MATH 4370/7370 – Linear Algebra and Matrix Analysis

Matrices are *everywhere*

Julien Arino

Fall 2023

University of Manitoba

Outline

We give a few examples that illustrate how ubiquitous matrices are in mathematics

In the process, we introduce some concepts that are used later

However, precise definitions are given in subsequent chapter; here concepts are introduced with very few explanations

Linear systems of difference equations

Ordinary differential equation

Linear systems of ODE – Brief theory

Discrete-time Markov chains

Discretisation of partial differential equations

Linear systems of difference equations

Let $x(t) \in \mathbb{R}^n$ be a **state variable** and $t \in \mathbb{N}$ be an **independent variable**, typically thought of as **time**. Let $A \in \mathcal{M}_n(\mathbb{R})$. An **autonomous homogeneous linear system of difference equations** is a sequence defined by

$$x(t+1) = Ax(t) \tag{1a}$$

$$x(0) = x_0 \in \mathbb{R}^n, \tag{1b}$$

where x_0 is called the **initial condition** (IC)

(Autonomous: A is a constant that does not depend on t. Homogeneous: the system is not of the form $x(t+1) = Ax(t) + b$)

Given the initial condition $x(0) = x_0$, we have from (1a) that

$$x(1) = Ax(0) = Ax_0$$
$$x(2) = Ax(1) = AAx_0 = A^2x_0$$
$$x(3) = Ax(2) = AA^2x_0 = A^3x_0$$

By induction,

$$x(t) = A^t x_0$$

for all $t \in \mathbb{N}$. Thus the behaviour of $x(t)$ as $t \to \infty$ depends on A^t

In order to understand what this behaviour could be, the following two questions should be answered

1. What is (if it exists) $\lim_{t\to\infty} x(t)$?
2. What is (if it exists) $\lim_{t\to\infty} A^t$?

Let v be an eigenvector associated to the eigenvalue λ of A, *i.e.*, λ be such that $v \neq 0$ satisfies $Av = \lambda v$. Then we have

$$\begin{aligned} A^2 v &= A(Av) \\ &= A(\lambda v) \\ &= \lambda A v \\ &= \lambda^2 v \end{aligned}$$

i.e., v is also an eigenvector of A^2 and is associated to the eigenvalue λ^2. By induction

$$A^k v = \lambda^k v$$

i.e., v is an eigenvector of the matrix A^k associated to the eigenvalue λ^k. $A^k v$ is a vector in $\mathbb{F}^n$; we can thus take its norm $\|A^k v\|$, where $\|\cdot\|$ is some norm on $\mathbb{F}^n$. It follows that if $|\lambda| < 1$, then $\|A^k v\| = |\lambda|^k \|v\|$ goes to zero as $k \to \infty$

Theorem 1.1

Let $A \in \mathcal{M}_n(\mathbb{R})$, consider the map Ax. TFAE:

1. *There exists a norm $\| \|_\alpha$ on $\mathbb{R}^n$ and a constant $0 < \mu < 1$ such that for any $x \in \mathbb{R}^n$, the iterates satisfy, for all $k \geq 0$*
$$\|A^k x\|_\alpha \leq \mu^k \|x\|_\alpha$$
2. *For any norm $\| \|_\beta$ on $\mathbb{R}^n$, there exists constants $0 < \mu < 1$ and $C \geq 1$ such that all $x \in \mathbb{R}^n$ and all $k \geq 0$*
$$\|Ax\|_\beta \leq C\mu^k \|x\|_\beta$$
3. *All the eigenvalues λ of A satisfy $|\lambda| < 1$*

To use the result above, we need to ensure all eigenvalues lie inside the unit disk in $\mathbb{C}$

For certain classes of matrices, this can be achieved without explicitly computing the eigenvalues

A linear map corresponding to a matrix with all eigenvalues of modulus less than 1 is a **linear contraction**, with the origin a **linear sink** or **attracting fixed point**. If all eigenvalues have modulus larger than 1, then the map induced by A is a **linear expansion**, and the origin is a **linear source** or **repelling fixed point**. The map Ax is a **hyperbolic linear map** if all eigenvalues of A have modulus different of 1

Definition 1.2

Let $\mathcal{M}_n$ be the set of square $n \times n$ matrices. For $A \in \mathcal{M}_n$, denote $\sigma(A)$ the set of its eigenvalues, *i.e.*,

$$\sigma(A) = \{\lambda \in \mathbb{C}; \exists v \neq 0,\ Av = \lambda v\}$$

which we call the **spectrum** of A. We call **spectral radius** of A the real number

$$\rho(A) = \max_{\lambda \in \sigma(A)} \{|\lambda|\}$$

Theorem 1.3

If $\rho(A) < 1$, *then* $\lim_{t \to \infty} x(t) = 0$ *for system* (1)

