Null Space, Subsets and Elementary Modes of Metabolic Networks

Prof Cristiana Sebu

University of Malta, Malta

Vectors and Vector Operations in $\mathbf{R}^n$

Definition 1.1.1 If n is a positive integer, then an ordered n-tuple is a sequence of real numbers $(x_1, x_2, \ldots, x_n)$. We denote the set of all such n-tuples by $\mathbf{R}^n$.

Let $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ be vectors in $\mathbf{R}^n$:

- ***Equality***: $\mathbf{u} = \mathbf{v} \quad \Leftrightarrow u_1 = v_1,\ u_2 = v_2, \ldots, u_n = v_n$
- ***Sum***: $\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n)$
- ***Scalar multiplication***: $k\mathbf{u} = (ku_1, ku_2, \ldots, ku_n),\ k \in \mathbf{R}$

The operations of addition are scalar multiplication defined above are called the ***standard operations*** on $\mathbf{R}^n$.

The zero vector in $\mathbf{R}^n$: $\mathbf{0} = (0, 0, \ldots, 0)$

If $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ is any vector in $\mathbf{R}^n$, the negative (or additive inverse) of $\mathbf{u}$ is

$$-\mathbf{u} = (-u_1, -u_2, \ldots, -u_n)$$

- ***Difference***: $\mathbf{v} - \mathbf{u} = \mathbf{v} + (-\mathbf{u}) = (v_1 - u_1, v_2 - u_2, \ldots, v_n - u_n)$

Example 1.1.2 Consider the following vectors in $\mathbf{R}^5$ $\mathbf{u} = (2, 3, 1, 0, 5)$ and $\mathbf{v} = (4, -1, 2, 1, 0)$, then evaluate $\mathbf{u} + \mathbf{v}$, $\mathbf{u} - \mathbf{v}$, $\mathbf{u} + \mathbf{v}$, $3\mathbf{u}$, $2\mathbf{u} - 3\mathbf{v}$.

Homogeneous Linear Equations

Definition 2.1 A **linear equation** in the n variables $x_1, x_2, \ldots, x_n$ can be expressed as:

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b,$$

The variables in a linear equation sometimes are called **unknowns**.

A **solution** of a linear equation is a set of numbers $s_1, s_2, \ldots, s_n$ such that the equation is satisfied when we substitute $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$. The set of all solutions is called its **solution set** or sometimes the **general solution** of the equation.

If $b = 0$, then the linear equation is **homogeneous**.

Example 2.2: Find the solution set of the following equations:

(i) $4x - 2y = 1$

(ii) $x_1 - 4x_2 + 7x_3 = 0$

Definition 2.3 A set of homogeneous linear equations in the variables $x_1, x_2, \ldots, x_n$ is called a **homogeneous linear system**:

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0$$
$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0$$
$$\cdots$$
$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0$$

A sequence of numbers $s_1, s_2, \ldots, s_n$ is called a **solution** of the system if $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ is a solution of every equation in the system.

Solutions
- Always at least one: $x_1 = x_2 = \ldots = x_n = 0$
- Can have infinitely many

Remark 2.4 A homogeneous system of linear equations with more unknowns than equations has infinitely many solutions.

The **augmented matrix** of a homogeneous linear system has the following form:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & 0 \\ a_{21} & a_{22} & \dots & a_{23} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & 0 \end{pmatrix}$$

Gaussian Elimination

The basic method for solving a system of linear equations is to replace the given system by a new system that has the same solution set but which is easier to solve by carrying out **elementary row operations** on the augmented matrix:

1. Interchange any two rows

2. Add a multiple of one row to another

3. Multiply every element in a row by a fixed number

Elementary row operations do not actually change the solutions to the equations.

The aim is to reduce the matrix to **row-echelon** form:

1. If a row does not consist entirely of zeros, then the first nonzero number in the row is a 1 (called **leading 1**).

2. If there are any rows that consist entirely of zeros, then they are grouped together at the bottom of the matrix.

3. In any two successive rows that do not consist entirely of zeros, the leading 1 in the lower row occurs farther to the right then the leading 1 in the higher row.

Definition 2.5 The step-by-step procedure to reduce any matrix to row-echelon form is called **Gaussian Elimination**.

