: Energy, Information, Life. W. H. Freeman and Co., 2004.
- [6] Howard C. Berg. Random Walks in Biology. Princeton University Press, 1983.
- [7] Peter Dayan and Lawrence F. Abbott. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. MIT Press, Cambridge, MA, 2001.
- [8] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. Nature, 403, 2000.
- [9] Nir Friedman. Inferring cellular networks using probabilistic graphical models. Science, 303, 2004.
- [10] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in escherichia coli. Nature, 403, 2000.
- [11] D. Heckerman. A tutorial on learning with bayesian networks, 1995.
- [12] E.F Keller. Revisiting ‘scale-free’ networks.
- [13] Fred Rieke, David Warland, Rob de Ruyter van Steveninck, and William Bialek. Spikes: Exploring the Neural Code. MIT Press, Cambridge, MA, 1996.

- [14] Gideon Schwartz. Estimating the dimension of a model. Ann. Stat., 6, 1978.