Modeling Gene Regulatory Network

Mahroo Hajimehdi

Jan 2022

Overview of the main concepts

1- Regulation of Gene Expression (a biological approach)

2- System Equation

2-1 Stoichiometric Matrix

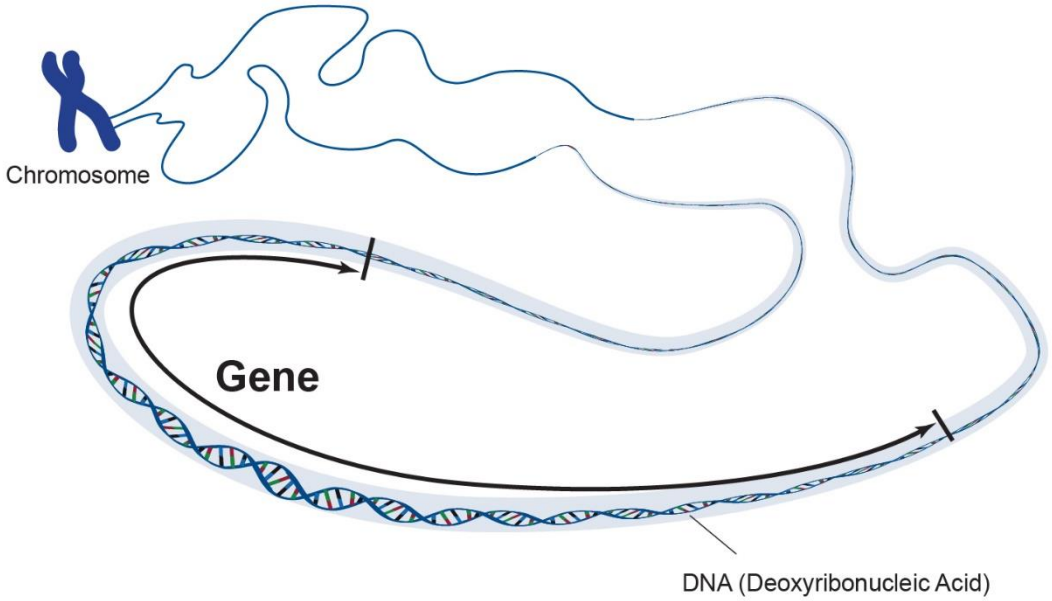
2-2 Information Encoded in the Stoichiometric Matrix N

3- Solving ODEs via Matrices

4- Hopf Bifurcations: Oscillatory Gene Expression

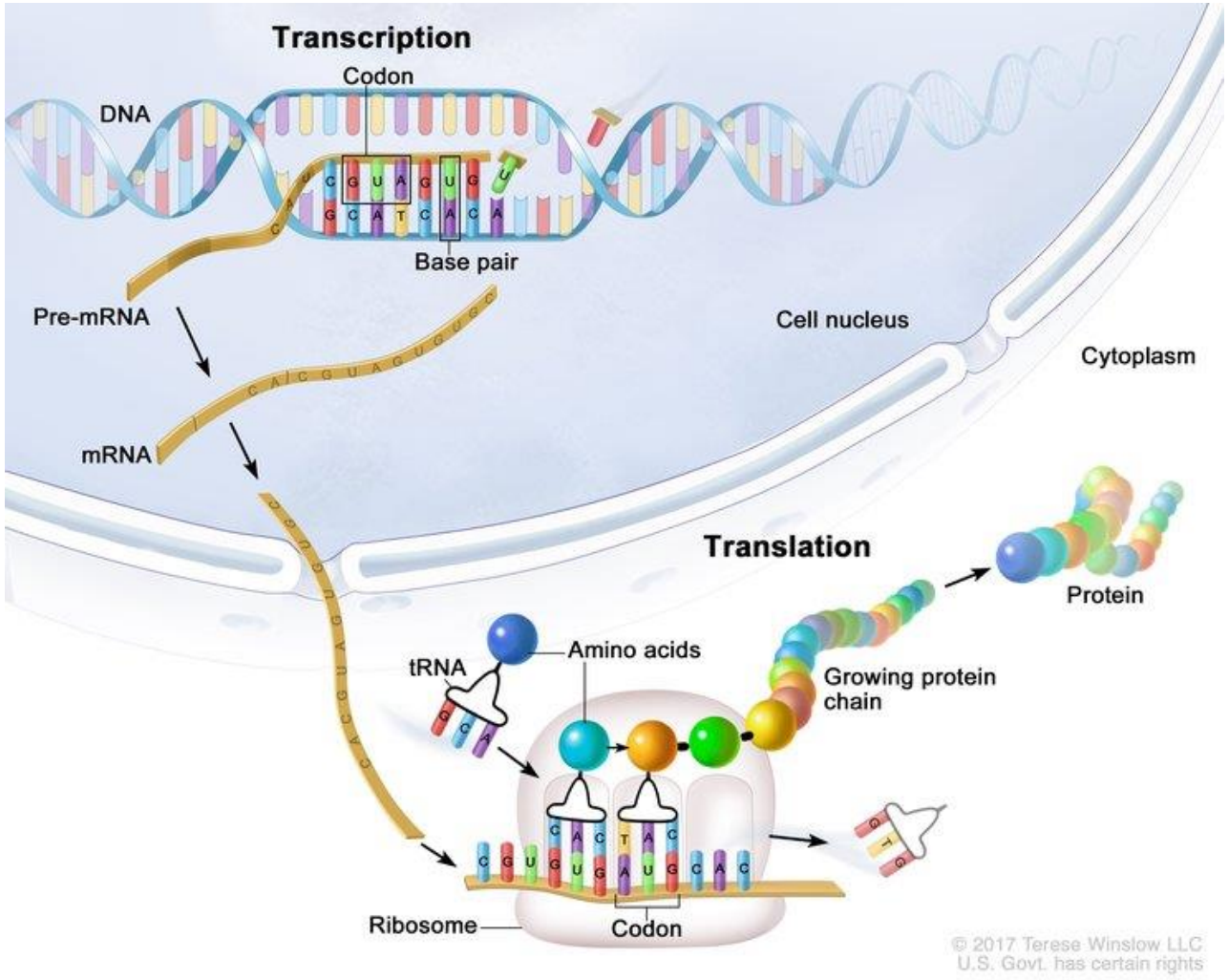
Regulation of Gene Expression
(a biological approach)

Classically, a **gene** is defined as the information encoded by the sequence of a DNA region that is required for the construction of a protein.



