Université Paris I Panthéon-Sorbonne

Lecture Notes

Master Methematical Models in Economics and Finance (MMEF)

BASIC NOTIONS OF LINEAR ALGEBRA

(short summary)

Michel GRABISCH

1 Vector spaces

A *vector space* over $\mathbb{R}$ is a set V closed under addition (associative and commutative, with a neutral element $\vec{0}$ (the zero vector), and additive inverses), and *scalar multiplication*, i.e., multiplication of a vector by a real number, satisfying the following properties for all $a, b \in \mathbb{R}$ and $x, y \in V$:

$$a(x+y) = ax + ay, \quad (a+b)x = ax + by, \quad a(bx) = (ab)x, \quad 1x = x.$$

In the whole document, we will restrict to vector spaces which are subsets of $\mathbb{R}^n$*, for some* $n \in \mathbb{N}$*.*

Vectors are represented as columns, e.g., $x = \begin{pmatrix} 1 \\ 4 \\ 0 \\ -2 \end{pmatrix}$.

A *subspace* of a vector space V is a subset of V which is a vector space.

A *linear combination* of vectors $x_1, \ldots, x_k \in V$ is any expression $\alpha_1 x_1 + \cdots + \alpha_k x_k$ with $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$. The *span* of $x_1, \ldots, x_k$ is the set of all their linear combinations:

$$\operatorname{span}\{x_1, \ldots, x_k\} = \{\alpha_1 x_1 + \cdots + \alpha_k x_k \;:\; \alpha_1, \ldots, \alpha_k \in \mathbb{R}\}.$$

$x_1, \ldots, x_k \in V$ are *linearly dependent* if there exist $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$, not all zero, such that

$$\sum_{i=1}^{k} \alpha_i x_i = \vec{0} \quad \text{(zero vector)}$$

$x_1, \ldots, x_k \in V$ are *linearly independent* if they are not linearly dependent, i.e., for all $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$,

$$\sum_{i=1}^{k} \alpha_i x_i = \vec{0} \Rightarrow \alpha_1 = \cdots = \alpha_k = 0.$$

$\{x_1, \ldots, x_k\}$ is a basis of V if $\operatorname{span}\{x_1, \ldots, x_k\} = V$ and $x_1, \ldots, x_k$ are linearly independent. Consequently, any $v \in V$ has a unique expression as a linear combination of $x_1, \ldots, x_k$. The *dimension* of V is the size (cardinality) of a basis of V.

2 Matrices

A $m \times n$ *matrix* is an array of numbers in $\mathbb{R}$ with m rows and n columns. The usual notation is $A = [a_{ij}]$, where a_{ij} is the entry of A at row i and column j.

The *transpose* of a $m \times n$ matrix $A = [a_{ij}]$ is the $n \times m$ matrix $A^T = [a_{ji}]$.

The *trace* of a $m \times n$ matrix $A = [a_{ij}]$ is defined by

$$\operatorname{tr} A = \sum_{i=1}^{k} a_{ii}, \quad \text{with } k = \min(m, n).$$

Any $m \times n$ matrix A defines a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ by:

$$x \in \mathbb{R}^n \mapsto Ax = \begin{bmatrix} \sum_{j=1}^{n} a_{1j} x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj} x_j \end{bmatrix} \in \mathbb{R}^m.$$

The *range* of A is the range (image) of the corresponding linear mapping, i.e.,

$$\text{range}A = \{y \in \mathbb{R}^m \ : \ y = Ax \text{ for some } x \in \mathbb{R}^n\}.$$

The *null space* or *kernel* of A is defined by

$$\text{Ker}A = \{x \in \mathbb{R}^n \ : \ Ax = \vec{0}\}$$

A fundamental result (called *rank-nullity theorem*) says that

$$\dim(\text{range}A) + \dim(\text{Ker}A) = n.$$

Matrix operations:

(i) For $A, B \in \mathbb{R}^{m\times n}$ with $A = [a_{ij}]$ and $B = [b_{ij}]$, $A + B = [a_{ij} + b_{ij}]$.

(ii) For $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, $AB = [\sum_{k=1}^{n} a_{ik}b_{kj}] \in \mathbb{R}^{m\times p}$.

The *identity matrix of order* n, denoted by I_n, is a $n \times n$ matrix given by

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Remark 1. (i) If $x, y \in \mathbb{R}^n$, $x^T y \in \mathbb{R}$ and $xy^T \in \mathbb{R}^{n\times n}$, as a vector is considered as an $n \times 1$ matrix.