Example 2.6 Reduce the following matrix to row-echelon form

$$\begin{pmatrix} 2 & 2 & -1 & 0 & 1 & 0 \\ -1 & -1 & 2 & -3 & 1 & 0 \\ 1 & 1 & -2 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Back-substitution

It is sometimes preferable to solve a system of linear equations by using Gaussian elimination to bring the augmented matrix into row-echelon form. The corresponding system of equations can be solved by a technique called **back-substitution**.

Example 2.7 Solve the following homogeneous linear system by Gaussian elimination

$$\begin{aligned}
2x_1 + 2x_2 - x_3 + x_5 &= 0 \\
-x_1 - x_2 + 2x_3 - 3x_4 + x_5 &= 0 \\
x_1 + x_2 - 2x_3 - x_5 &= 0 \\
x_3 + x_4 + x_5 &= 0
\end{aligned}.$$

Example 2.8 Determine a basis for the solution space of the homogeneous linear system in Example 2.7.

Gauss-Jordan elimination

Same as Gaussian elimination but it includes a final step to get the reduced row-echelon form of the augmented matrix of a system, i.e. each column containing a leading 1 has zeros everywhere else.

Null Space of a Matrix

Let $\mathbf{A} \in M_{m,n}(\mathbf{R})$ and $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathrm{R}^n$, $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$.

Definition 3.1 The **kernel** of $\mathbf{A}$, (also called the **null space** of $\mathbf{A}$), is defined to be

$$\mathrm{Ker}(\mathbf{A}) = \left\{ \mathbf{x} \in \mathbf{R}^n : \ \mathbf{A}\mathbf{x} = \mathbf{0} \right\}.$$

Observation 3.2 The kernel of $\mathbf{A}$ (the null space of $\mathbf{A}$) is simply the set of solutions to the homogeneous system of linear equations $\mathbf{Ax} = \mathbf{0}$, and so is clearly a subset of $\mathbf{R}^n$.

Example 3.3 Consider the following matrix

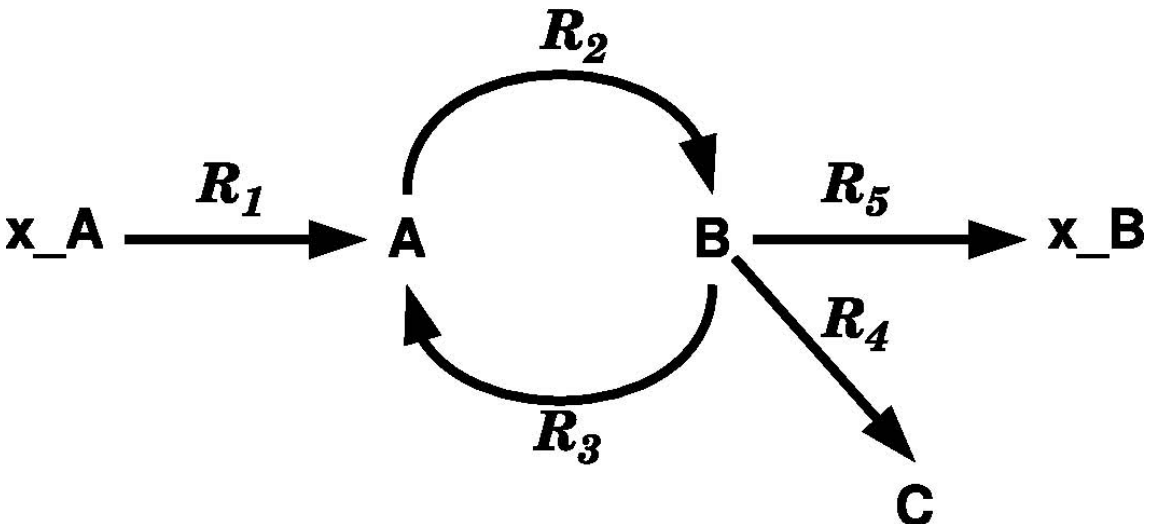
$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \in M_{3,5}(\mathbf{R}).$$

Find a basis for Ker$(\mathbf{A})$.