The term **gene expression** commonly refers to the whole process during which the information of a particular gene is translated into a particular protein. This process involves several steps:

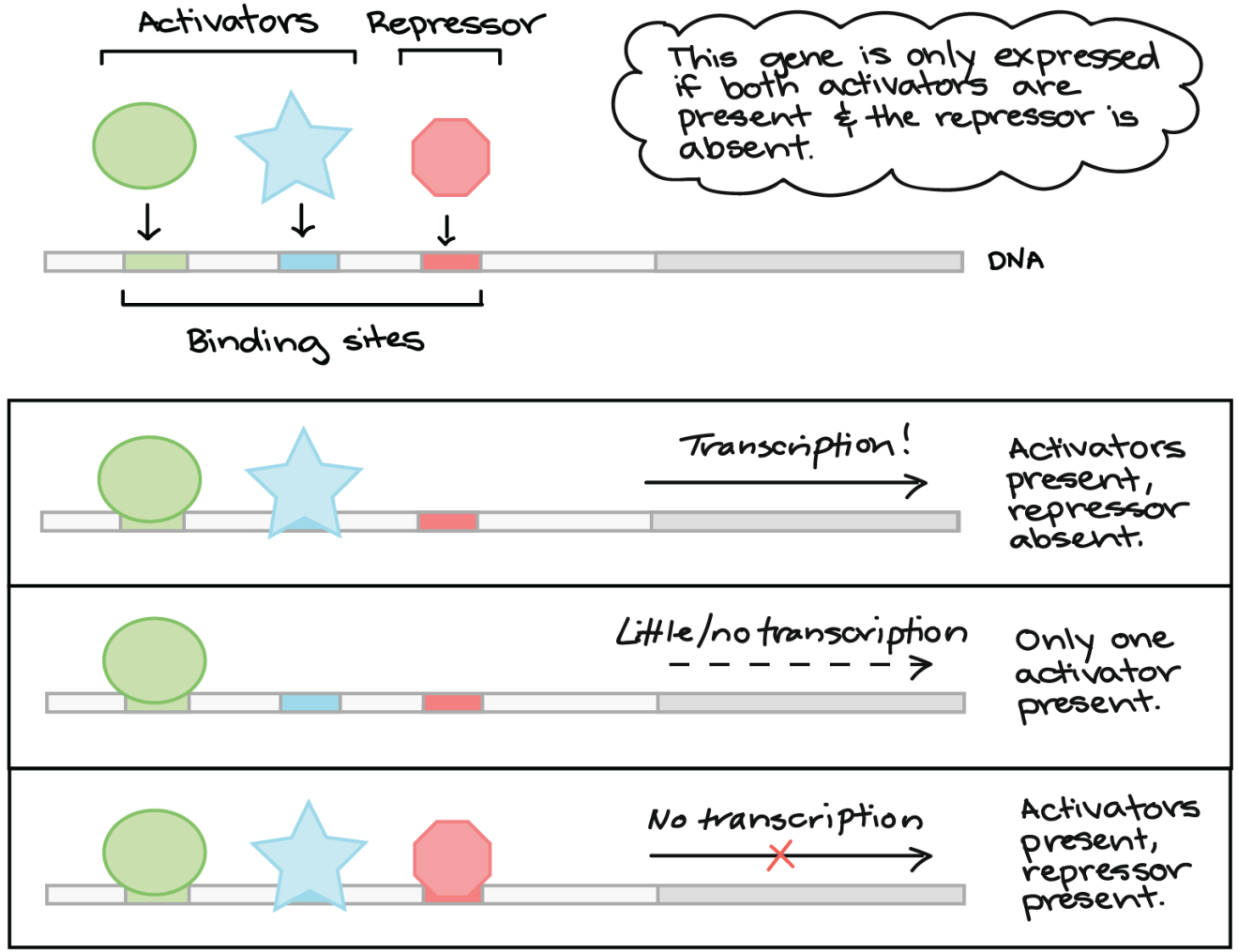
1. Transcription
2. mRNA processing (in eukaryotic cells)
3. Translation



Gene expression in eukaryotes is controlled at six different steps:

1. **Transcriptional** control: when and how often is a gene transcribed.
2. **RNA processing** control: how is the RNA transcript spliced.
3. **RNA transport and localization** control: which mRNAs in the nucleus are exported to cytosol and where in the cytosol are they localized.
4. **Translational** control: which mRNAs in the cytosol are translated by ribosomes
5. **mRNA degradation** control: which mRNAs in the cytosol are destabilized.
6. **Protein activity** control: decide upon activation, inactivation, compartmentalization, and degradation of the translated protein.

The expression level of the majority of genes is controlled by **transcription factors**. **Transcription factors** are proteins that **bind to DNA regulatory sequences** *upstream* of the site at which transcription is initiated.



Advances in microarray and assay technologies are facilitating increasingly large amounts of laboratory data for analysis of these networks.

