E600 Mathematics

Chapter 2: Matrix Algebra

Martin Reinhard

September 16, 2020

- Linear equation systems are a fundamental building block of advanced mathematics
- Optimization: first and second derivatives of more general functions typically have matrix form
- Statistical programming (Applied Econometrics) and also Quantitative Macro build heavily on matrix algebra – ever wondered why economists prefer MATLAB?
- Matrices facilitate otherwise complex problems (recall: LI test)

Consider the system

$$\begin{array}{ccccccc} x_1 & + & x_2 & + & x_3 & = & 6 \\ & & x_2 & - & x_3 & = & 0 \\ 5 \cdot x_1 & & & + & x_3 & = & 1 \end{array}$$

- Important questions of general interest:
  1. Does a solution exist?
  2. If so, is the solution unique? Otherwise, how many solutions are there?
  3. How can we (or our computer) efficiently derive the (set of) solutions?
- Matrix representation "$Ax = b$":

$$\begin{pmatrix} 1 \cdot x_1 + 1 \cdot x_2 + 1 \cdot x_3 \\ 0 \cdot x_1 + 1 \cdot x_2 + (-1) \cdot x_3 \\ 5 \cdot x_1 + 0 \cdot x_2 + 1 \cdot x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \\ 1 \end{pmatrix} \quad \Leftrightarrow \quad \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 5 & 0 & 1 \end{pmatrix}}_{=A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{=x} = \underbrace{\begin{pmatrix} 6 \\ 0 \\ 1 \end{pmatrix}}_{=b}$$

$\Rightarrow$ we can use A (and b) to answer 1. – 3.!

Definition (Matrix of dimension $n \times m$)

Let $(a_{ij} : i \in \{1, \dots, n\}, j \in \{1, \dots, m\})$ be a collection of elements from a basis set X, i.e. $\forall i \in \{1, \dots, n\}, j \in \{1, \dots, m\} : a_{ij} \in X$. Then, the matrix A of these elements is the ordered two-dimensional collection

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}.$$

We call n the row dimension and m the column dimension of A. We write $A \in X^{n \times m}$, and $A = (a_{ij})_{i \in \{1, \dots, n\}, j \in \{1, \dots, m\}}$, or, if $n = m$, $A = (a_{ij})_{i,j \in \{1, \dots, n\}}$.

- Real Matrix ($X = \mathbb{R}$) as a vector of vectors:
  - $a_i^r = (a_{i1}, \ldots, a_{im})$ i-th row of A (row vector)
  - $a_j^c = (a_{1j}, \ldots, a_{mj})' = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$ j-th column of A (column vector)

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} = \begin{pmatrix} a_1^r \\ \vdots \\ a_n^r \end{pmatrix} = (a_1^c, \ldots, a_m^c)$$

  - Typically: leave r/c away if clear from context
- Vector space of real matrices: not vectors of real *numbers* (as $\mathbb{R}^n$), but vectors of real *vectors*!

Set of real $m \times n$-Matrices: $\mathcal{M}_{n\times m} = \{A \in \mathbb{R}^{n\times m}\}$

Definition (Addition of Matrices)

For matrices A, B of identical dimension, i.e. $\exists n, m \in \mathbb{N} : A, B \in \mathcal{M}_{n\times m}$, their sum is obtained from addition of their elements, that is

$$A + B = (a_{i,j} + b_{i,j})_{i\in\{1,\ldots,n\}, j\in\{1,\ldots,m\}}.$$

Definition (Scalar Multiplication of Matrices)

Let $A = (a_{i,j})_{i\in\{1,\ldots,n\}, j\in\{1,\ldots,m\}}$ and $\lambda \in \mathbb{R}$. Then, scalar multiplication of A with λ is defined element-wise, that is

$$\lambda A := (\lambda a_{i,j})_{i=1,\ldots,n, j=1,\ldots,m}.$$

Self-study: $(\mathcal{M}_{n\times m}, +, \cdot)$ is a real vector space!