Example – Leslie matrices

Leslie matrices arise when considering the *age-structured* dynamics of populations reproducing every year, such as most fish

Assume n is the maximum age observed for that species. Time t is taken in years and is *discrete*. Let $x(t) = (x_1(t), \ldots, x_n(t))^T$ be the vector of distribution of the population in ages, *i.e.*, $x_i(t)$ is the population of fish of age a between $i-1$ and i at time t

A proportion $s_i \in [0,1]$ of individuals of age $i-1 \leq a < i$ survive to the next year. As years progress at the same rate as age, this means that $x_{i+1}(t+1) = s_i x_i(t)$

When individuals reproduce, they give birth to f_i individuals in the first age class, *i.e.*, birth in the first age class takes the form

$$x_1(t+1) = f_1 x_1(t) + f_2 x_2(t) + \cdots + f_n x_n(t)$$

Population growth model with a Leslie matrix

$$
\begin{aligned}
x_1(t+1) &= f_1 x_1(t) + f_2 x_2(t) + \cdots + f_n x_n(t) \\
x_2(t+1) &= s_1 x_1(t) \\
x_3(t+1) &= s_2 x_2(t) \\
&\vdots \\
x_n(t+1) &= s_{n-1} x_{n-1}(t)
\end{aligned}
$$

Suppose that on the first year counts are taken, we observe an initial age distribution $x(0) = (x_1(0), \ldots, x_n(0))^T$

Assume the survival rates s_i and fecundity constants f_i are known

How can we expect the population to evolve over the course of time? In particular, is the species likely to survive or will it become extinct?

Write the system in vector form

$$\begin{pmatrix} x_1(t+1) \\ x_2(t+1) \\ x_3(t+1) \\ \vdots \\ x_n(t+1) \end{pmatrix} = \begin{pmatrix} f_1 & f_2 & f_3 & \cdots & f_{n-1} & f_n \\ s_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & s_2 & 0 & \cdots & 0 & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & & & & s_{n-1} & 0 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ \vdots \\ x_n(t) \end{pmatrix} \tag{2}$$

which we summarize as

$$x(t+1) = Lx(t) \tag{3}$$

with L the matrix in (2), *i.e.*, a **nonnegative** matrix with only the first row and the first sub-diagonal nonzero

$$L = \begin{pmatrix} f_1 & f_2 & f_3 & \cdots & f_{n-1} & f_n \\ s_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & s_2 & 0 & \cdots & 0 & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & & & & s_{n-1} & 0 \end{pmatrix}$$

By Theorem 1.3, if $\rho(L) < 1$ then $x(t) \to 0$ at $t \to \infty$, so in this case, the population would become extinct

By the discussion earlier, if $\rho(L) = 1$, then $\|x(t)\|$ (the total population) stays constant

If $\rho(L) > 1$, the population increases

ODE

An **ordinary differential equation** (ODE) is an equation of the form

$$\frac{d}{dt}x(t) = f(x(t)) \tag{4}$$

where $x(t) \in \mathbb{R}^n$ is a *function* and $f : \mathbb{R}^n \to \mathbb{R}^n$

An **initial value problem** (IVP) is the consideration of (4) together with an initial condition $x(t_0) = x_0 \in \mathbb{R}^n$

A solution to (4) is a function $\phi(t)$ that satisfies (4). A solution to the IVP with $x(t_0) = x_0$ associated to (4) is, among all solutions to (4), the one (typically the only one) that additionally satisfies the initial condition, *i.e.*, such that $\phi(t_0) = x_0$

Linear systems

Consider the autonomous linear system

$$x'(t) = Ax(t) \tag{5}$$

where $A \in \mathcal{M}_n$ is a constant

Let $\sigma(A) = \{\lambda_1, \ldots, \lambda_n\}$ be the spectrum of A. Let $w_j = u_j + iv_j$ be a generalized eigenvector of A corresponding to an eigenvalue $\lambda_j = a_j + ib_j$, with $v_j = 0$ if $b_j = 0$, and

$$B = \{u_1, \ldots, u_k, u_{k+1}, v_{k+1}, \ldots, u_m, v_m\}$$

be a basis of $\mathbb{R}^n$, with $n = 2m - k$, where k is the number of real eigenvalues in $\sigma(A)$

Definition 1.4 (Stable, unstable and center subspaces)

The **stable**, **unstable** and **center** subspaces of the linear system (5) are given, respectively, by

$$E^s = \mathrm{Span}\{u_j, v_j : a_j < 0\},$$

$$E^u = \mathrm{Span}\{u_j, v_j : a_j > 0\}$$

and

$$E^c = \mathrm{Span}\{u_j, v_j : a_j = 0\}$$

Definition 1.5

The mapping $e^{At} : \mathbb{R}^n \to \mathbb{R}^n$ is the **flow** of the linear system (5)

