

The notes below review basic concepts and theorems from linear algebra and Markov chains that will be used in the course. For detailed discussion of the definitions and results, for examples, and for proofs, the reader is referred to standard textbooks in linear algebra and probability theory.

The notes below assume familiarity with the following notions: fields, vectors, scalars, vector spaces, and matrices.

1 Inner and outer products

A vector $x \in \mathbb{R}^n$ is thought of as a *column vector*. x^T is the same vector, but thought of as a *row vector*.

The *inner product* of two vectors $x, y \in \mathbb{R}^n$ is defined as:

$$\langle x, y \rangle = x^T y = \sum_{i=1}^n x_i y_i.$$

Note that the inner product of two vectors is a scalar.

The *outer product* of two vectors $x, y \in \mathbb{R}^n$ is defined as xy^T . Note that the outer product is an $n \times n$ matrix. The (i, j) -th entry of this matrix consists of the scalar $x_i y_j$.

Example 1. We denote the all-ones n -dimensional vector by $\mathbf{1}$ and the all-ones $n \times n$ matrix by \mathbf{J} . Hence, the inner product $\mathbf{1}^T \mathbf{1}$ equals the scalar n , while the outer product $\mathbf{1} \mathbf{1}^T$ equals the matrix \mathbf{J} .

2 Norms

A *norm* is a function $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}$. It gives each vector $x \in \mathbb{R}^n$ a measure of *length*. There are many different types of norms, but all of them satisfy the following axioms:

1. Non-negativity: $\forall x \in \mathbb{R}^n, \|x\| \geq 0$.
2. $\|x\| = 0$ if and only if $x = \mathbf{0}$. ($\mathbf{0}$ denotes the all-zeroes vector.)
3. Scalability: $\forall x \in \mathbb{R}^n, \forall a \in \mathbb{R}, \|a \cdot x\| = |a| \cdot \|x\|$.
4. Triangle inequality: $\forall x, y \in \mathbb{R}^n, \|x + y\| \leq \|x\| + \|y\|$.

The *distance* between two vectors $x, y \in \mathbb{R}^n$ induced by the norm $\|\cdot\|$ is defined as:

$$\forall x, y \in \mathbb{R}^n, d(x, y) = \|x - y\|.$$

The *unit circle* under norm $\|\cdot\|$ is defined as:

$$\{x \in \mathbb{R}^n \mid \|x\| = 1\}.$$

2.1 Euclidean norm

The *Euclidean norm* (a.k.a. the L_2 norm) of a vector $x \in \mathbb{R}^n$ is defined as:

$$\|x\|_2 = \sqrt{x^T x} = \sqrt{\sum_{i=1}^n x_i^2}.$$

The Euclidean norm induces the standard, geometric, measure of distance all of us are familiar with:

$$\|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

The unit circle (see Figure 1(a)) under the Euclidean norm is the standard circle whose center is at 0 and whose radius is 1.

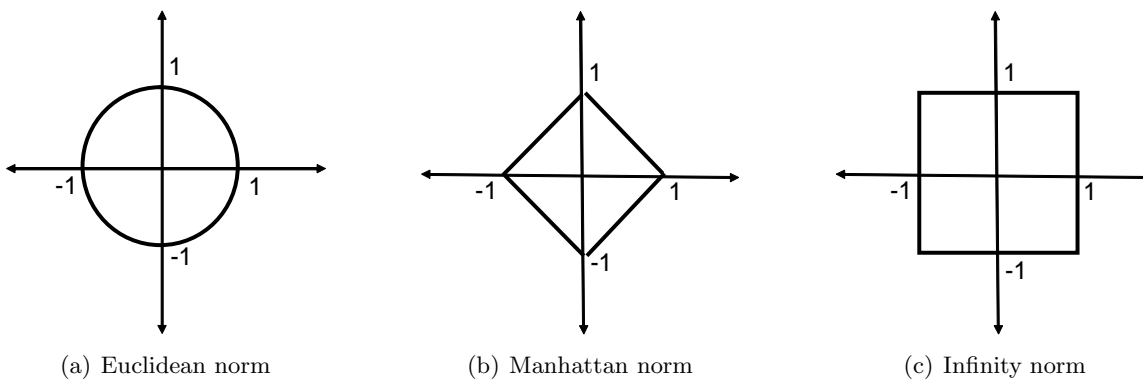


Figure 1: The unit circle under the Euclidean, Manhattan, and Infinity norms.