(ii) Let $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. Then $Ax \in \mathbb{R}^m$ is a linear combination of the columns of A, while $y^T A \in \mathbb{R}^n$ is a linear combination of the rows of A.

3 Determinants

For square matrices, *determinants* are defined inductively by:

- For a 1×1 matrix $[a_{11}]$: $\det[a_{11}] = a_{11}$.
- Otherwise,

$$\det A = \sum_{k=1}^{n} (-1)^{i+k} a_{ik} \det A_{ik} = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det A_{kj},$$

for arbitrary i, j, and A_{ik} is the matrix A without row i and column k.

For example, with $n = 2$:

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Important results:

(i) $\det A^T = \det A$

(ii) $\det AB = \det A \, \det B$

(iii) $\det I_n = 1$

(iv) $\det A = 0$ if and only if a subset of the row vectors (equiv., column vectors) of A are linearly dependent.

(v) If a row of A is $\vec{0}^T$, then $\det A = 0$.

To each matrix $A \in \mathbb{R}^{m\times n}$ corresponds a unique *reduced row echelon form (RREF)* (also called *Hermite normal form*) such that:

(i) Any zero row occurs at the bottom of the matrix

(ii) The *leading entry* (i.e., the first nonzero entry) of any nonzero row is 1

(iii) All other entries in the column of a leading entry are zero

(iv) The leading entries occur in a stair step pattern, from left to right: leading entry $a_{ik} \Rightarrow$ leading entry $a_{i+1,\ell}$ (if it exists) with $\ell > k$.

Example of a RREF:

$$A = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & -3 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The RREF is obtained from a matrix by

(i) interchanging rows

(ii) multiply a row by a nonzero scalar

(iii) a row is replaced by the sum of itself and another row multiplied by a scalar.

Important result: for a matrix $A \in \mathbb{R}^{n\times n}$, $\det A \neq 0$ if and only if its RREF is I_n.

4 Rank and nonsingularity ; inverse

The *rank* of a matrix $A \in \mathbb{R}^{m\times n}$ is the dimension of its range, i.e., the cardinality of a largest linearly independent set of columns (equiv., of rows) of A.

Important results:

- $\text{rank}A = \text{rank}A^T$
- $\text{rank}A$ is the rank of its RREF, which is the number of leading entries.

Theorem 1 (characterization of the rank). Let A be a $m \times n$ matrix. The following are equivalent:

(i) $\text{rank}A = k$

(ii) k, and no more than k, rows of A are linearly independent

(iii) k, and no more than k, columns of A are linearly independent

(iv) Some $k\times k$ submatrix of A has a nonzero determinant, and any $(k+1)\times(k+1)$ submatrix has a zero-determinant

(v) $k = n - \dim(\text{Ker}A)$ (rank-nullity theorem).

A matrix $A \in \mathbb{R}^{m\times n}$ is *nonsingular* if $Ax = \vec{0} \Leftrightarrow x = \vec{0}$. Otherwise, A is *singular*. Observe that if $m < n$ then A is singular.

A matrix $A \in \mathbb{R}^{n\times n}$ is *invertible* if there exists a matrix $A^{-1} \in \mathbb{R}^{n\times n}$ such that $A^{-1}A = AA^{-1} = I_n$. Note that $\det A^{-1} = \dfrac{1}{\det A}$.

Theorem 2 (characterization of nonsingularity). Let $A \in \mathbb{R}^{n\times n}$. The following are equivalent:

(i) A is nonsingular

(ii) A^{-1} exists

(iii) $\text{rank}A = n$

(iv) rows are linearly independent

(v) columns are linearly independent

(vi) $\det A \neq 0$

(vii) $\dim(\text{range}A) = n$

(viii) $\dim(\text{Ker}A) = 0$

5 Linear systems

A *linear system of equalities* has the form

$$\begin{cases} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ & \vdots & & & \vdots & & \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & = & b_m \end{cases}$$

with $a_{ij}, b_j \in \mathbb{R}$ for all i, j. Using matrix notation, this can be rewritten as

$$Ax = b$$

with $A = [a_{ij}]$, $b^T = \begin{bmatrix} b_1 & \cdots & b_m \end{bmatrix}$, and $x^T = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$.