- Maximum Norm of $(\mathcal{M}_{n\times m}, +, \cdot)$:

$$\|A\|_\infty = \max\{|a_{ij}| : i \in \{1, \ldots, n\}, j \in \{1, \ldots, m\}\}$$

1. $\|A\|_\infty \geq 0$ and $(\|A\|_\infty = 0 \Leftrightarrow \forall i, j : a_{ij} = 0$, i.e. $A = \mathbf{0}_{n\times m})$
2. Triangle inequality transfers from $|\cdot|$
3. Absolute homogeneity easily checked as well $\Rightarrow$ norm ✓

- Can define norm-induced metrics, convergence, etc. like for $\mathbb{R}^n$!

- $A \in \mathbb{R}^{n \times m}$ is **equal** to the matrix B if (and only if) $\in \mathbb{R}^{n \times m}$ and $\forall i,j : a_{ij} = b_{ij}$ (e.g. $\mathbf{0}_{2\times 2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \mathbf{0}_{2\times 3}$!)
- **Transposed** matrix $A'/A^T/A^t$: swap of row and column index, e.g.

$$\begin{pmatrix} 1 & 0 & 4 \\ 3 & 1 & 2 \end{pmatrix}' = \begin{pmatrix} 1 & 3 \\ 0 & 1 \\ 4 & 2 \end{pmatrix}$$

- **Vector** of length n: (column) vector $a \in \mathbb{R}^{n \times 1}$, row vector $a \in \mathbb{R}^{1 \times n}$
- **Zero matrix**: $\mathbf{0}_{n \times m} \in R^{n \times m}$: $\mathbf{0}_{n \times m} = (a_{ij})_{i \in \{1,\ldots,n\}, j \in \{1,\ldots,m\}}$ where $\forall i,j : a_{ij} = 0$ (neutral element of addition!)

- Square matrix: $A \in \mathbb{R}^{n \times m}$ with $n = m$, i.e. $A \in \mathbb{R}^{n \times n}$
- Symmetric matrix: *square* matrix $A \in \mathbb{R}^{n \times n}$ with $a_{ij} = a_{ji}\ \forall i, j$, e.g.

$$\begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix} \quad \text{but not} \quad \begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix}$$

- Diagonal matrix: $A = (a_{ij})_{i,j \in \{1,\dots,n\}}$ with $(i \neq j \Rightarrow a_{ij} = 0)$, e.g.

$$\begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

  We write $A = diag\{a_{11}, a_{22}, \dots, a_{nn}\}$
- Identity matrix: $\mathbf{I_n} = diag\{1, 1, \dots, 1\}$ (write down $\mathbf{I_3}$)

Note: identity/zero matrix $<$ diagonal $<$ symmetric $<$ square

Definition (Upper and Lower Triangular Matrix)

A **square** *matrix* $A = (a_{ij})_{i,j \in \{1,\dots,n\}} \in \mathbb{R}^{n\times n}$ *is said to be upper triangular if* $(i > j \Rightarrow a_{ij} = 0)$*, i.e. when the entry* a_{ij} *equals zero whenever it lies below the diagonal. Conversely,* A *is said to be lower triangular if* A' *is upper triangular, i.e.* $(i < j \Rightarrow a_{ij} = 0)$*.*

Rather than studying the definition, the concept may be more straightforward to grasp by just looking at an upper triangular matrix:

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & -4 & 3 \\ 0 & 0 & 0 & 2 \end{pmatrix} \quad \text{for which} \quad A' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 1 & -4 & 0 \\ 4 & 0 & 3 & 2 \end{pmatrix}.$$

Definition (Matrix Product)

Consider $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{m \times k}$ where the column dimension of A is equal to the row dimension of B. Then, the matrix $C \in \mathbb{R}^{n \times k}$ of column dimension equal to the one of A and row dimension equal to the one of B is called the product of A and B, denoted $C = A \cdot B$, if

$$\forall i \in \{1, \ldots, n\}, j \in \{1, \ldots, k\} : c_{ij} = \sum_{l=1}^{m} a_{il} b_{lj}.$$

What is $A \cdot B$ if

- $A = \begin{pmatrix} 1 & 0 \\ 3 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} -1 & 2 & 4 \\ 0 & 1 & 3 \end{pmatrix}$?
- $A = \begin{pmatrix} -1 & 2 & 4 \\ 0 & 1 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 3 & 2 \end{pmatrix}$? Commutativity?