2.2 Manhattan norm

The *Manhattan norm* (a.k.a. the L_1 norm) of a vector $x \in \mathbb{R}^n$ is defined as:

$$\|x\|_1 = \sum_{i=1}^n |x_i|.$$

The unit circle under the Manhattan norm is depicted in Figure 1(b).

2.3 Infinity norm

The *Infinity norm* (a.k.a. the L_∞ norm) of a vector $x \in \mathbb{R}^n$ is defined as:

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

The unit circle under the Infinity norm is depicted in Figure 1(c).

2.4 p -norms

For $p > 0$, the p -norm (a.k.a. the L_p norm) of a vector $x \in \mathbb{R}^n$ is defined as:

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}.$$

The p -norm generalizes the Euclidean norm ($p = 2$) and the Manhattan norm ($p = 1$).

2.5 0-norm

The 0-norm (a.k.a. the L_0 norm) of a vector $x \in \mathbb{R}^n$ is defined as:

$$\|x\|_0 = \sum_{i=1}^n x_i^0,$$

where 0^0 is defined as 0. That is, $\|x\|_0$ is simply the number of non-zero entries in x . Unlike the previous examples the 0-norm is not strictly a norm, because it does not scale when multiplied by a scalar.

2.6 Relationships among the norms

There are many known relationships among the above norms. We review some of them here. (For others, like Hölder's inequality, refer to a textbook in algebra.)

Proposition 2. For every vector $x \in \mathbb{R}^n$,

$$\frac{\|x\|_1}{n} \leq \|x\|_\infty \leq \|x\|_2 \leq \|x\|_1.$$

2.7 Normalization

A vector $x \in \mathbb{R}^n$ is called *normalized* (under norm $\|\cdot\|$), if $\|x\| = 1$. Given any vector $x \in \mathbb{R}^n$, $x \neq \mathbf{0}$, the *normalized form* of x is the vector $y = \frac{x}{\|x\|}$. y equals x , up to a multiplication by a scalar, and it is normalized:

$$\|y\| = \left\| \frac{x}{\|x\|} \right\| = \frac{1}{\|x\|} \cdot \|x\| = 1.$$

2.8 Angles

The *angle* between two vectors $x, y \in \mathbb{R}^n$, $x, y \neq \mathbf{0}$, is defined as:

$$\angle(x, y) = \arccos\left(\frac{x^T y}{\|x\|_2 \|y\|_2}\right).$$

That is, $\cos(x, y)$ is the inner product of the normalized forms of x and y . To show that this inner product is indeed always in the interval $[-1, 1]$, we use the following inequality:

Theorem 3 (Cauchy-Schwartz inequality). *For any two vectors $x, y \in \mathbb{R}^n$,*

$$\sum_{i=1}^n |x_i y_i| \leq \sqrt{\sum_{i=1}^n x_i^2} \cdot \sqrt{\sum_{i=1}^n y_i^2}.$$

In other words, if $x \cdot y$ denotes the vector obtained by pointwise multiplying x and y , then the Cauchy-Schwartz inequality can be written as:

$$\|x \cdot y\|_1 \leq \|x\|_2 \cdot \|y\|_2.$$

It follows from the triangle inequality and from the Cauchy-Schwartz inequality that:

$$\left| \frac{x^T y}{\|x\|_2 \|y\|_2} \right| \leq \frac{\|x \cdot y\|_1}{\|x\|_2 \|y\|_2} \leq 1.$$

Hence, the inner product of $\frac{x}{\|x\|_2}$ and $\frac{y}{\|y\|_2}$ is indeed in the interval $[-1, 1]$.