The *Gauss-Jordan elimination method*, which leads to the set of solutions of the system, consists in putting the *augmented matrix* $[A\ b]$ in RREF. Indeed, $A_1x = b_1$ and $A_2x = b_2$ have the same set of solutions $\Leftrightarrow$ $[A_1\ b_1]$ and $[A_2\ b_2]$ have the same RREF.

A linear system is *consistent* if there exists at least one solution. Otherwise, the linear system is *inconsistent*.

Theorem 3 (characterization of consistency). Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. The linear system $Ax = b$ is consistent if and only if $\text{rank}[A\ b] = \text{rank}A$.

(set of solutions) Suppose $Ax = b$ is consistent, with solution x_0. Observe that x_0' is solution iff $Ax_0' = b = Ax_0$ iff $A(x_0' - x_0) = 0$ iff $x_0' - x_0 \in \text{Ker}A$. Consequently, the set of solutions has the form

$$\{x_0\} + \text{Ker}A$$

where "+" is understood in the sense of subspaces. Therefore, the dimension of the (affine) subspace of solutions is $\dim(\text{Ker}A)$.

Theorem 4 (characterization of consistent square linear systems). Let $A \in \mathbb{R}^{n\times n}$. The following are equivalent:

(i) $Ax = b$ is consistent for each $b \in \mathbb{R}^n$

(ii) $Ax = \vec{0}$ has a unique solution, which is $x = \vec{0}$

(iii) $Ax = b$ has a unique solution for each $b \in \mathbb{R}^n$

(iv) A is nonsingular

(v) A^{-1} exists

(vi) $\text{rank}A = n$.

If one of the above assertions holds, then the unique solution is $x = A^{-1}b$.

(back to Gauss-Jordan elimination) suppose $[A\ b]$ of the consistent linear system $Ax = b$ has been put in RREF. According to Theorem 3, the number of leading variables (entries) is the rank of A, the remaining variables are the *free variables*, whose number gives the dimension of $\text{Ker}A$.

$Ax = b$ is inconsistent if and only if in the RREF of $[A\ b]$ there is a row of the form $\begin{bmatrix}0 & \cdots & 0 & a\end{bmatrix}$ with $a \neq 0$.

Example 1. Consider the linear system

$$\begin{cases} 2x + y - z + 3t = 1 \\ 4x + 2y - z + 4t = 5 \\ 2x + y + t = 4 \end{cases}$$

The augmented matrix is

$$[A\ b] = \begin{bmatrix} 2 & 1 & -1 & 3 & 1 \\ 4 & 2 & -1 & 4 & 5 \\ 2 & 1 & 0 & 1 & 4 \end{bmatrix}$$

Let us put it in echelon form[1]. Subtracting 2 times row 1 from row 2, and subtracting row 1 from row 3 yield

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

Now, adding row 2 and minus row 3 yields

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

Finally, adding the two first rows yields

$$\begin{bmatrix} 2 & 1 & 0 & 1 & 4 \\ 0 & 0 & 1 & -2 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

[1] In putting in RREF, to solve linear systems, it is not necessary to make the leading entry equal to 1.

Then the system has a solution. There are two free variables y and t, therefore the dimension of the subspace of solutions is 2. Let us express the set of solutions. The system is

$$\begin{cases} 2x + y + t = 4 \\ z - 2t = 3 \end{cases}$$

Putting the free variables on the right handside yields:

$$\begin{cases} 2x = 4 - y - t \\ z = 3 + 2t \end{cases}$$

Hence, finally the set of solutions is given by

$$\{(x, y, z, t) \in \mathbb{R}^4 \ : \ x = 2 - \frac{1}{2}y - \frac{1}{2}t, \ z = 3 + 2t, \ y, t \in \mathbb{R}\}.$$

In particular, $(2, 0, 3, 0)$ is a solution.

6 Introduction to eigenvalues and eigenvectors

Basic definitions. Given a square matrix $A \in \mathbb{C}^{n\times n}$, if there exists a scalar $\lambda \in \mathbb{C}$ and a nonzero vector $x \in \mathbb{C}^n$ such that

$$Ax = \lambda x,$$

then λ is an *eigenvalue* of A and x is an *eigenvector* of A. We say that (λ, x) is an *eigenpair* and the *eigenspace* of A associated to λ is the vector subspace $\{x \in \mathbb{C}^n \ : \ Ax = \lambda x\}$.

Eigenvectors of distinct eigenvalues are linearly independent.

$y \in \mathbb{C}^n$ is a *left eigenvector* of A associated to λ if $y^*A = \lambda y^*$, where $()^*$ indicates the conjugate transpose.