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \; B = \begin{pmatrix} b_1 & \cdots & b_k \end{pmatrix} \text{ (row/column notation). Then,}$$

$$AB = \begin{pmatrix} a_1' \cdot b_1 & a_1' \cdot b_2 & \cdots & a_1' \cdot b_k \\ a_2' \cdot b_1 & a_2' \cdot b_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1}' \cdot b_k \\ a_n' \cdot b_1 & \cdots & a_n' \cdot b_{k-1} & a_n' \cdot b_k, \end{pmatrix}$$

Example:

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 4 & 4 & 0 \\ -2 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 5 & -3 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 2 \cdot 0 + 3 \cdot 5 & 1 \cdot 0 + 2 \cdot 1 + 3 \cdot (-3) \\ 0 \cdot 1 + 1 \cdot 0 + 0 \cdot 5 & 0 \cdot 0 + 1 \cdot 1 + 0 \cdot (-3) \\ 4 \cdot 1 + 4 \cdot 0 + 0 \cdot 5 & 4 \cdot 0 + 4 \cdot 1 + 0 \cdot (-3) \\ (-2) \cdot 1 + 4 \cdot 0 + 1 \cdot 5 & (-2) \cdot 0 + 4 \cdot 1 + 1 \cdot (-3) \end{pmatrix} = \begin{pmatrix} 16 & -7 \\ 0 & 1 \\ 4 & 4 \\ 3 & 1 \end{pmatrix}$$

Theorem (Associativity and Distributivity of the Product)

Assuming that A, B, C are matrices of appropriate dimension, the product for matrices is

(i) *Associative:* $(AB)C = A(BC)$ (*order of multiplication is irrelevant!*)

(ii) *Distributive over matrix addition:* $A(B + C) = AB + AC$ *and* $(A + B)C = AC + BC$

Theorem (Zero and Identity matrix)

Let $A \in \mathbb{R}^{n \times m}$. *Then,*

(i) $A + \mathbf{0}_{n \times m} = A$.

(ii) *For any* $k \in \mathbb{N}$, $A \cdot \mathbf{0}_{m \times k} = \mathbf{0}_{n \times k}$ *and* $\mathbf{0}_{k \times n} \cdot A = \mathbf{0}_{k \times m}$.

(iii) *For any* $k \in \mathbb{N}$, $A \cdot \mathbf{I_m} = A$ *and* $\mathbf{I_n} \cdot A = A$.

Take-away: product (almost) behaves naturally! (exception: commutativity)

Theorem (Transposition, Sum, and Product)

(i) *Let* $A, B \in \mathbb{R}^{n\times m}$. *Then,* $(A+B)' = A' + B'$

(ii) *Let* $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{m\times k}$. *Then,* $(AB)' = B'A'$

(iii) *If* $A \in \mathbb{R}^{1\times 1}$, *then* A *is actually a scalar and* $A' = A$. *(?!)*

- Recall: scalar product compactly displays sum of products: $x \cdot y = x'y = \sum_{i=1}^{n} x_i y_i$
- More generally: when $A = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ and $B = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ with $a_j \in \mathbb{R}^{1\times m}$, $b_j \in \mathbb{R}^{1\times k}$. then $A'B = \sum_{i=1}^{n} a_i' b_i$ where $a_i' b_i \in \mathbb{R}^{m\times k}$
- Recall intuition: vectors of vectors vs. vectors of real numbers