The term **flow** is used since e^{At} describes the motion of points $x_0 \in \mathbb{R}^n$ along trajectories of (5)

Definition 1.6

If all eigenvalues of A have nonzero real part, *i.e.*, if $E^c = \emptyset$, then the flow e^{At} of system (5) is a **hyperbolic flow** and the system (5) is a **hyperbolic linear system**

Definition 1.7

A subspace $E \subset \mathbb{R}^n$ is **invariant with respect to the flow** e^{At}, or **invariant under the flow** of (5), if $e^{At}E \subset E$ for all $t \in \mathbb{R}$

Theorem 1.8

Let E be the generalized eigenspace of A associated to the eigenvalue λ. Then $AE \subset E$

Theorem 1.9

Let $A \in \mathcal{M}_n(\mathbb{R})$. Then

$$\mathbb{R}^n = E^s \oplus E^u \oplus E^c$$

Furthermore, if the matrix A is the matrix of the linear autonomous system (5), then E^s, E^u and E^c are invariant under the flow of (5), i.e., let $x_0 \in E^S$, $y_0 \in E^C$ and $z_0 \in E^U$, then $e^{At}x_0 \in E^S$, $e^{At}y_0 \in E^C$ and $e^{At}z_0 \in E^U$

Nonlinear systems of ODE

There is no general theory allowing to obtain explicit solutions of a nonlinear IVP

Instead of seeking explicit solutions, we use **qualitative analysis**, which uses analysis to establish properties of the solutions without needing to actually find their explicit expression

Suppose you are given a system of ordinary differential equations

$$x' = f(x), \tag{6}$$

where $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^n$ is C^1. A standard step when studying (6) **qualitatively** is to seek **equilibria** of (6), *i.e.*, points $x^\star \in \mathbb{R}^n$ such that

$$f(x^\star) = 0. \tag{7}$$

At such a point, $x' = 0$, meaning that system (6) is at rest. If you were to consider solutions to (6) with an initial condition $x(0) = x^\star$, then there would hold that $x(t) = x^\star$ for all $t \geq 0$

What would happen if instead of starting at $x^\star$, you were to choose an initial condition $x(0)$ close to but distinct from $x^\star$?

Simplified version of Hartman-Grobman

Definition 1.10

An equilibrium point $x^\star$ is **hyperbolic** if the Jacobian matrix Df of (6) evaluated at $x^\star$, denoted $Df(x^\star)$, has no eigenvalues with zero real part, *i.e.*, is invertible

Theorem 1.11 (Hartman-Grobman)

Let $x^\star$ be a hyperbolic equilibrium point of (6)*. Then in some neighbourhood $\mathcal{N}(x^\star)$ of $x^\star$, the flow of* (6) *is topologically equivalent to the flow of the linear system*

$$x' = Df(x^\star)(x - x^\star) \tag{8}$$

where $Df(x^\star)$ is the Jacobian matrix Df of f evaluated at $x^\star$

Theorem 1.12 (Stable manifold theorem)

Let E be an open subset of $\mathbb{R}^n$ containing the origin, let $f \in C^1(E)$, and let ϕ_t be the flow of the nonlinear system (6)*. Suppose that $f(0) = 0$ and that $Df(0)$ has k eigenvalues with negative real part and $n - k$ eigenvalues with positive real part. Then there exists a k-dimensional differentiable manifold S tangent to the stable subspace E^s of the linear system (8) at 0 such that for all $t \geq 0$, $\phi_t(S) \subset S$ and for all $x_0 \in S$*

$$\lim_{t \to \infty} \phi_t(x_0) = 0$$

and there exists an $(n - k)$-dimensional differentiable manifold U tangent to the unstable subspace E^u of (8) at 0 such that for all $t \leq 0$, $\phi_t(U) \subset U$ and for all $x_0 \in U$

$$\lim_{t \to -\infty} \phi_t(x_0) = 0$$

Example – A chemostat model

System of 2 nonlinear DE modelling a biological device called a **chemostat**

$$\frac{dS}{dt} = D\left(S^0 - S\right) - \mu(S)x \tag{9a}$$

$$\frac{dx}{dt} = (\mu(S) - D)x \tag{9b}$$

Parameters S^0 and D, respectively the **input concentration** and the **dilution** rate, are real and positive. The function μ is the **growth** function. It is generally assumed to satisfy $\mu(0) = 0$, $\mu' > 0$ and $\mu'' < 0$

One can verify that the positive quadrant is positively invariant under the flow of (9), *i.e.*, that for $S(0) \geq 0$ and $x(0) \geq 0$, solutions remain nonnegative for all positive times, and similar properties

But since we are here only interested in applications of the stable manifold theorem, we proceed to a very crude analysis, and will not deal with this point