Another corollary of the Cauchy-Schwartz inequality is the following connection between the Euclidean norm and the Manhattan norm:

Proposition 4. *For every vector $x \in \mathbb{R}^n$,*

$$\|x\|_1 \leq \sqrt{n} \cdot \|x\|_2.$$

Two vectors $x, y \in \mathbb{R}^n$ are called *orthogonal*, if $\angle(x, y) = \frac{\pi}{2}$. That is, x, y are orthogonal if and only if $x^T y = 0$. We have the following fact about the sum of orthogonal vectors:

Theorem 5 (Pythagorean theorem). *If $x, y \in \mathbb{R}^n$ are orthogonal, then*

$$\|x + y\|_2^2 = \|x\|_2^2 + \|y\|_2^2.$$

3 Bases

Suppose V is a vector space over some field \mathbb{F} . (V can be, for example, \mathbb{R}^n or some subspace of \mathbb{R}^n .) A *linear combination* of a given set of vectors $x_1, \dots, x_k \in V$ is the weighted sum vector

$$\sum_{i=1}^k a_i x_i,$$

where $a_1, \dots, a_k \in \mathbb{F}$ are scalars. If a vector x can be written as a linear combination of x_1, \dots, x_k , we say that x *linearly depends* on $\{x_1, \dots, x_k\}$. The set of vectors that linearly depend on $\{x_1, \dots, x_k\}$ is the *span* of x_1, \dots, x_k :

$$\text{span}(x_1, \dots, x_k) = \{x \in V \mid \exists a_1, \dots, a_k \in \mathbb{F}, x = \sum_{i=1}^k a_i x_i\}.$$

In principle, the representation of a vector $x \in \text{span}(x_1, \dots, x_k)$ as a linear combination of x_1, \dots, x_k need not be unique.

Proposition 6. *For any set of vectors $x_1, \dots, x_k \in V$, $\text{span}(x_1, \dots, x_k)$ is a linear subspace of V .*

A set of vectors x_1, \dots, x_k is called *independent*, if none of these vectors linearly depends on the others. That is, for all $i \in \{1, \dots, k\}$, $x_i \notin \text{span}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$. If x_1, \dots, x_k is independent, then every vector $x \in \text{span}(x_1, \dots, x_k)$ has a *unique* representation as a linear combination of x_1, \dots, x_k .

A *basis* for V is a maximally independent set of vectors. That is, x_1, \dots, x_k is a basis, if: (1) x_1, \dots, x_k are independent; (2) no set of vectors that strictly contains $\{x_1, \dots, x_k\}$ is independent.

Example 7. A canonical example of a basis for \mathbb{R}^n is the *standard basis*: $\mathbf{e}_1, \dots, \mathbf{e}_n$, where \mathbf{e}_i is the *standard unit vector*:

$$\mathbf{e}_i[j] = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

Fact 8. *All bases for V are of the same size.*

The size of (any) basis for V is called the *dimension* of V .

Theorem 9. *Let x_1, \dots, x_n be a basis for V . Then, every vector $x \in V$ has a unique representation as a linear combination of x_1, \dots, x_n .*

It follows from the theorem that if x_1, \dots, x_n is a basis, then $V = \text{span}(x_1, \dots, x_n)$.

4 Matrices

Let A be an $m \times n$ matrix over \mathbb{R} . The m rows of A are denoted A_1, \dots, A_m and the n columns of A are denoted A^1, \dots, A^n . A is called a *square matrix*, if $n = m$.

4.1 Vector-matrix product

The *product* of an $m \times n$ matrix A and a column vector $x \in \mathbb{R}^n$ is defined as:

$$Ax = (A_1^T x, \dots, A_m^T x).$$

That is, Ax is an m -dimensional column vector, whose i -th entry is the inner product between the i -th row of A and x .