Elementary Modes of Metabolic Networks

The proportions with which the substances in a reaction system are interconverted can be written in the form of a stoichiometry matrix, $\mathbf{N}$.

For illustration, let us consider the following reaction scheme.



$$\frac{dA}{dt} = R_1 - R_2 + R_3,$$
$$\frac{dB}{dt} = R_2 - R_3 - R_4 - R_5,$$
$$\frac{dC}{dt} = R_4.$$

or,

$$\begin{pmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{pmatrix}.$$

In the long term, the rates of change in the concentrations are assumed to tend to zero (steady state). Hence,

$$\begin{pmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}}_{\mathbf{N}} \underbrace{\begin{pmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{pmatrix}}_{\mathbf{v}} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\mathbf{Nv} = \mathbf{0}$$

$$\mathbf{v} \in \mathrm{Ker}(\mathbf{N})$$

Since the matrix $\mathbf{N}$ is identical to the matrix $\mathbf{A}$ in Example 3.3,

$\{(0,1,1,0,0),(1,1,0,0,1)\}$ is a basis for Ker$(\mathbf{N})$.

The kernel (or null-space) matrix is a 5×2 matrix

$$\mathbf{K}=\begin{pmatrix}0 & 1\\ 1 & 1\\ 1 & 0\\ 0 & 0\\ 0 & 1\end{pmatrix},$$

Each reaction is associated with a 2-dimensional row vector.

All admissible steady-state reaction rates can be expressed (uniquely) as a linear combination of the vectors $\{(0,1,1,0,0),(1,1,0,0,1)\}$:

$$\mathbf{v}=\begin{pmatrix}R_1\\R_2\\R_3\\R_4\\R_5\end{pmatrix}=w_1\begin{pmatrix}0\\1\\1\\0\\0\end{pmatrix}+w_2\begin{pmatrix}1\\1\\0\\0\\1\end{pmatrix}=\begin{pmatrix}0&1\\1&1\\1&0\\0&0\\0&1\end{pmatrix}\begin{pmatrix}w_1\\w_2\end{pmatrix}=\mathbf{Kw}=\begin{pmatrix}0w_1+1w_2\\1w_1+1w_2\\1w_1+0w_2\\0w_1+0w_2\\0w_1+1w_2\end{pmatrix}\begin{matrix}\leftarrow\text{ subset}\\ \\ \\ \leftarrow\text{ dead}\\ \leftarrow\text{ subset}\end{matrix}$$

$\mathbf{w}=(w_1,w_2)$ is any vector in $\mathbf{R}^2$.

Significance of subsets

Ratios of fluxes carried by reactions in a subset are constant, regardless of actual flux carried.

Removal of one or more reactions from a subset causes the remaining reactions to become dead.

The maximum sustainable net flux through a subset is limited to the maximum flux that can be carried by the reaction with the lowest maximum flux in the subset.

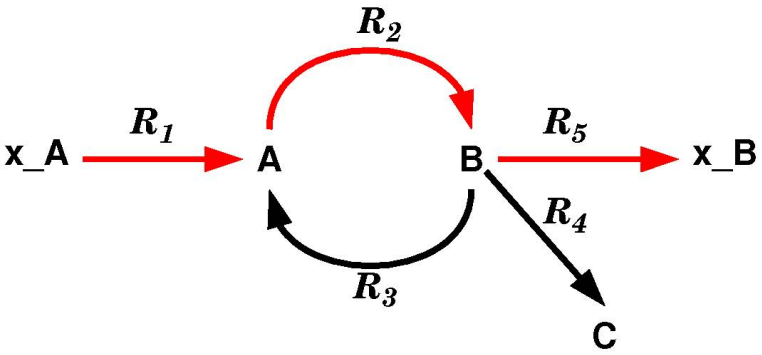
Subsets can be regarded as metabolic modules.

Definition 4.1 An **elementary mode** is a set of reactions in a system that balance all internal metabolites, respect reversibility and cannot be decomposed (i.e. a minimal set of reactions).