Write the system in vector form as

$$\xi' = f(\xi),$$

with $\xi = (S, x)^T$ and

$$f(\xi) = \begin{pmatrix} D(S^0 - S) - \mu(S)x \\ (\mu(S) - D)x \end{pmatrix}$$

EP of the system are found by solving $f(\xi) = 0$. We find one situated on one of the boundaries of the positive quadrant

$$\xi_T^\star = (S_T^\star, x_T^\star) = \left(S^0, 0\right)$$

and the second one in the interior of $\mathbb{R}_+^2$,

$$\xi_I^\star = (S^\star, x^\star) = \left(\lambda, S^0 - \lambda\right)$$

where λ is such that $\mu(\lambda) = D$. Note that this implies that if $\lambda \geq S^0$, $\xi_T^\star$ is the only equilibrium of the system since in that case, $\xi_I^\star \not\geq 0$, which is not biologically realistic

At an arbitrary point $\xi = (S, x)$, the Jacobian matrix of (9) is given by

$$Df(\xi) = \begin{pmatrix} -D - \mu'(S)x & -\mu(S) \\ \mu'(S)x & \mu(S) - D \end{pmatrix}$$

Thus, at the trivial equilibrium $\xi_T^\star$

$$Df(\xi_T^\star) = \begin{pmatrix} -D & -\mu(S^0) \\ 0 & \mu(S^0) - D \end{pmatrix}$$

We have two eigenvalues, $-D$ and $\mu(S^0) - D$. Since $-D < 0$, we focus on the eigenvalue $\mu(S^0) - D$

Assume that $\mu(S^0) - D < 0$. This implies that $\xi_T^\star$ is the only equilibrium, since $\xi_I^\star$ is not feasible if $\lambda > S^0$

System has dimensionality 2 and $Df(\xi_T^\star)$ has two negative eigenvalues $\implies$ the stable manifold theorem (Theorem 1.12) states that there exists a 2-dimensional differentiable manifold $\mathcal{M}$ such that

- $\phi_t(\mathcal{M}) \subset \mathcal{M}$,
- for all $\xi_0 \in \mathcal{M}$, $\lim_{t\to\infty} \phi_t(\xi_0) = \xi_T^\star$.
- At $\xi_T^\star$, $\mathcal{M}$ is tangent to the stable subspace E^S of the linearized system $\xi' = Df(\xi_T^\star)(\xi - \xi_T^\star)$.

Since there are no eigenvalues with positive real part, there does not exist an unstable manifold in this case

Let us now characterize the nature of the stable subspace E^S. It is obtained by studying the linear system

$$
\begin{aligned}
\xi' &= Df(\xi_T^\star)(\xi - \xi_T^\star) \\
&= \begin{pmatrix} -D & -\mu(S^0) \\ 0 & \mu(S^0) - D \end{pmatrix} \begin{pmatrix} S - S^0 \\ x \end{pmatrix} \\
&= \begin{pmatrix} -D(S - S^0) - \mu(S^0)x \\ (\mu(S^0) - D)x \end{pmatrix}
\end{aligned}
\tag{10}
$$

Of course, the Jacobian matrix associated to this system is the same as that of the nonlinear system (at $\xi_T^\star$). Associated to the eigenvalue $-D$ is the eigenvector $v_1 = (1, 0)^T$, to $\mu(S^0) - D$ is $v_2 = (-1, 1)^T$

The stable subspace is thus given by $\text{span}(v_1, v_2)$, *i.e.*, the whole of $\mathbb{R}^2$

In fact, the stable manifold of $\xi_T^\star$ is the whole positive quadrant, since all solutions limit to this equilibrium

The same type of analysis can be conducted at the interior equilibrium $\xi_I^\star$. It is a little harder in this case, since $x^\star > 0$ there and therefore the Jacobian matrix $Df(\xi_I^\star)$ does not have the same upper triangular structure as $Df(\xi_T^\star)$

Linear ODEs

Definition 1.13 (Linear ODE)

A *linear* ODE is a differential equation taking the form

$$\frac{d}{dt}x = A(t)x + B(t), \tag{LNH}$$

where $A(t) \in \mathcal{M}_n(\mathbb{R})$ with continuous entries, $B(t) \in \mathbb{R}^n$ with real valued, continuous coefficients, and $x \in \mathbb{R}^n$. The associated IVP takes the form

$$\begin{aligned} \frac{d}{dt}x &= A(t)x + B(t) \\ x(t_0) &= x_0. \end{aligned} \tag{11}$$

Types of systems

- $x' = A(t)x + B(t)$ is linear nonautonomous ($A(t)$ depends on t) nonhomogeneous (also called *affine* system).
- $x' = A(t)x$ is linear nonautonomous homogeneous.
- $x' = Ax + B$, that is, $A(t) \equiv A$ and $B(t) \equiv B$, is linear autonomous nonhomogeneous (or affine autonomous).
- $x' = Ax$ is linear autonomous homogeneous.