If the network is operating sufficiently close to a steady-state, have shown that multiple linear regressions can be applied to this data to derive **a linear ordinary differential equation (ODE)** model of the form:

$$\dot{x} = Ax + Bu$$

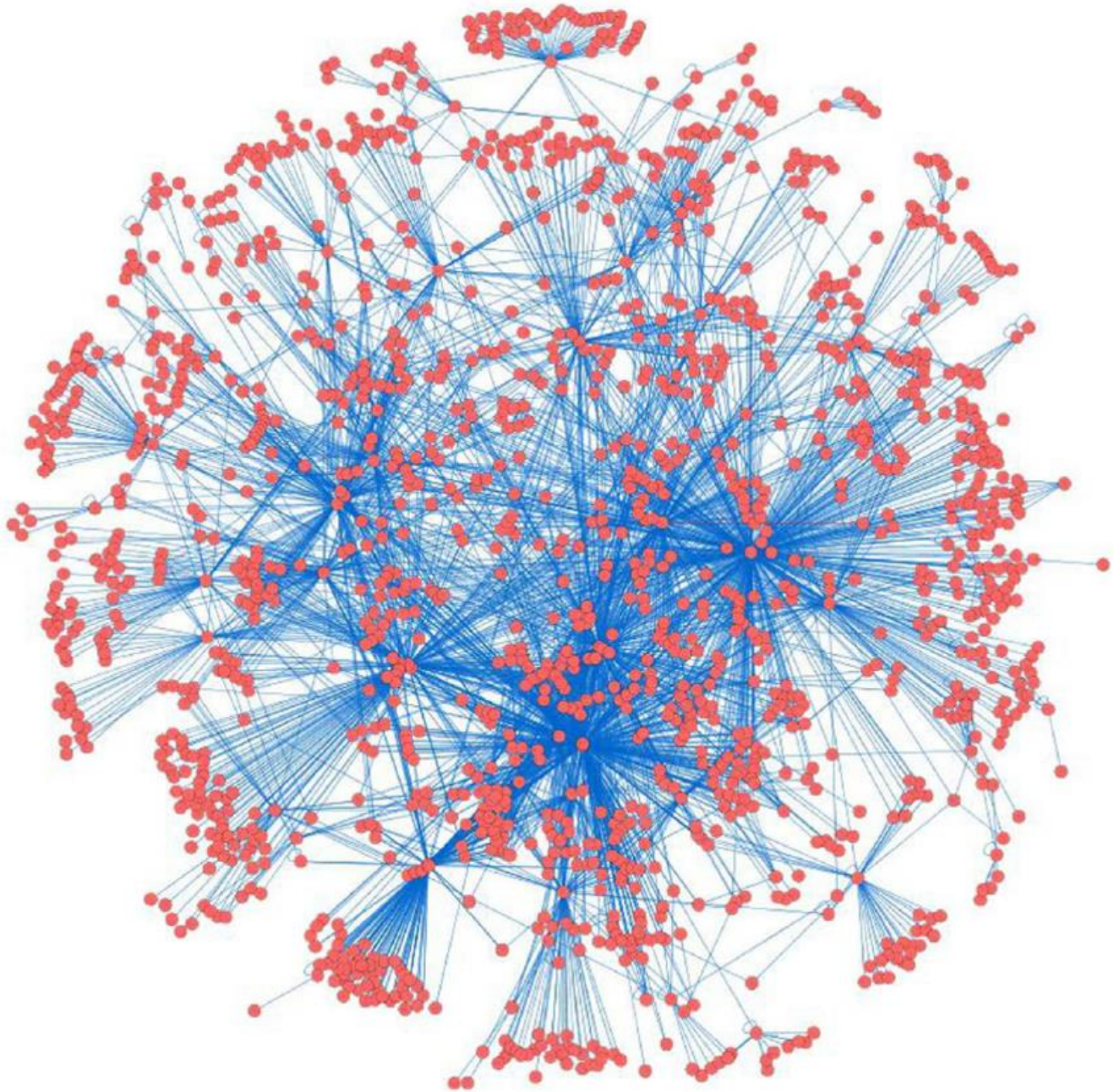
where *A* and *B* are real-valued matrices of suitable sizes, *x* is the vector of gene expression values, and *u* is the vector (or matrix) of exciting inputs.

Steady-state?
When a reaction involves one or more intermediates, if **concentration of one of the intermediates remains constant** at some stage of the reaction, the system has reached a steady-state

We can describe the dynamics of the TF binding network by a set of ordinary differential equations.

Back

Activator TF	Inhibitor TF
$TF + DNA \underset{k_2}{\overset{k_1}{\rightleftharpoons}} TFDNA \overset{k_3}{\rightarrow} P + TFDNA$	$TF + DNA \underset{k_2}{\overset{k_1}{\rightleftharpoons}} TFDNA$ $DNA \overset{k_3}{\rightarrow} P + DNA$
$\frac{d[TF]}{dt} = -k_1[TF][DNA] + k_2[TFDNA]$	$\frac{d[TF]}{dt} = -k_1[TF][DNA] + k_2[TFDNA]$
$\frac{d[DNA]}{dt} = -k_1[TF][DNA] + k_2[TFDNA] + k_3[TFDNA]$	$\frac{d[DNA]}{dt} = -k_1[TF][DNA] + k_2[TFDNA]$
$\frac{d[TFDNA]}{dt} = k_1[TF][DNA] - k_2[TFDNA]$	$\frac{d[TFDNA]}{dt} = k_1[TF][DNA] - k_2[TFDNA]$
$\frac{d[P]}{dt} = k_3[TFDNA]$	$\frac{d[P]}{dt} = k_3[DNA]$



Stoichiometric Matrix

Stoichiometric coefficients denote the proportion of substrate and product molecules involved in a reaction. For example, for the reaction

$$S_1 + S_2 \rightarrow 2P$$

the stoichiometric coefficients of S_1, S_2, and P are 1, 1, and 2, respectively.