The *product* of a row vector $x^T \in \mathbb{R}^m$ and A is defined as:

$$x^T A = (x^T A^1, \dots, x^T A^n).$$

That is, $x^T A$ is an n -dimensional row vector, whose j -th entry is the inner product between x and the j -th column of A .

4.2 Matrix product

Let A be an $m \times n$ matrix and let B be an $n \times k$ matrix. The *product* of A and B is an $m \times k$ matrix AB , whose (i, j) -th entry is the inner product of the i -th row of A and the j -th column of B :

$$(AB)_{ij} = A_i^T B^j.$$

4.3 Rank

The *row rank* of A is the dimension of $\text{span}(A_1, \dots, A_m)$. The *column rank* of A is the dimension of $\text{span}(A^1, \dots, A^n)$.

Theorem 10. *For any matrix A , its row rank equals its column rank.*

Hence, the row rank and column rank of A are simply called the *rank* of A , denoted $rk(A)$. The following is immediate:

Fact 11. *For any $m \times n$ matrix A , $rk(A) \leq \min\{n, m\}$.*

A is said to have *full rank*, if $rk(A) = \min\{m, n\}$.

A square $n \times n$ matrix A is called *nonsingular* (or *regular*), if it has full rank, i.e., $rk(A) = n$.

4.4 Matrix inverse

The $n \times n$ *identity matrix*, denoted \mathbf{I} , is the matrix all of whose entries are 0, except for the diagonal, which is all-ones.

The *inverse* of a square $n \times n$ matrix A is an $n \times n$ matrix A^{-1} satisfying:

$$AA^{-1} = A^{-1}A = \mathbf{I}.$$

Not every square matrix has an inverse:

Theorem 12. *A is invertible if and only if it is nonsingular.*

4.5 Systems of linear equations

A *system of linear equations* with n variables (called also “unknowns”) is specified by an n -dimensional column vector x of variables, by an $m \times n$ matrix A of *coefficients*, and by an m -dimensional column vector of *free coefficients*. The goal is to solve the n linear equations specified compactly as:

$$Ax = y.$$

The *solution space* of the system is the following set of vectors:

$$\{z \in \mathbb{R}^n \mid Az = y\}.$$

If y is the $\mathbf{0}$ vector, the system is called *homogeneous* and the solution space is called the *null space* of A .

Fact 13. *The null space of A is a linear subspace of \mathbb{R}^n .*

Theorem 14. *If A is a square nonsingular matrix, then for every y , the solution space of $Ax = y$ contains exactly one solution. If A is either non-square or square but singular, then the solution space may be either empty or contain multiple solutions.*

The standard algorithm for finding the solutions for a system of linear equation is *Gaussian Elimination*. If A is an $n \times n$ matrix, the complexity of this algorithm is $O(n^3)$ scalar operations. Using fast matrix multiplication, it is possible to solve a system in $O(n^{2.36\dots})$ time.

4.6 Basis change matrix

Suppose x_1, \dots, x_n and y_1, \dots, y_n are two different bases of \mathbb{R}^n . Now, we are given a vector $x \in \mathbb{R}^n$ represented as a linear combination $\sum_{i=1}^n a_i x_i$ of x_1, \dots, x_n and we wish to find its representation as a linear combination $\sum_{i=1}^n b_i y_i$ of y_1, \dots, y_n . How do we do it?

Let $a = (a_1, \dots, a_n)$ be the vector of coefficients of x in its representation in the basis x_1, \dots, x_n . Let $b = (b_1, \dots, b_n)$ be the (unknown) representation of x in the basis y_1, \dots, y_n . Let X be the $n \times n$ matrix whose columns are x_1, \dots, x_n and let Y be the $n \times n$ matrix whose columns are y_1, \dots, y_n . Then, we have:

$$Xa = Yb.$$

Therefore,

$$b = Y^{-1}Xa.$$

(Note that Y has a full column rank and is therefore invertible.) The matrix $Y^{-1}X$ is called the *basis change matrix* from x_1, \dots, x_n to y_1, \dots, y_n .