Existence and uniqueness of solutions

Theorem 1.14 (Existence and Uniqueness)

Solutions to (11) *exist and are unique on the whole interval over which A and B are continuous.*
In particular, if A, B are constant, then solutions exist on $\mathbb{R}$.

Autonomous linear systems

Consider the autonomous affine system

$$\frac{d}{dt}x = Ax + B, \tag{A}$$

and the associated homogeneous autonomous system

$$\frac{d}{dt}x = Ax. \tag{L}$$

Exponential of a matrix

Definition 1.15 (Matrix exponential)

Let $A \in \mathcal{M}_n(\mathbb{F})$ with $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. The *exponential* of A, denoted e^{At}, is a matrix in $\mathcal{M}_n(\mathbb{F})$, defined by

$$e^{At} = \mathbb{I} + \sum_{k=1}^{\infty} \frac{t^k}{k!} A^k,$$

where $\mathbb{I}$ is the identity matrix in $\mathcal{M}_n(\mathbb{F})$.

Properties of the matrix exponential

- $e^{At_1} e^{At_2} = e^{A(t_1+t_2)}$ for all $t_1, t_2 \in \mathbb{R}$. 1
- $Ae^{At} = e^{At}A$ for all $t \in \mathbb{R}$.
- $(e^{At})^{-1} = e^{-At}$ for all $t \in \mathbb{R}$.
- The unique solution ϕ of (L) with $\phi(t_0) = x_0$ is given by

$$\phi(t) = e^{A(t-t_0)}x_0.$$

Computing the matrix exponential

Let P be a nonsingular matrix in $\mathcal{M}_n(\mathbb{R})$. We transform the IVP

$$\begin{aligned} \frac{d}{dt}x &= Ax \\ x(t_0) &= x_0 \end{aligned} \tag{LIVP}$$

using the transformation $x = Py$ or $y = P^{-1}x$.

The dynamics of y is $y' = P^{-1}APy$. The initial condition is $y_0 = P^{-1}x_0$.

We have thus transformed IVP (LIVP) into

$$\begin{aligned} \frac{d}{dt}y &= P^{-1}APy \\ y(t_0) &= P^{-1}x_0 \end{aligned} \qquad \text{(L\_IVP\_y)}$$

From the earlier result, we then know that the solution of (L_IVP_y) is given by

$$\psi(t) = e^{P^{-1}AP(t-t_0)}P^{-1}x_0,$$

and since $x = Py$, the solution to (LIVP) is given by

$$\phi(t) = Pe^{P^{-1}AP(t-t_0)}P^{-1}x_0.$$

So everything depends on $P^{-1}AP$.

The cases

- $P^{-1}AP$ is diagonal, the solution to (LIVP) is given by

$$\phi(t) = P \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix} P^{-1} x_0.$$

- $P^{-1}AP$ is not diagonal, then use Jordan form (slightly more complicated).

Theorem 1.16

For all $(t_0, x_0) \in \mathbb{R} \times \mathbb{R}^n$*, there is a unique solution* $x(t)$ *to* (LIVP) *defined for all* $t \in \mathbb{R}$*. Each coordinate function of* $x(t)$ *is a linear combination of functions of the form*

$$t^k e^{\alpha t} \cos(\beta t) \quad and \quad t^k e^{\alpha t} \sin(\beta t)$$

where $\alpha + i\beta$ *is an eigenvalue of* A *and* k *is less than the algebraic multiplicity of the eigenvalue.*

Generalized eigenvectors, nilpotent matrix

Definition 1.17 (Generalized eigenvectors)

Let $A \in \mathcal{M}_r(\mathbb{R})$. Suppose λ is an eigenvalue of A with multiplicity $m \leq n$. Then, for $k = 1, \ldots, m$, any nonzero solution v of

$$(A - \lambda \mathbb{I})^k v = 0$$

is called a *generalized eigenvector* of A.

Definition 1.18 (Nilpotent matrix)

Let $A \in \mathcal{M}_n(\mathbb{R})$. A is *nilpotent* (of order k) if $A^j \neq 0$ for $j = 1, \ldots, k-1$, and $A^k = 0$.

Jordan normal form

Theorem 1.19 (Jordan normal form)

Let $A \in \mathcal{M}_n(\mathbb{R})$ have eigenvalues $\lambda_1, \ldots, \lambda_n$, repeated according to their multiplicities.

- *Then there exists a basis of generalized eigenvectors for $\mathbb{R}^n$.*
- *And if $\{v_1, \ldots, v_n\}$ is any basis of generalized eigenvectors for $\mathbb{R}^n$, then the matrix $P = [v_1 \cdots v_n]$ is invertible, and A can be written as*

  $$A = S + N,$$

  where

  $$P^{-1}SP = \mathrm{diag}(\lambda_j),$$

  the matrix $N = A - S$ is nilpotent of order $k \leq n$, and S and N commute, i.e., $SN = NS$.