- Let's verify our matrix representation $Ax = b$:

$$\begin{array}{ccccccc} 1 \cdot x_1 & + & 1 \cdot x_2 & + & 1 \cdot x_3 & = & 6 \\ 0 \cdot x_1 & + & 1 \cdot x_2 & + & (-1) \cdot x_3 & = & 0 \\ 5 \cdot x_1 & + & 0 \cdot x_2 & + & 1 \cdot x_3 & = & 1 \end{array} \Leftrightarrow \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 5 & 0 & 1 \end{pmatrix}}_{=A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{=x} = \underbrace{\begin{pmatrix} 6 \\ 0 \\ 1 \end{pmatrix}}_{=b}$$

- Now: (how) use A to characterize the solutions and determine them?
- Easiest scenarios: "square system" – as many equations as unknowns
- Simplest of all: one equation in one unknown: $ax = b$
  - If $a \neq 0$, then unique solution $x = a^{-1}b = b/a$
  - If $a = 0$ and $b = 0$: infinitely many solutions $x \in \mathbb{R}$
  - If $a = 0$ and $b \neq 0$: no solution ("b never reached by $a \cdot x$")

$\rightarrow$ how to generalize this for matrices?

- 1×1 system: generalization step
  - Unique solution if and only if a is invertible, i.e.

$$\exists a^{-1} \in \mathbb{R} : a^{-1}a = aa^{-1} = 1$$

$\rightarrow x^* = a^{-1}b$

  - Otherwise: infinitely many/no solutions; $b \in \text{im}(f_a)$, $f_a(x) = ax$?
- For $n \times n$ systems ($Ax = b$):
  - Unique solution if and only if A is invertible, i.e.

$$\exists A^{-1} \in \mathbb{R}^{n \times n} : A^{-1}A = AA^{-1} = \mathbf{I_n}$$

$\rightarrow x^* = A^{-1}b$ (why?)

  - Otherwise: infinitely many/no solutions; $b \in \text{im}(f_A)$, $f_A(x) = Ax$?
- Now: (i) when $\exists A^{-1}$? (ii) when $b \in \text{im}(f_A)$?

- Inverse matrix of square matrix $A \in \mathbb{R}^{n \times n}$: $A^{-1} \in \mathbb{R}^{n \times n}$ so that

$$A^{-1}A = AA^{-1} = \mathbf{I_n}$$

- The inverse matrix is unique!
  Proof. Let $B, C \in \mathbb{R}^{n \times n}$ and assume that $BA = AB = \mathbf{I_n}$ and $CA = AC = \mathbf{I_n}$. Then, it holds that

$$C = C\mathbf{I_n} = C(AB) = \underbrace{(CA)}_{=\mathbf{I_n}} B = \mathbf{I_n} B = B,$$

  i.e. $C = B$.

- (1) How to (dis)prove intertability? (2) How to find the inverse?

Proposition (Invertability: Transpose and Product)

Suppose that $A, B \in \mathbb{R}^{n\times n}$ are invertible. Then,

- *A' is invertible and $(A')^{-1} = (A^{-1})'$,*
- *AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$,*
- *$\forall \lambda \in \mathbb{R}$, $\lambda \neq 0$, λA is invertible and $(\lambda A)^{-1} = 1/\lambda A^{-1}$.*

Corollary: if $A_1, \ldots, A_k$ are invertible matrices, then

$$A_k^{-1}A_{k-1}^{-1}\ldots \underbrace{A_2^{-1}\underbrace{A_1^{-1}A_1}_{=\mathbf{I_n}}A_2}_{=\mathbf{I_n}}\ldots A_k = \mathbf{I_n} \quad \Rightarrow \quad \left(\Pi_{i=1}^{k}A_i\right)^{-1} = \Pi_{i=k}^{1}A_i^{-1}$$

. . . but how to proceed more fundamentally?

Definition (Elementary Matrix Operations)

$A = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ *with rows* $a_1', \dots, a_n' \in \mathbb{R}^n$, $\tilde{A} = \begin{pmatrix} \tilde{a}_1 \\ \vdots \\ \tilde{a}_n \end{pmatrix}$. *The three elementary matrix operations are mappings* $A \to \tilde{A}$ *given by*

(E1) *Interchange of two rows* i, j: $\tilde{a}_i = a_j$, $\tilde{a}_j = a_i$ *and* $\tilde{a}_k = a_k$ $\forall k \notin \{i, j\}$,

(E2) *Scalar multiplication of a row* i *with* $\lambda \neq 0$: $\tilde{a}_i = \lambda a_i$ *and* $\tilde{a}_j = a_j$ $\forall j \neq i$,

(E3) *Addition of a multiple of row* j *to row* i: $\tilde{a}_i = a_i + \lambda a_j$, $\lambda \in \mathbb{R}$, *and* $\tilde{a}_j = a_j$ *for all* $j \neq i$.