The change of concentrations in time can be described using ODEs:

$$\frac{dS_1}{dt} = -v, \qquad \frac{dS_2}{dt} = -v \quad and \quad \frac{dP}{dt} = 2v$$

This means that the decay of S_1 with rate v is accompanied by the decay of S_2 with the same rate and by the production of P with the double rate.

For a metabolic network consisting of *m* substances and *r* reactions, the system dynamics is described by the system equations:

$$\frac{dS_i}{dx} = \sum_{j=1}^{r} n_{ij} v_j \quad i=1,2,\ldots,m$$

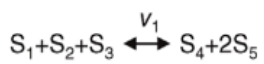
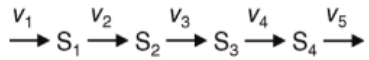
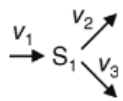
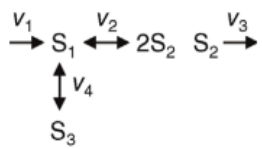
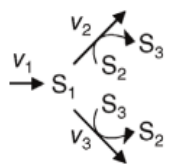
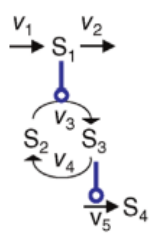
The quantities n_{ij} are the stoichiometric coefficients of the *ith* metabolite in the *jth* reaction. The stoichiometric coefficients n_{ij} assigned to the compounds S_i and the reactions v_j can be combined into the **stoichiometric matrix N** with

$$\mathbf{N} = \{n_{ij}\} \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, r,$$

$$\mathbf{N} = \begin{array}{c} \begin{matrix} v_1 & v_2 & \cdots & v_r \end{matrix} \\ \begin{pmatrix} n_{11} & n_{12} & \cdots & n_{1r} \\ n_{21} & n_{22} & \cdots & n_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ n_{m1} & n_{m2} & \cdots & n_{mr} \end{pmatrix} \end{array} \begin{matrix} S_1 \\ S_2 \\ \vdots \\ S_m \end{matrix},$$

↓ each *column* belongs to a reaction

→ each *row* to a compound.

	Network	Stoichiometric matrix	ODE system
N1		$\mathbf{N} = \begin{pmatrix} -1 \\ -1 \\ -1 \\ 1 \\ 2 \end{pmatrix}$	$\frac{dS_1}{dt} = \frac{dS_2}{dt} = \frac{dS_3}{dt} = -v_1$ $\frac{dS_4}{dt} = v_1$ $\frac{dS_5}{dt} = 2v_1$
N2		$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$	$\frac{dS_1}{dt} = v_1 - v_2$ $\frac{dS_2}{dt} = v_2 - v_3$ $\frac{dS_3}{dt} = v_3 - v_4$ $\frac{dS_4}{dt} = v_4 - v_5$
N3		$\mathbf{N} = \begin{pmatrix} 1 & -1 & -1 \end{pmatrix}$	$\frac{dS_1}{dt} = v_1 - v_2 - v_3$
N4		$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & -1 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	$\frac{dS_1}{dt} = v_1 - v_2 - v_4$ $\frac{dS_2}{dt} = 2v_2 - v_3$ $\frac{dS_3}{dt} = v_4$
N5		$\mathbf{N} = \begin{pmatrix} 1 & -1 & -1 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix}$	$\frac{dS_1}{dt} = v_1 - v_2 - v_4$ $\frac{dS_2}{dt} = -v_2 + v_3$ $\frac{dS_3}{dt} = v_2 - v_3$
N6		$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\frac{dS_1}{dt} = v_1 - v_2$ $\frac{dS_2}{dt} = v_4 - v_3$ $\frac{dS_3}{dt} = -v_4 + v_3$ $\frac{dS_4}{dt} = v_5$

the mathematical description of the metabolic system consists of a vector $S = (S_1, S_2, \dots, S_n)^T$ of concentration values, a vector $v = (v_1, v_2, \dots, v_r)^T$ reaction rates, a parameter vector $p = (p, p_2, \dots, p_m)^T$, and the stoichiometric matrix **N**.

If the system is in steady state, we can also consider the vector $J = (J_1, J_2, \dots, J_r)^T$ containing the steady-state fluxes.

With these notions, the balance equation reads:

$$\frac{dS}{dx} = Nv$$

Flux?
Flux or metabolic flux is the rate of **turnover** of molecules through a metabolic pathway.

Turnover rate?
The rate of substrate molecules **transformed** per minute by a single enzyme molecule.

2-2 Information Encoded in the Stoichiometric Matrix N

The stoichiometric matrix contains important information about the structure of the metabolic network;

- Calculate which combinations of individual fluxes are possible in steady state (i.e., calculate the admissible steady-state flux space)
- Find out dead ends and unbranched reaction pathways
- Discover the conservation relations for the included reactants.

In steady state, it holds that

$$\frac{dS}{dx} = Nv = 0$$

Note that 0 is a vector with length n, that is, $0 = (0, 0, \ldots, 0)^T$

By definition, the set

$$\mathcal{N}(N) = \{u \mid Nu = 0\}$$

Contains all valid flux vector

The null space is a linear vector space, so all properties of linear vector spaces follow,

- Contains the zero vector, and closed under linear combination
- It has a basis $\{k_i, \ldots, k_{r-Rank(N)}\}$

This equation has nontrivial solutions only for Rank(N) < r , A kernel matrix K fulfilling

$$NK = 0$$

expresses the respective linear dependencies between the columns of the stoichiometric matrix. K consists of r - Rank(N) basis vectors as columns and can be determined using the Gauss algorithm.

Every possible set J of steady-state fluxes can be expressed as linear combination of the columns k_i of K:

$$J = \sum_{j=1}^{r-Rank(N)} a_i k_i$$

The coefficients must have units corresponding to the units of reaction rates (e.g., mMs^{-1}).