Theorem 1.20

Under conditions of the Jordan normal form Theorem, the linear system $x' = Ax$ with initial condition $x(0) = x_0$, has solution

$$x(t) = P\text{diag}\left(e^{\lambda_j t}\right) P^{-1} \left(\mathbb{I} + Nt + \cdots \frac{t^k}{k!} N^k\right) x_0.$$

The result is particularly easy to apply in the following case.

Theorem 1.21 (Case of an eigenvalue of multiplicity n)

Suppose that λ is an eigenvalue of multiplicity n of $A \in \mathcal{M}_n(\mathbb{R})$. Then $S = \text{diag}(\lambda)$, and the solution of $x' = Ax$ with initial value x_0 is given by

$$x(t) = e^{\lambda t} \left(\mathbb{I} + Nt + \cdots \frac{t^k}{k!} N^k\right) x_0.$$

In the simplified case, we do not need the matrix P (the basis of generalized eigenvectors).

Discrete-time Markov chain

A **discrete-time Markov chain** is a stochastic process

Consider a system with n **states** denoted $S_1, \ldots, S_n$. The system starts in a given state. Every time step, it switches to a different state. (Transition from one state to itself is also allowed)

We assume that the system is *stochastic*, *i.e.*, that the transitions happen at random. In discrete-time Markov chains, the instants at which transitions occur (or not) are in a discrete set, typically rescaled to be $\mathbb{N}$

(In *continuous-time* Markov chains, the time of the switches themselves is a continuous random variable, so times are in $\mathbb{R}$)

Let $p_i(t)$ be the probability that state S_i occurs on the t^{th} time step, $1 \leq i \leq n$

One of the key assumptions in Markov chains is that the process is **memoryless**: the transition that occurs from time t to time $t+1$ depends only on the state of the system at time t

Since one the states S_i must occur on the t^{th} time step,

$$p_1(t) + p_2(t) + \cdots + p_n(t) = 1$$

The transition matrix

Let

$$P = [p_{ij}]$$

be the **transition matrix** (or **projection matrix**) of the Markov chain, where

$$p_{ij} = \mathbb{P}(S_i | S_j)$$

i.e., the probability of making a transition from state j to state i

Many texts, *e.g.*, [KS83], define $p_{ij} = \mathbb{P}(S_j | S_i)$. This works the same way, except it leads to P^T instead of P (or row vectors instead of column vectors)

Let $p_i(t+1)$ be the probability that state S_i, $1 \leq i \leq n$, occurs on the $(t+1)^{th}$ step. There are n ways to be in state S_i at step $t+1$:

1. Step t is S_1. The probability of getting S_1 on the t^{th} step is $p_1(t)$ and the probability of having S_i after S_1 is p_{i1}, so $\mathbb{P}(S_i|S_1) = p_{i1}p_1(k)$
2. We get S_2 on step t and S_i on step $t+1$ and $\mathbb{P}(S_i|S_2) = p_{i2}p_2(t)$

..

n. The probability of occurrence of S_i at step $t+1$ if S_n at step t is $\mathbb{P}(S_i|S_n) = p_{in}p_n(t)$

$$
\begin{aligned}
\implies p_i(t+1) &= \mathbb{P}(S_i|S_1) + \cdots + \mathbb{P}(S_i|S_n) \\
&= p_{i1}p_1(t) + \cdots + p_{in}p_n(t)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
p_1(t+1) &= p_{11}p_1(t) + p_{12}p_2(t) + \cdots + p_{1n}p_n(t) \\
&\vdots \\
p_k(t+1) &= p_{k1}p_1(t) + p_{k2}p_2(t) + \cdots + p_{kr}p_r(t) \\
&\vdots \\
p_n(t+1) &= p_{n1}p_1(t) + p_{n2}p_2(t) + \cdots + p_{nn}p_r(t)
\end{aligned}
$$

In matrix form,

$$p(t+1) = Pp(t), \quad t = 1, 2, 3, \ldots \tag{12}$$

where $p(t) = (p_1(t), p_2(t), \ldots, p_n(t))^T$ is a (column) probability vector and $P = [p_{ij}] \in \mathcal{M}_n$ is the transition matrix

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} \tag{13}$$

Two observations about P

- Entries of P being probabilities, they are all in $[0,1]$. So in particular, they are all nonnegative. We say P is a **nonnegative matrix** and write $P \geq 0$

- The *column* sums of P all equal 1. Take for instance the first column: its entries represent the probabilities of transition from state 1. Since an event must always happen, the sum of these probabilities *must* be 1. We say P is a (column) **stochastic matrix**

Definition 1.22 (Stochastic matrix)