- Crucial tool for finding the inverse using Gauß-Jordan's algorithm
- Do not confuse with *basis* operations $+$ and $\cdot$!

- Elementary Operations (EO) can be represented by left-multiplication of operator matrix E: $\tilde{A} = EA$ (formal definition: see script)
- Generally: $e_i' A$ extracts the i-th row of A, and Ae_j the j-th column! $\rightarrow$ E stack extraction operators
- Give examples for 2×2 system
- Can perform multiple operations at once using the same logic!
- Example: 4×4 system, what do these EO matrices do?

$$E^1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad E^2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}, \quad E^3 = \begin{pmatrix} -3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

- How do EOs help in finding the inverse?

- Suppose we can bring A to an identity matrix using only EOs $E_1, \dots E_k$, i.e. $E_k E_{k-1} \dots E_1 A = \mathbf{I_n}$
- Then, $E = E_k E_{k-1} \dots E_1$ **fits the definition of A^{-1}!**
- $A^{-1} = E = E \cdot \mathbf{I_n} \rightarrow$ Gauß-Jordan algorithm
- General systems ($n \neq m$): EOs help characterize the number of solutions! Key concept: (generalized) upper triangular form

Definition (Generalized Upper Triangular Matrix)

Let $A \in \mathbb{R}^{n \times m}$. If $n \geq m$, A has generalized upper triangular form if $A = \begin{pmatrix} A_T \\ \mathbf{0_{n-m \times m}} \end{pmatrix}$ where $A_T \in \mathbb{R}^{m \times m}$ is upper triangular. If $n \leq m$, A has generalized upper triangular form if $A = (A_T, X)$ where $A_T \in \mathbb{R}^{m \times m}$ is upper triangular and X is an arbitrary $n \times m - n$ matrix.

Examlpes Generalized Upper Triangular Form: A, B, D, but not C

$$A = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 4 & 7 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 1 & 4 & 7 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Theorem (Triangulizability of a Matrix)

Consider a matrix $A \in \mathbb{R}^{n \times m}$. Then, if $n = m$, A can be brought to upper triangular form applying only elementary operations to A. Generally, A can be brought to generalized upper triangular form.

(Quite extensive) proof can be found in script

- Recall: If we can bring A to identity form, then we know how to invert it
- From the upper triangular form, the identity form is usually not far away: take for instance

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

- Condition (first step): non-zero diagonal
- Let's find more general (and direct) invertability conditions; concepts will also help characterize general equation systems
- Afterwards: solving (we actually already know how to)

- Consider $f_A(x) = Ax$ for a **square** matrix $A \in \mathbb{R}^{n \times n}$
- A invertible:
  - For any basis $B = \{b_1, \ldots, b_n\}$ of $\mathbb{R}^n$, $T_B(A) = \{Ab_1, \ldots, Ab_n\}$ is a basis of $f_A[\mathbb{R}^n] = \mathbb{R}^n$
  - Ax is proportional to x, magnitude is scaled by basis change
  - Determinant $\det(A) =$ *scaling factor* for basis; summarizes expansion/compression of $\|Ax\|$ relative to $\|x\|$
  - E.g.: $A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ has $\det(A) = 6 > 1$ (expansion); induces basis change $B_{can} = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \rightarrow T_{B_{can}}(A) = \left\{ \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \end{pmatrix} \right\}$
- A non-invertible:
  - $\dim(f_A[\mathbb{R}^n]) < n$, i.e., Ax "loses" dimensions

$\Rightarrow$ "Infinite compression", $\det(A) = 0$

- A_{-ij}: A w/o column i and row j, e.g.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 4 \end{pmatrix} \quad \rightarrow \quad A_{-23} = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$$

Definition (Determinant)

Let $A \in \mathbb{R}^{n\times n}$. Then, we define the determinant of A, denoted $\det(A)$ *or* $|A|$, *as*

(i) *if $n = 1$ and $A = (a)$ is a scalar,* $\det(A) = \det(a) := a$.