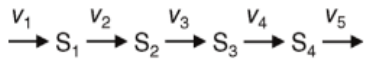
Three properties of the metabolic network can be found directly from the kernel matrix

1. Dead ends in metabolism (reactions that cannot carry a flux in any steady state): correspond to identically zero rows in the kernel
2. Enzyme subsets (reactions that are forced to operate in lock step in any steady state): correspond to kernel rows that are scalar multiples of each other
3. Independent components (groups of reactions that can carry flux independently from reactions outside the group): block-diagonal structure in the kernel

Dead end?

A dead-end metabolite (DEM) is defined as **a metabolite that is produced by the known metabolic reactions of an organism and has no reactions consuming it**.

N2

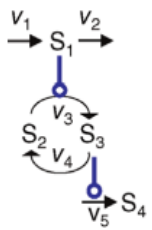


$$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

$$\frac{dS_1}{dt} = v_1 - v_2$$
$$\frac{dS_2}{dt} = v_2 - v_3$$
$$\frac{dS_3}{dt} = v_3 - v_4$$
$$\frac{dS_4}{dt} = v_4 - v_5$$

For the network N2 in Table 3.1, we have $r = 5$ reactions and $\text{Rank}(\mathbf{N}) = 4$. The kernel matrix contains just $1 = 5 - 4$ basis vectors, which are multiples of $\mathbf{k}_1 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix}^T$. This means that in steady state the flux through all reactions must be equal.

N6



$$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\frac{dS_1}{dt} = v_1 - v_2$$
$$\frac{dS_2}{dt} = v_4 - v_3$$
$$\frac{dS_3}{dt} = -v_4 + v_3$$
$$\frac{dS_4}{dt} = v_5$$

Network N6 can present a small signaling cascade. It has five reactions and $\text{Rank}(\mathbf{N}) = 3$. Two basis vectors of the kernel are

$$\mathbf{k}_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \end{pmatrix}^T,$$
$$\mathbf{k}_2 = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \end{pmatrix}^T. \tag{3.11}$$

If we calculate the possible steady-state fluxes according to Eq. (3.10), we can easily see that in every steady state it holds that production and degradation of S_1 are balanced ($J_1 = J_2$) and that the fluxes through the cycle are equal ($J_3 = J_4$). In addition, J_5 must be equal to zero, otherwise S_4 would accumulate. One could prevent the last effect by also including the degradation of S_4 into the network.

Solving ODEs via Matrices

Flashback!

Consider the following example:

$$\frac{dx_1}{dt} = -2x_1 + 2x_2 \quad , \quad \frac{dx_2}{dt} = 2x_1 + x_2$$

1. Find the critical points.

$$\begin{cases} -2x_1 + 2x_2 = 0 \\ 2x_1 + x_2 = 0 \end{cases} \rightarrow \quad x_2 = 0\,, x_1 = 0$$

2. Write the equation in matrix form: $\frac{dx}{dt} = Ax(t)$

$$A = \begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix} \quad , \quad x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

$$\frac{dx}{dt} = \begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

3 Find eigenvalues.

$$Ax = \lambda x \rightarrow Ax - \lambda x = 0 \rightarrow (A - I\lambda)x = 0 \xrightarrow{(A-I\lambda)=D} Dx = 0$$

$$\det(A - I\lambda) = 0$$

$$\det\begin{vmatrix} -2-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} = 0$$

$$\lambda^2 - \tau\lambda + \det(A) = 0$$

$$\lambda^2 + \lambda - 6 = 0$$

$$\lambda_1 = 2\,, \lambda_1 = -3$$

4 Find eigenvectors.

$$\begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \lambda \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\lambda = 2 \quad \begin{cases} -2u_1 + 2u_2 = 2u_1 \\ 2u_1 + u_2 = 2u_2 \end{cases} \rightarrow \quad u = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$\lambda = -3 \quad \begin{cases} -2u_1 + 2u_2 = -3u_1 \\ 2u_1 + u_2 = -3u_2 \end{cases} \rightarrow \quad u = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$$

5 Find the ODE system solution.

$$x(t) = c_1 e^{2t} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + c_2 e^{-3t} \begin{pmatrix} 1 \\ 5 \end{pmatrix}$$