The nonnegative $n \times n$ matrix M is

- **row stochastic** if $\sum_{j=1}^{n} m_{ij} = 1$ for all $i = 1, \ldots, n$
- **column stochastic** if $\sum_{i=1}^{n} m_{ij} = 1$ for all $j = 1, \ldots, n$
- **stochastic** if it is row or column stochastic
- **doubly stochastic** if it is row *and* column stochastic

An interesting property of stochastic matrices

Theorem 1.23

Let M be a stochastic matrix. Then all eigenvalues λ of M are such that $|\lambda| \leq 1$. Furthermore, $\lambda = 1$ is an eigenvalue of M

Assume w.l.o.g. that M is row stochastic, *i.e.*, M has row sums 1. In vector form, $M\mathbb{1} = \mathbb{1}$. Now remember that λ is an eigenvalue of M, with associated eigenvector $v \neq 0$, if and only if $Mv = \lambda v$. So, in the expression $M\mathbb{1} = \mathbb{1}$, we read an eigenvector, $\mathbb{1}$, and an eigenvalue, 1

Proving the first conclusion of Theorem 1.23 involves a theorem called the Perron-Frobenius Theorem, which we will see in much detail later

Long time behaviour

Let $p_0 := p(0) = (p_1(0), \dots, p_n(0))^T$ be the initial probability distribution vector, with $\mathbb{1}^T p_0 = 1$, *i.e.*, such that the sum of the entries of p_0 be 1; we could also write $\langle p_0, \mathbb{1} \rangle = 1$. Then

$$
\begin{aligned}
p(1) &= Pp(0) = Pp_0 \\
p(2) &= Pp(1) \\
&= P(Pp_0) \\
&= P^2 p_0
\end{aligned}
$$

Iterating, for any $t \in \mathbb{N}$,

$$p(t) = P^t p_0$$

Rings a bell? (cf Section 1)

So the long time evolution of the system is governed by

$$\lim_{t\to+\infty} p(t) = \lim_{t\to+\infty} P^t p_0 = \left(\lim_{t\to+\infty} P^t\right) p_0 \tag{14}$$

if the latter limit exists

So if we can characterize the nature of matrix P^t and in particular, the existence of the limit $\lim_{t\to\infty} P^t$, we will know the long time behaviour of the Markov chain

Theorem 1.24

If M, N are nonsingular stochastic matrices, then MN is a stochastic matrix

Corollary 1.25

If M is a nonsingular stochastic matrix, then for any $t \in \mathbb{N}$, M^t is a stochastic matrix

$\implies$ matrix P^t in (14) is stochastic; so, in particular, it is a nonnegative matrix with column sums all equal to 1

Regular Markov chains

Definition 1.26 (Regular Markov chain)

A **regular Markov chain** is one in which P^k is positive for some integer $k > 0$, i.e., P^k has only positive entries, no zero entries

Definition 1.27 (Primitive matrix)

A nonnegative matrix M is **primitive** if, and only if, there is an integer $k > 0$ such that M^k is (entry-wise) positive

Theorem 1.28

A Markov chain is regular $\iff$ P *is primitive*

Regular Markov chains are well-behaved

Theorem 1.29

If P is the transition matrix of a regular Markov chain, then

1. *the powers P^t approach a stochastic matrix W*
2. *each column of W is the same vector $w = (w_1, \ldots, w_n)^T$*
3. *the components of w are positive*

So if the Markov chain is regular, (14) becomes

$$\lim_{t \to +\infty} p(t) = \lim_{t \to +\infty} P^t p_0 = W p_0$$

Let $M \in \mathcal{M}_n$, u, v be two column vectors, $\lambda \in \mathbb{C}$. Then, if

$$Mu = \lambda u$$

u is the (right) eigenvector corresponding to λ, and if

$$v^T M = \lambda v^T$$

then v is the left eigenvector corresponding to λ

To a given eigenvalue there corresponds one left and one right eigenvector (to multiples)

$(v^T M) = (\lambda v^T)^T \iff M^T v = \lambda v$, so if your numeric/symbolic solver spits out right eigenvectors, to get left ones, compute eigenvectors of M^T

Back to the regular MC

We already know that the left eigenvector corresponding to 1 is $\mathbb{1}^T$, since $\mathbb{1}^T P = \mathbb{1}^T$, *i.e.*, the column sums of P all equal 1

The vector w in Theorem 1.29 is the right eigenvector corresponding to the eigenvalue 1 of P

To see this, remark that, if $p(t)$ converges, then $p(t+1) = Pp(t)$ in the limit for large t, so w is a fixed point of the system. We thus write

$$w = Pw$$

and solve for w, which amounts to finding w as the (right) eigenvector corresponding to the eigenvalue 1

When you compute an eigenvector, the result is to a multiple and often the expression needs normalising (you want a probability vector)