(ii) *for all $n \in \mathbb{N} : n \geq 2$, when* $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$,

$\det(A) := \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{-ij})$ *with* $i = 1$.

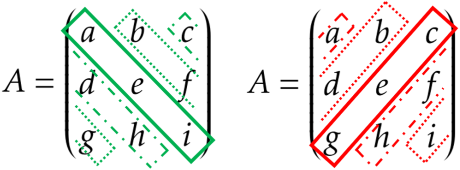
- Laplace Expansion Theorem: It holds that
  - $\det(A) := \sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det(A_{-ij})$ for arbitrary $i \in \{1, \dots, n\}$ (expansion by the i-th row)
  - $\det(A) := \sum_{i=1}^{n}(-1)^{i+j}a_{ij}\det(A_{-ij})$ for arbitrary $j \in \{1, \dots, n\}$ (expansion by the j-th column)

  $\rightarrow$ Look for rows/columns with many zeros (e.g. A on slide before)!
- Determinant of "small" matrices:
  (i) If $n = 2$ and $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then $\det(A) = ad - bc$.

  (ii) If $n = 3$ and $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$, then

  $\det(A) = aei + bfg + cdh - (ceg + bdi + afh)$.

- Usually: do Laplace expansions until one may use the 3x3 rule
- We typically consider at most 4x4 matrices analytically
- 3x3 graphically:



- Example: $\det \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ -1 & 0 & 4 \end{pmatrix}$

- The determinant of triangular matrices is equal to the trace (product of diagonal elements)

$$\det(A) = a_{11} \det \begin{pmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{32} & \cdots & a_{3n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix} = a_{11} a_{22} \det \begin{pmatrix} a_{33} & a_{34} & \cdots & a_{3n} \\ 0 & a_{43} & \cdots & a_{4n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix}$$

$$= \ldots = \Pi_{i=1}^{n} a_{ii} = \mathrm{tr}(A)$$

- $\det(A) \neq 0 \Leftrightarrow (\forall i \in \{1, \ldots, n\} : a_{ii} \neq 0)$ (does this look familiar?)
- Triangular A is invertible if and only if $\det(A) \neq 0$!
- This holds also for non-triangular A, let's see why

Theorem (Determinant and Elementary Operations)

Let $A \in \mathbb{R}^{n \times n}$ and $\tilde{A}$ the resulting matrix for the respective EO. Then,

(i) *for (E1) (interchange of two rows), we have* $\det(\tilde{A}) = -\det(A)$,

(ii) *for (E2) (row multiplication with a scalar $\lambda \neq 0$),* $\det(\tilde{A}) = \lambda \det(A)$,

(iii) *for (E3) (addition of multiple of row to another row),* $\det(\tilde{A}) = \det(A)$, *i.e. (E3) does not change the determinant.*

- EOs do not affect the property "$\det \neq 0$"!
  $\Rightarrow$ $\det(A) \neq 0 \Leftrightarrow (\det(\tilde{A}) \neq 0$ for the triangularized matrix $\tilde{A})$

$\Rightarrow$ Identity transformation of A exists if and only if $\det(A) \neq 0$

- Indeed: **we can invert $\mathbf{A}$ if and only if** $\det(\mathbf{A}) \neq \mathbf{0}$**!**
- Square equation systems $Ax = b$: unique solution $x^* = A^{-1}b$ if and only if $\det(A) \neq 0$

- Product: $\det(AB) = \det(A)\det(B)$ (if A, B are square!)
- Inverse Matrix: $\det(A^{-1}) = 1/\det(A)$
- Scalar Product: If $A \in \mathbb{R}^{n \times n}$, then $\det(\lambda A) = \lambda^n \det(A)$ (follows from Product rule with $\lambda A = \text{diag}\{\lambda, \ldots, \lambda\}A$)

Conclusion: Square matrix is invertible if and only if it has non-zero determinant! In a Square System, we (almost) always use the determinant invertability criterion!