$$x_1(t) = c_1 e^{2t} + c_2 e^{-3t}$$

$$x_2(t) = 2c_1 e^{2t} + 5c_2 e^{-3t}$$

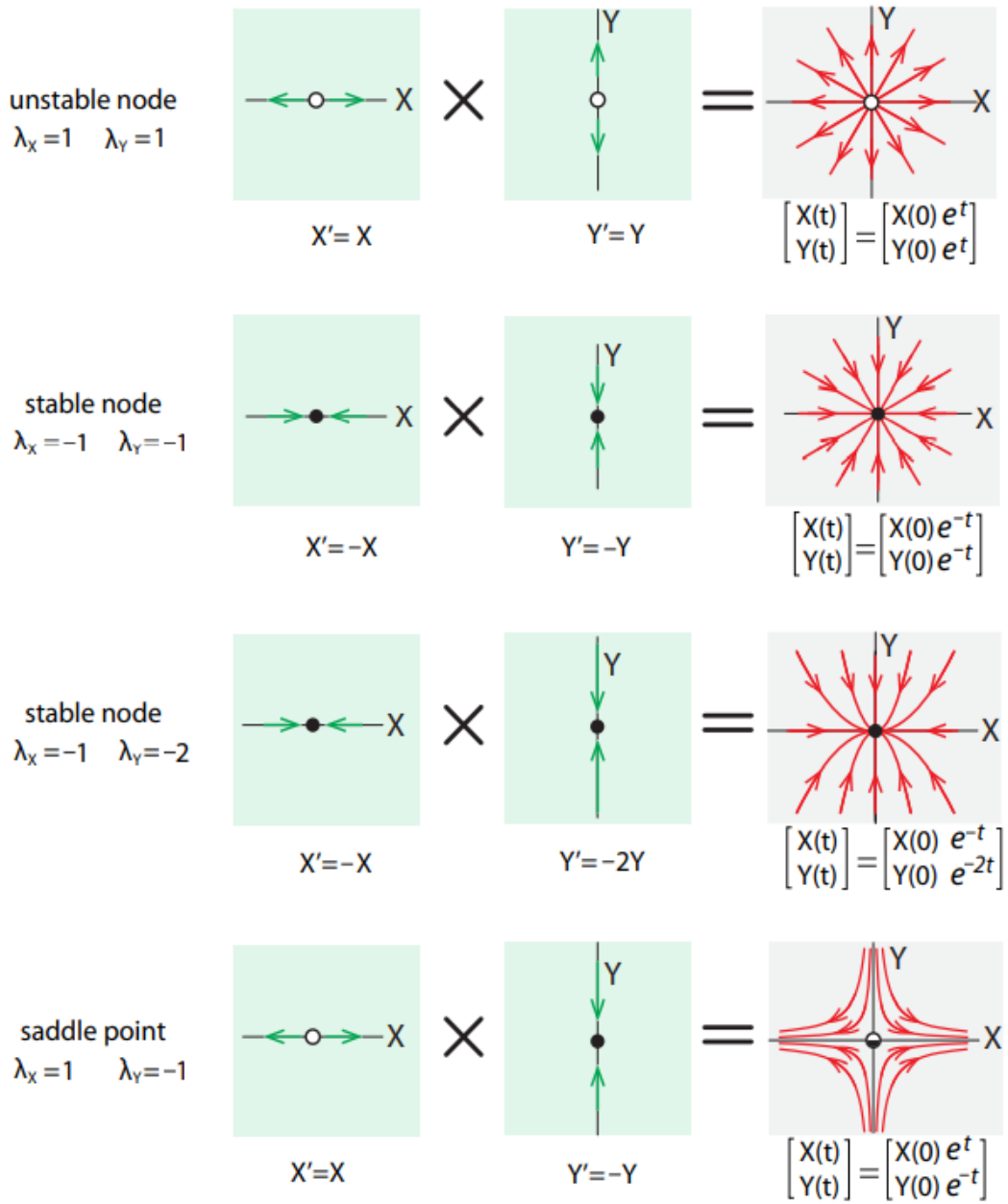
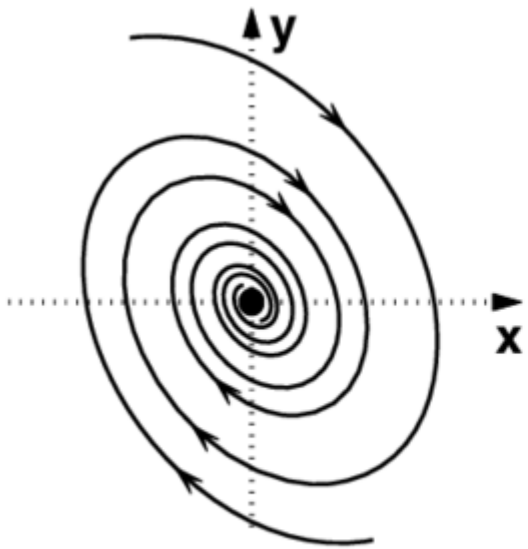


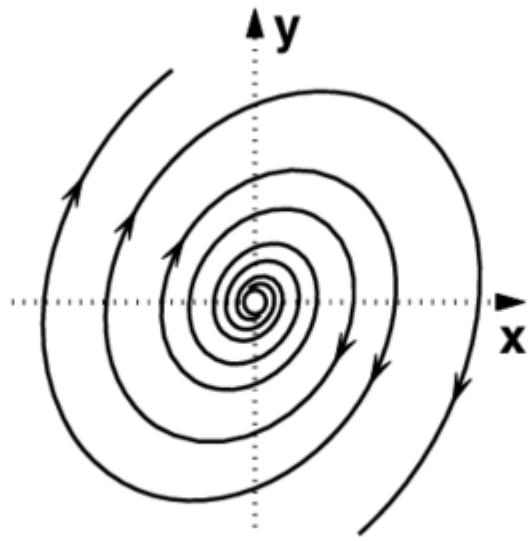
Figure 6.39: Equilibrium points and flows in 2D uncoupled systems.



stable spiral

$\lambda = \alpha \pm i\beta$

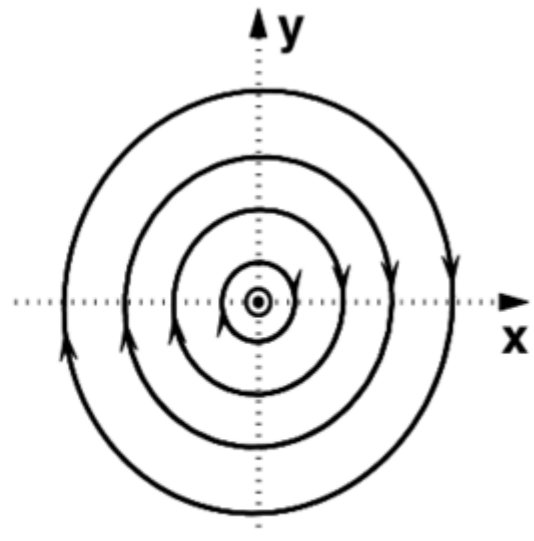
$\alpha < 0$



unstable spiral

$\lambda = \alpha \pm i\beta$

$\alpha > 0$



center

$\lambda = \alpha \pm i\beta$

$\alpha = 0$

4 Hopf Bifurcations: Oscillatory Gene Expression

Genes are often under regulation that causes them to express in an oscillatory pattern, with cycles ranging from hours to days. Oscillatory gene expression has been detected in many genes, including Hes1, which is critical in neural development, and p53, the "guardian angel gene," which is critical in cancer regulation.