Once you obtain w, check that the norm

$$\|w\| = w_1 + \cdots + w_n$$

is equal to one. If not, normalise as

$$\frac{w}{\|w\|}$$

Absorbing Markov chains

Suppose now that the matrix is not only not primitive but also **reducible**

Definition 1.30 (Reducible/irreducible matrices)

$0 \leq M \in \mathcal{M}_n$ is **reducible** if $\exists P \in \mathcal{M}_n$, permutation matrix, s.t.

$$P^T MP = \begin{pmatrix} S & R \\ 0 & Q \end{pmatrix}$$

If no such matrix exists, M is **irreducible**

Let $\mathcal{G}(M)$ be the digraph induced by M

Theorem 1.31

$0 \leq M \in \mathcal{M}_n$ *irreducible* $\iff$ $\mathcal{G}(M)$ *strongly connected*

So if the transition matrix P is reducible, $\mathcal{G}(P)$ is not strongly connected, *i.e.*, there are states of the chain that are not accessible from others

When in a state that does not have access to other states, we are stuck.. we say **absorbed**

An **absorbing Markov chain** is one where at least one state is absorbing

Theorem 1.29 does not apply here, but we get a lot of interesting properties by observing that P can be put in the form

$$\begin{pmatrix} \mathbb{I} & R \\ 0 & Q \end{pmatrix}$$

and considering the **fundamental matrix** $N = (\mathbb{I} - Q)^{-1}$

Following [Win89]: $\Omega \subset \mathbb{R}^d$ $(d \geq 1)$ a bounded connected region, $u = u(x)$ for $x \in \Omega$; a linear elliptic BVP on Ω takes the form

$$Lu = f, \quad \Omega \tag{15a}$$

$$\alpha u + \beta \frac{\partial u}{\partial v} = g, \quad \partial\Omega \tag{15b}$$

where L is a linear differential operator of the form

$$Lu = -\nabla(k\nabla u + bu) + qu$$

Suppose that the scalar functions $k(x)$, $q(x)$ and the vector function $b(x) = (b_1(x), \ldots, b_d(x))^T$ for $d > 1$, otherwise a scalar function $b(x)$, are sufficiently smooth over the region Ω. Further, let $k(x) \geq k_0 = \mathrm{const} > 0$ and $q(x) \geq 0$ for each $x \in \Omega$

L together with corresponding BC on $\partial\Omega$ is a linear elliptic differential operator

As a particular example, suppose $\Omega = (0, 1)$ is a one-dimensional domain, i.e., an interval, and suppose

$$\bar{\omega}_h = \{0 = x_1 < x_2 < \ldots < x_n = 1\} = \omega_h + \gamma_h$$

is a discretisation of the closure of said interval, with $\gamma_h = \{x_0, x_n\}$ and $h_i = h = 1/n$ be a uniform grid, i.e., $x_i = ih$. Consider the following problem

$$Lu = -u'' + b(x)u' = 0, \quad x \in \Omega \tag{16a}$$

$$u(0) = u_0, \quad u(1) = u_1 \tag{16b}$$

with $b(x)$ bounded in Ω

Let us approximate on $\bar{\omega}$ using *centred differences* (a finite difference scheme):

$$y_0 = u_0 \tag{17a}$$

$$-D_+ D_- y_i + b_i D_0 y_i = 0, \quad i = 1, \ldots, n-1 \tag{17b}$$

$$y_n = u_1 \tag{17c}$$

where $b_i = b(x_i)$. The notation $D_+ D_-$ and D_0 refer to *difference operators*:

$$D_+ y_i = \frac{y_{i+1} - y_i}{h} \tag{18a}$$

$$D_- y_i = \frac{y_i - y_{i-1}}{h} \tag{18b}$$

$$D_0 y_i = \frac{y_{i+1} - y_{i-1}}{2h} \tag{18c}$$

$$D_+ D_- y_i = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} \tag{18d}$$

These operators “encode” derivatives on the discrete grid used

For $i = 1, \ldots, n-1$, define $\gamma_i = hb_i/2$. Then the problem can be written in matrix form as

$$Ay = f \tag{19}$$

where $f = (u_0, 0, \ldots, 0, u_1)^T = u_0 e_1 + u_1 e_{n+1}$ and

$$A = \begin{pmatrix} 1 & 0 & & & & & \\ -(1+\gamma_1) & 2 & -(1-\gamma+1) & & & & \\ & -(1+\gamma_2) & 2 & -(1-\gamma_2) & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & & -(1+\gamma_{n-1}) & 2 & -(1-\gamma_{n-1}) \\ & & & & & 0 & 1 \end{pmatrix}$$

Properties of the (approximate) solution then depend on the properties of the M-matrix A

References I

J.G. Kemeny and J.L. Snell, *Finite markov chains*, Springer-Verlag, 1983.

G. Windisch, *M-matrices in numerical analysis*, Springer, 1989.