- Why another criterion for characterizing the solution set? Not all systems are square!
- Recall: a solution exists if we can "reach" b using $f_A(x) = Ax$
- Example:

$$x_1 + 2x_2 = 2$$
$$x_1 - x_2 = 0$$

Matrix form

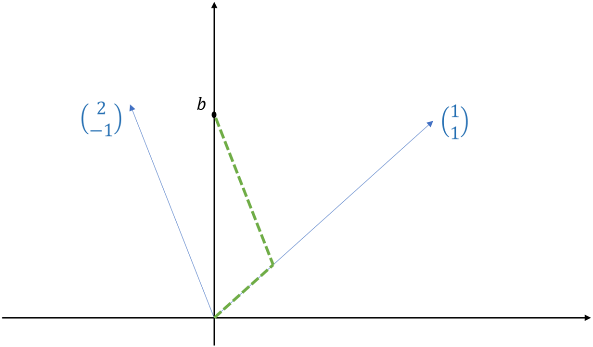
$$Ax = \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} x_1 + \begin{pmatrix} 2 \\ -1 \end{pmatrix} x_2 = \begin{pmatrix} 2 \\ 0 \end{pmatrix} = b$$

Column space criterion: A solution exists if and only if

$$b \in Span\left(\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \end{pmatrix}\right\}\right) =: Co(A)$$

Column space: spanned by the columns of A; x's as LC multipliers!

Column space criterion: “reaching” the point b



Note: condition for *existence*, not *uniqueness*!

- $b \notin f_A[\mathbb{R}^n]$: no solution = information conflict
  - Why "information conflict"? Rows provide contradicting info, e.g.

$$\begin{array}{rcrcl} x_1 & + & x_2 & = & 2 \\ 2x_1 & + & 2x_2 & = & 5 \end{array}$$

  where

$$Co(A) = Span\left(\left\{\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}1\\2\end{pmatrix}\right\}\right) = \left\{\begin{pmatrix}x\\2x\end{pmatrix} : x \in \mathbb{R}\right\} \not\ni \begin{pmatrix}2\\5\end{pmatrix} = b$$

  - Can occur in any (potentially non-square) system with ≥ 2 equations
  - Can be more complex/non-obvious (especially for larger systems)
- Number of LI columns of A = dimension of column space
  - LI columns of A are basis of $Co(A)$
  - The more dimensions $Co(A)$ has, the more likely $b \in Co(A)$

- Column rank: Number of LI columns of A
- Row rank: Number of LI rows of A
- Column rank is crucial for solution existence
- Theorem: Column rank = row rank =: Rank, denoted $\text{rk}(A)$ or $\text{rk } A$
- Corollary: rank bound $\text{rk } A \leq \min\{m, n\}$ for $n \times m$ matrix A
- $n \times m$ matrix A: full row (column) rank if $\text{rk } A = n$ ($\text{rk } A = m$)
- If $m = n = \text{rk } A$, then A has full rank
- **The rank is unchanged by elementary operations (E1) to (E3)!**

Proposition (Elementary Operations and the Set of Solutions)

Consider a system $Ax = b$ of linear equations. Then, for an elementary operation characterized $\tilde{A} = EA$ with operation matrix E, the system $\tilde{A}x = \tilde{b}$ with $\tilde{b} = Eb$ is equivalent to $Ax = b$ in terms of the solutions x.

In other words, EOs do not change the set of solutions!

Theorem (Rank Condition)

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then, the system $Ax = b$ has a unique solution if and only if $b \in Co(A)$ and $rk(A) = n$.