4.7 Determinants

Let A be an $n \times n$ square matrix. A *generalized diagonal* of A is a sequence of n entries of A s.t. no two of them belong to the same row and no two of them belong to the same column. It is easy

to check that the family of generalized diagonals of A corresponds exactly to all the permutations in S_n . Each permutation $\sigma \in S_n$ corresponds to the generalized diagonal $\{A_{i,\sigma(i)} \mid i = 1, \dots, n\}$.

The *determinant* of A is the following sum over all the generalized diagonals of A :

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \prod_{i=1}^n A_{i,\sigma(i)}.$$

$\operatorname{sgn}(\sigma)$ is the *sign* of the permutation σ . It is defined as follows. Any permutation can be transformed into the identity permutation using a sequence of transpositions (i.e., swaps between two elements), similarly to how bubblesort sorts an array. While there can be many transposition sequences leading from σ to the identity permutation, it turns out that these sequences are either always of odd length or always of even length. The sign is then defined as follows:

$$\operatorname{sgn}(\sigma) = \begin{cases} -1, & \text{if the number of transpositions needed to transform } \sigma \text{ to the identity is odd,} \\ 1, & \text{otherwise.} \end{cases}$$

Although $\det(A)$ is defined as a sum over exponentially many diagonals, it can be computed in $O(n^3)$ time using Gaussian Elimination.

Theorem 15. $\det(A) \neq 0$ if and only if A is nonsingular.

5 Orthonormal bases

A basis x_1, \dots, x_n is called *orthonormal*, if:

1. For every $i = 1, \dots, n$, $\|x_i\| = 1$ (i.e., x_i is normalized).
2. For every $1 \leq i \neq j \leq n$, $x_i^T x_j = 0$ (i.e., x_i and x_j are orthogonal).

For example, the standard basis of \mathbb{R}^n is orthonormal.

Fact 16. *Every vector space has an orthonormal basis.*

In fact, a vector space may have many orthonormal bases.

Proposition 17. *Let x_1, \dots, x_n be an orthonormal basis, and let X be an $n \times n$ matrix whose columns are x_1, \dots, x_n . Then, $X^{-1} = X^T$.*

Proof. We show that $X^T X = \mathbf{I}$. That would imply $X^{-1} = X^T$.

The (i, j) -th entry of $X^T X$ is the inner product of the i -th row of X^T and the j -th column of X . Yet, the i -th row of X^T is exactly the i -th column of X . Hence, the (i, j) -th entry of $X^T X$ is the inner product of x_i and x_j . Since x_1, \dots, x_n is an orthonormal basis, this inner product is 1, if $i = j$, and is 0, if $i \neq j$. This means that $X^T X = \mathbf{I}$. \square

Suppose $x \in \mathbb{R}^n$ is represented in the standard basis. Let x_1, \dots, x_n be some orthonormal basis. Let $x = \sum_{i=1}^n a_i x_i$ be the representation of x in this basis. We next show that the vector $a = (a_1, \dots, a_n)$ of coefficients has a special form.

The $n \times n$ matrix whose columns are the standard basis vectors is the identity matrix \mathbf{I} . Let X be the $n \times n$ matrix whose columns are x_1, \dots, x_n . By what we've shown in Section 4.6, $a = X^{-1}\mathbf{I}x = X^{-1}x$. By Proposition 17, $X^{-1} = X^T$, and thus $a = X^T x$. Therefore, a_i equals the inner product of the i -th row of X^T and x . Yet, the i -th row of X^T is the i -th column of X , which equals x_i . We conclude:

$$a_i = x_i^T x.$$

The following theorem shows that changing from one orthonormal basis to another does not change the norm:

Theorem 18 (Bessel's identity). *Let x_1, \dots, x_n and y_1, \dots, y_n be two orthonormal bases of \mathbb{R}^n . Let $\sum_{i=1}^n a_i x_i$ and $\sum_{i=1}^n b_i y_i$ be the representations of the same vector x in these two bases. Then,*

$$\|a\|_2 = \|b\|_2.$$

6 Eigenvalues

Let A be a square $n \times n$ matrix. A scalar $\lambda \in \mathbb{R}$ is said to be an *eigenvalue* of A , if there exists a vector $x \in \mathbb{R}^n$, $x \neq \mathbf{0}$, s.t.

$$Ax = \lambda x.$$

x is called a *right eigenvector* of A corresponding to the eigenvalue λ .

If there exists a vector $y \in \mathbb{R}^n$, $y \neq \mathbf{0}$ s.t.

$$y^T A = \lambda y^T,$$

then y is called a *left eigenvector* of A corresponding to the eigenvalue λ .

Some facts:

1. There exists a right eigenvector of A corresponding to an eigenvalue λ if and only if there exists a left eigenvector of A corresponding to λ . Therefore, there is no need to distinguish between "right eigenvalues" and "left eigenvalues". A right eigenvector, though, is not necessarily a left eigenvector and vice versa.
2. If A is symmetric (i.e., $A^T = A$), then every right eigenvector of A is also a left eigenvector of A and vice versa.

The rest of the discussion focuses on right eigenvectors. Everything we say is applicable to left eigenvectors as well.

Some properties of eigenvalues and eigenvectors:

1. If x is a right eigenvector of A corresponding to λ , then for any scalar $a \in \mathbb{R}$, $a \neq 0$, also ax is a right eigenvector of A corresponding to λ .
2. If x is a right eigenvector of A corresponding to λ , then for any integer $k \geq 1$, x is also a right eigenvector of A^k corresponding to the eigenvalue λ^k .
3. If x_1, \dots, x_k are right eigenvectors of A corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_k$, respectively, then x_1, \dots, x_k are linearly independent.
4. 0 is an eigenvalue of A if and only if A is singular.

Fix any eigenvalue λ of A , and let V_λ be the set of right eigenvectors of A corresponding to λ :

$$V_\lambda = \{x \in \mathbb{R}^n \mid Ax = \lambda x\}.$$

Lemma 19. *For any eigenvalue λ , V_λ is a linear subspace of \mathbb{R}^n .*

V_λ is called the *eigenspace* corresponding to λ . The dimension of this eigenspace is called the *multiplicity* of λ .

Theorem 20. *Any $n \times n$ matrix A has at most k distinct eigenvalues. The sum of their multiplicities equals n .*

For a general matrix, not all eigenvalues are necessarily real (some may be complex). However,

Theorem 21. *If A is a symmetric matrix, then all its eigenvalues are real.*

Furthermore, if A is symmetric, then there exists an orthonormal basis of \mathbb{R}^n consisting of right eigenvectors of A .

The eigenvalue of A of largest absolute value is called the *principal eigenvalue* of A . Any eigenvector corresponding to this eigenvalue is called a *principal eigenvector*.

7 Markov chains

Let $V = \{1, 2, \dots, n\}$ be a finite *state space*. Suppose we can place a “token” on one of these states and then move it around from state to state. A *Markov chain* on V is a stochastic process specifying an infinite sequence of token moves. The Markov chain is specified by an $n \times n$ *probability transition matrix* P . P_{ij} is the probability to move the token from state i to state j , given that the token is at state j . Note that all entries of P are nonnegative and the sum of each row of P is 1. Hence, P is a *stochastic matrix*.

A Markov chain is a *memoryless* process: the probability to move from i to j is independent of the history of the process (i.e., what was the route that led to state i).

Example 22. A *simple random walk* on a graph is an example of a Markov chain. Let $G = (V, E)$ be a directed graph. We define a matrix P as follows:

$$P_{ij} = \begin{cases} \frac{1}{d_{\text{out}}(i)}, & \text{if } (i, j) \in E, \\ 0, & \text{otherwise.} \end{cases}$$

Here, $d_{\text{out}}(i)$ is the *out-degree* of i (i.e., the number of nodes to which there is an edge from i). Note that if G is undirected, then P is a symmetric matrix.