The *spectrum* of A, denoted by $\sigma(A)$, is the set of all eigenvalues. Remark that $\vec{0} \in \sigma(A)$ iff A is singular. The *spectral radius* of A is defined by

$$\rho(A) = \max\{|\lambda|, \lambda \in \sigma(A)\}.$$

Finding eigenvalues. Eigenvalues can be found as the solutions of the *characteristic polynomial* of A:

$$\det(\lambda I_n - A) = 0.$$

Indeed, $Ax = \lambda x$ is equivalent to the system $(A - \lambda I_n)x = \vec{0}$, which can admit a nonzero solution x if and only if $A - \lambda I_n$ is singular, i.e., with zero determinant.

The characteristic polynomial is a polynomial in λ of degree n, which therefore admits n (complex) solutions, not necessarily distinct.

Important properties:

- The trace of A is the sum of the eigenvalues: $\text{tr}A = \sum_{i=1}^{n} \lambda_i$
- The determinant of A is the product of the eigenvalues: $\det A = \prod_{i=1}^{n} \lambda_i$

Multiplicities. Let $\lambda_1, \ldots, \lambda_q$ be the distinct eigenvalues of A. The *algebraic multiplicity* α_i of λ_i is the multiplicity of λ_i as a root of the characteristic polynomial. It holds

$$\sum_{i=1}^{q} \alpha_i = n.$$

The *geometric multiplicity* γ_i of λ_i is the dimension of its eigenspace (dimension of the kernel of $A - \lambda_i I_n$). We have

$$1 \leqslant \gamma_i \leqslant \alpha_i, \quad i = 1, \ldots, q.$$

The eigenvalue λ_i is *simple* if $\alpha_i = 1$. It is *semi-simple* if $\alpha_i = \gamma_i$.

Diagonalization. A is said to be *diagonalizable* if there exists a nonsingular matrix $S \in \mathbb{C}^{n\times n}$ such that SAS^{-1} is a diagonal matrix.

Let $A \in \mathbb{C}^{n\times n}$, and $\lambda_1, \ldots, \lambda_q$ be its (distinct) eigenvalues. Then A is diagonalizable iff $\sum_{i=1}^{q} \gamma_i = n$.

This amounts to say that A is diagonalizable if and only if there exists n linearly independent eigenvectors $x^1, \ldots, x^n$, in which case $S = [x^1 \ \cdots \ x^n]$, and

$$SAS^{-1} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

Jordan decomposition. If a matrix is not diagonalizable, it can be always put into its Jordan form. A *Jordan block* of size m is a $m \times m$ matrix of the form

$$J_m(\lambda) = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix}$$

and $J_1(\lambda) = [\lambda]$. A *Jordan matrix* is block-diagonal and each block is a Jordan block.

Theorem 5 (Jordan decomposition). Let $T \in \mathbb{C}^{n\times n}$. There exists a nonsingular matrix $S \in \mathbb{C}^{n\times n}$ such that

$$T = S \begin{bmatrix} J_{m_1}(\lambda_1) & & & \\ & J_{m_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{m_q}(\lambda_q) \end{bmatrix} S^{-1}$$

with $\sum_{i=1}^{q} m_i = n$, $\lambda_1, \ldots, \lambda_q$ are the eigenvalues of T, and the geometric multiplicity of λ_i is equal to the number of blocks $J_{m_i}(\lambda_i)$, while the algebraic multiplicity is the sum of the sizes of blocks $J_{m_i}(\lambda_i)$.

If all eigenvalues are semi-simple, then the columns of S are the right eigenvectors, while the rows of S^{-1} are the left eigenvectors.

Let $T \in \mathbb{C}^{m\times m}$ with Jordan reduction SJ_TS^{-1}. Then

$$T^k = SJ_T^kS^{-1}.$$

A square matrix A is *semi-convergent* if $\lim_{k\to\infty} A^k$ exists, and it is *convergent* if in addition this limit is the matrix 0.

We have the following properties:

- A Jordan block is convergent iff $|\lambda| < 1$;
- A Jordan block of size 1 is semi-convergent iff $|\lambda| < 1$ or $\lambda = 1$.

From this we deduce the convergence of T^k:

Theorem 6.

- T is convergent iff $\rho(T) < 1$;
- T is semi-convergent iff either $\rho(T) < 1$ or 1 is a semisimple eigenvalue and all other eigenvalues have modulus less than 1.