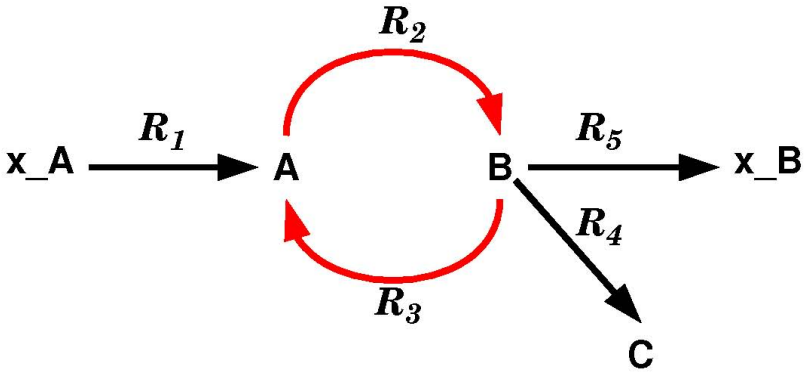
The elementary modes of metabolic networks are the simplest relevant way of connecting the inputs to the outputs of the system and could be represented by a proper choice of basis vectors of the null space since all admissible steady-state reaction rates are superpositions of elementary modes.

Elementary modes

1. $(1,1,0,0,1)$



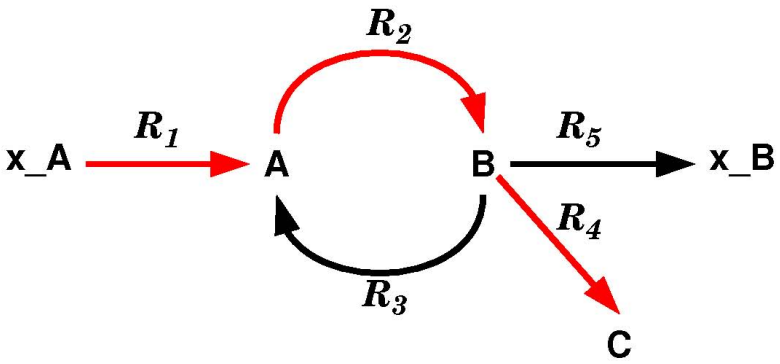
2. $(0,1,1,0,0)$

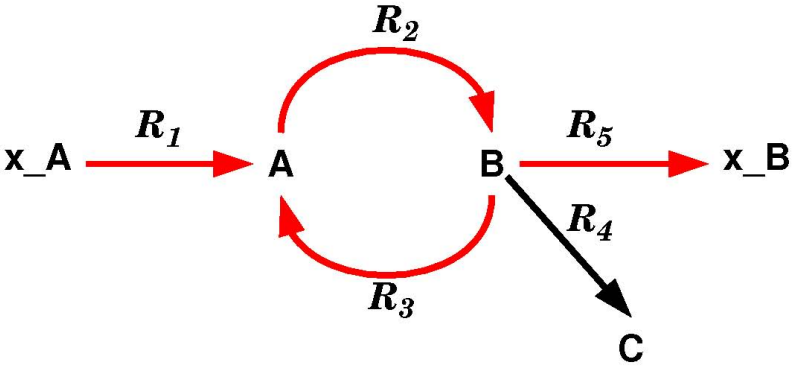


3. $(1,0,-1,0,1)$

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

Non Elementary Modes





Significance of null-space analysis

Encapsulates all possible steady state behaviour of a system.

Allows identification of relationships between fluxes.

Forms the starting point for most (if not all) structural analysis of metabolic networks.

Can also be used to establish similar relationships between concentrations.

Limitations of null-space analysis

However, it does not take into account the thermodynamics.

It is rather hard to integrate experimental flux observations.

Significance of elementary modes

Define all possible pathways (and hence net stoichiometries) through a network.

Represent minimal subsystems in a network.

Flux assignment to elementary modes gives a picture of how the system is utilising resources and allows an estimate as to how close an observed system is to steady-state.

A fundamental concept in the structural analysis of metabolic networks.

Restrictions of elementary mode analysis

Elementary modes might be impossible to calculate for large systems.

Even if possible, the number of elementary modes will be very large ($>10^6$).

The process might be much too slow for repeated and/or interactive analysis.

Identify only elementary modes of particular interest.

Statistical descriptions of the set of elementary modes.