Suppose ϕ_0 is some *initial distribution* over the state space V . Suppose we choose a random state according to ϕ_0 and then start a Markov chain from this state. What will be the distribution over the states after one step? Let us call this distribution ϕ_1 . It is easy to check that, if we view ϕ_0, ϕ_1 as n -dimensional vectors, we have:

$$\phi_1^T = \phi_0^T P.$$

Similarly, for any step $t = 1, 2, \dots$,

$$\phi_t^T = \phi_{t-1}^T P.$$

Therefore, by induction we have:

$$\phi_t^T = \phi_0^T P^t.$$

Thus, any initial distribution ϕ_0 induces an infinite sequence of distributions $\phi_0, \phi_0 P, \dots, \phi_0 P^t, \dots$. We next show that under certain conditions on the matrix P , this sequence is guaranteed to converge.

Given a Markov chain $MC = (V, P)$, we define the *Markov chain graph*, G_{MC} , as follows:

- The vertex set of G_{MC} is V .
- (i, j) is an edge if and only if $P_{ij} > 0$.

Note that if MC is a simple random walk on a graph G , then $G_{MC} = G$.

We now define the following special types of Markov chains:

- A Markov chain MC is called *irreducible*, if the graph G_{MC} is strongly connected (that is, there exists a directed path from every node $i \in V$ to every other node $j \in V$).
- A Markov chain MC is called *aperiodic*, if for every $i, j \in V$, the g.c.d. of the lengths of the paths leading from i to j in G_{MC} is 1. In other words, starting from some integer k , there exists a path from i to j whose length is $\ell \geq k$.

Examples of graphs that are periodic are: (1) An undirected bipartite graph (the length of each path between two nodes on the same side of the graph is always even); (2) A directed cycle (the length of each path between i and itself must divide by the length of the cycle).

It can be shown that an undirected graph is aperiodic if and only if it is non-bipartite.

- A Markov chain MC is called *ergodic*, if it is both irreducible and aperiodic.

The following is probably the most basic theorem in the theory of Markov chains:

Theorem 23. If $MC = (V, P)$ is ergodic, then there exists a distribution π on V (called the limit distribution, the equilibrium distribution, or the stationary distribution), s.t. for every initial distribution ϕ_0 on V ,

$$\lim_{t \rightarrow \infty} \phi_0^T P^t = \pi^T.$$

That is, *irrespective of the initial distribution*, the Markov chain converges to the *same* limit distribution!

Markov chains are then a useful tool for sampling. Suppose we wish to sample from the state space V according to the distribution π , yet direct sampling is hard (e.g., because the state space is large and not known in its entirety in advance). If we have a Markov chain MC whose limit distribution is π , we can use it to sample from π as follows: We pick an arbitrary state $v_0 \in V$, and start the Markov chain from v_0 . If we run it for sufficiently many steps, we know that the distribution of the reached state is close to π . We can thus stop the Markov chain and output the reached state as a sample.

The limit distribution π is also stationary, meaning:

$$\pi^T = \pi^T P.$$

That is, if the current state is already distributed according to π , then subsequent states will also be distributed according to π .

Example 24. If $MC = (V, P)$ is a simple random walk on an undirected, connected, and non-bipartite graph G , then the stationary distribution of MC is the following:

$$\pi(i) = \frac{d_i}{2|E|},$$

where d_i is the degree of the i -th node. In particular, if G is a regular graph (i.e., all nodes have the same degree), then π is the uniform distribution on V .

Algebraically, π is a left eigenvector of P corresponding to the eigenvalue 1. It can be checked that since P is a stochastic matrix, 1 is its principal eigenvalue. Therefore, π is a principal left eigenvector of P .