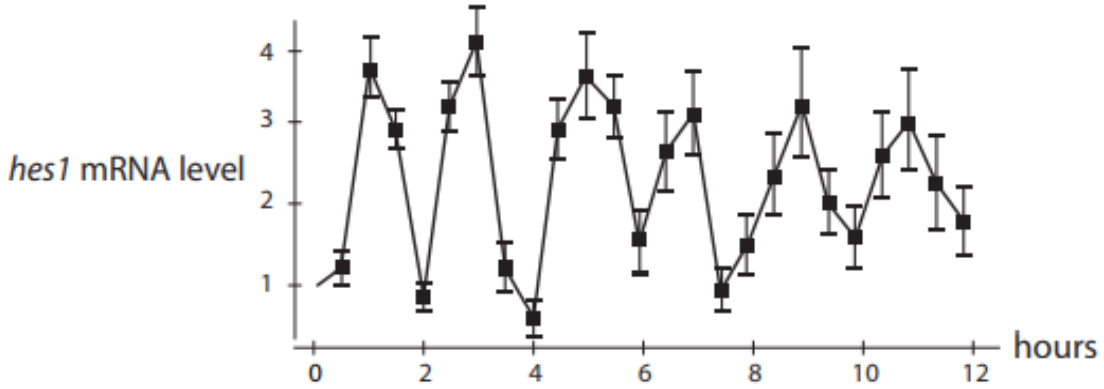


Figure 4.4: Two-hour oscillations in the expression of the gene *Hes1*. Redrawn from "Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop," by H. Hirata, S. Yoshiura, T. Ohtsuka, Y. Bessho, T. Harada, K. Yoshikawa, and R. Kageyama, (2002), *Science* 298(5594):840–843. Reprinted with permission from AAAS.

Rhythmic gene expression has to be coordinated to, and in some cases actually drive, these rhythmic processes. Therefore, cells have evolved mechanisms to produce oscillatory gene expression. Most of these mechanisms depend on some kind of **negative feedback**, where the gene produces a product that **inhibits** that very gene.

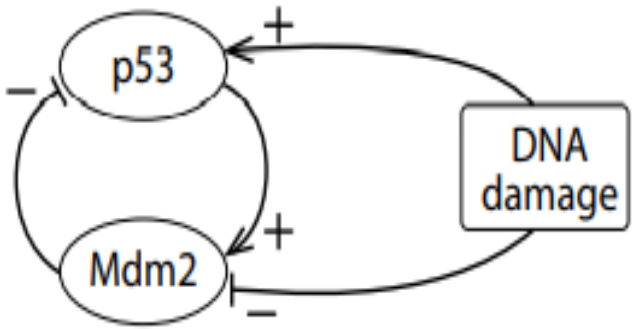


Figure 4.24: Negative feedback in the p53-mdm2 system.

The occurrence of oscillations is generally caused by the presence of a **negative feedback loop** in the regulatory network.

Hopf bifurcation theorem (approximately)

Consider an equilibrium point of a vector field that depends on a parameter. Let J be the Jacobian matrix representing the linear approximation to the vector field at that equilibrium point. Suppose that a pair of conjugate eigenvalues of J, $a \pm bi$ passes from $a < 0$ to $a > 0$ as a parameter passes a critical value. In this case, the behavior changes from a stable equilibrium to an unstable equilibrium surrounded by a stable limit cycle attractor.

$$A' = \frac{kA^2}{A^2 + 10(1 + \frac{R}{0.2})} - A + 0.4$$
$$R' = \frac{0.3A^2}{A^2 + 10(1 + \frac{R}{0.2})} - 0.2R$$

We will use k as our control parameter. The Jacobian matrix can be expressed in terms of A, R, and k:

$$M = \begin{bmatrix} -2kA^3b^2 + 2kAb - 1 & -50A^2kb^2 \\ 0.6Ab - 0.6A^3b^2 & -15A^2b^2 - 0.2 \end{bmatrix} \quad \text{where } b = \frac{1}{A^2 + 10(1 + \frac{R}{0.5})}$$

Because of the complexity of this model, the only way to study the system is by plugging different k values into the system and calculating the corresponding equilibrium points and the Jacobian matrix around that equilibrium point to determine its stability.

First of all, let's find the equilibrium points when $k = 9.5$. Solving $A' = R' = 0$, we get

$$(A, R)\big|_{k=9.5} = (1, 0.1)$$

Plugging in the k value as well as the equilibrium point, we get the Jacobian matrix

$$M\big|_{k=9.5} = \begin{bmatrix} 0.14 & -1.9 \\ 0.036 & -0.26 \end{bmatrix}$$

The corresponding eigenvalues are solutions to

$$det\left(M\big|_{k=9.5} - \lambda\mathbb{I}\right) = 0 \Longrightarrow \lambda = -0.6 \pm 0.17\,\boldsymbol{i}$$

These are complex conjugate eigenvalues with negative real part. Therefore, this equilibrium point is a stable spiral (Figure 7.42).

Now let's consider the case $k = 10.5$. The equilibrium points can be found by setting $A' = R' = 0$. We get

$$(A, R)\big|_{k=10.5} = (2.5, 0.3)$$

Similarly, plugging in the k value as well as the equilibrium point, we get the Jacobian matrix

$$M\big|_{k=10.5} = \begin{bmatrix} 0.34 & -3.4 \\ 0.038 & -0.3 \end{bmatrix}$$

And the corresponding eigenvalues are

$$det\left(M\big|_{k=10.5} - \lambda \mathbb{I}\right) = 0 \Longrightarrow \lambda = +0.024 \pm 0.16\, \boldsymbol{i}$$

which are complex conjugate eigenvalues with positive real part. Therefore, this equilibrium point when $k = 10.5$ is an unstable spiral. And by the Hopf bifurcation theorem, there is a stable limit cycle attractor surrounding the equilibrium point (Figure 7.42).

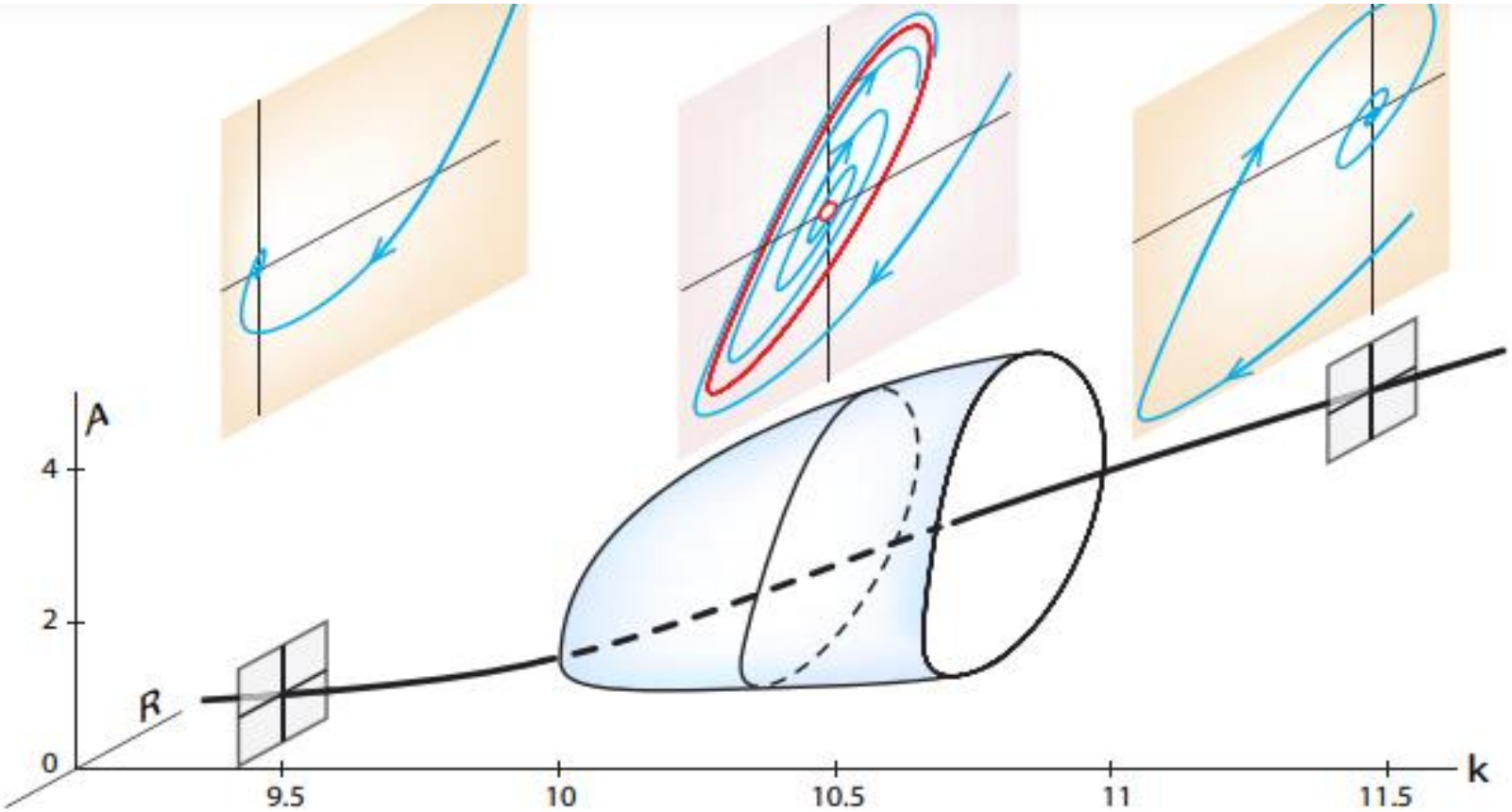


Figure 7.42: A 3D Hopf bifurcation diagram for the gene expression model.

References

[1] Klipp, E., Liebermeister, W., Wierling, C., & Kowald, A. (2016). *Systems Biology: A Textbook* (2nd ed.). Wiley-Blackwell.

[2] Garfinkel, A., Shevtsov, J., & Guo, Y. (2017). *Modeling Life: The Mathematics of Biological Systems* (1st ed. 2017 ed.). Springer.

[3] Li, W., Krishna, S., Pigolotti, S., Mitarai, N., & Jensen, M. H. (2012). Switching between oscillations and homeostasis in competing negative and positive feedback motifs. *Journal of Theoretical Biology*, *307*, 205–210. https://doi.org/10.1016/j.jtbi.2012.04.011

[4] Alon, U. (2019). *An Introduction to Systems Biology: Design Principles of Biological Circuits (Chapman & Hall/CRC Computational Biology Series)* (2nd ed.). Chapman and Hall/CRC.

[5] Neuhauser, C. (2022). *Calculus For Biology and Medicine (Calculus for Life Sciences Series) 3th (third) edition*. Pearson.

[6] Vignoni, A. (2020, June). *Modeling: ODEs and Hill Functions Section 1: ODEs, Law of mass action and the central dogma* [Poster session]. Synthetic Biology and BiosystemsControl Lab, Valencia, Valencia.