- m equations in n unknowns (!! $m \times n$-matrices from now on !!)
- Intuition rk $A = n$: one LI column for every direction into which b expands $\to$ unique LC of columns

- Definitions:
  - Under-identified system: $m < n$, strictly less equations than unknowns
  - Over-identified system: $m > n$, strictly more equations than unknowns
- Under-identified system: $\text{rk } A \leq m < n \Rightarrow$ no unique solution
  - Intuition: Column space interpretation; e.g.

$$\lambda_1 \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + \lambda_2 \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} + \lambda_3 \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

  - More LC coefficients than dimensions in $b \rightarrow$ no unique LC!
  - "Information deficit" in pinning down n-dim. LC

- Over-identified system: $m > n$: $A = \begin{pmatrix} A_{n\times n} \\ A_{(m-n)\times n} \end{pmatrix}$
  - Recall: we can bring A to upper triangular form with EOs E:

$$\tilde{A} = \begin{pmatrix} A_T \\ \mathbf{0}_{(m-n)\times n} \end{pmatrix} = EA$$

  - $\tilde{A}x = Eb = \tilde{b}$ features the **same solutions!** When are there any? E.g.

$$\begin{pmatrix} 2 & 3 \\ 0 & 4 \\ 0 & 0 \end{pmatrix} x = \begin{pmatrix} 2 \\ -1 \\ c \end{pmatrix} \quad \rightarrow \quad c \stackrel{!}{=} 0$$

  - *No* solution if $\tilde{b}$ has non-zero entries below index n
  - Otherwise: omit "$0 = 0$" rows ("delete redundant information") and investigate square system

- Under-identified systems: insufficient information to pin down a unique solution
  - $\Rightarrow$ Usually not interesting
- Over-identified systems
  - Include contradicting statements, or
  - are actually a square system with added redundant information
  - In latter case: triangularize system and solve the square remainder (example)
  - $\Rightarrow$ **Any system with a unique solution is or can be reduced to a square system**
- Square system: rk $A = n \Leftrightarrow Co(A) = \mathbb{R}^n$
  - $b \in Co(A)$ guaranteed $\rightarrow$ no information conflict
  - Unique solution exists if and only if rk $A = n$ (cf. rank condition)

- As a corollary of what we did so far, we get this powerful result:

Corollary (Invertability, Rank and Determinant of Square Matrices)

$A \in \mathbb{R}^{n \times n}$ (square matrix). Then, the following statements are equivalent:

(i) *A is invertible,*

(ii) $\det(A) \neq 0$,

(iii) *$rk\ A = n$,*

(iv) *for any $b \in \mathbb{R}^n$, $Ax = b$ has a unique solution for x,*

(v) *Any triangular matrix $\tilde{A}$ obtained from applying EOs to A has only non-zero diagonal entries.*

Theorem (Testing Linear Independence)

A equivalent condition for the set of vectors $B = \{b_1, b_2, ..., b_k\}$, $b_j \in \mathbb{R}^n$, *to be linearly independent is that*

$$\sum_{j=1}^{k} \lambda_j b_j = \mathbf{0} \Rightarrow (\forall j \in \{1, ..., k\} : \lambda_j = 0).$$

- Ch. 1 promised a LI check using matrices
- $\sum_{j=1}^{k} \lambda_j b_j = \mathbf{0}$ is an equation system (LC of columns)!
- Matrix form: $B_{mat}\lambda = \mathbf{0}$ with $\lambda = (\lambda_1, \ldots, \lambda_k)'$, $B_{mat} = (b_1, b_2, ..., b_k)$
- We need (i) a unique solution (ii) equal to $\lambda = \mathbf{0}$

Theorem (Testing Linear Independence)

A equivalent condition for the set of vectors $B = \{b_1, b_2, ..., b_k\}$, $b_j \in \mathbb{R}^n$, to be linearly independent is that

$$\sum_{j=1}^{k} \lambda_j b_j = \mathbf{0} \Rightarrow (\forall j \in \{1, ..., k\} : \lambda_j = 0).$$

- $k > n$: solution can not be unique (more vectors than dimensions)
- $k = n$: B_{mat} is square $\rightarrow$ apply determinant criterion
  - If $\det(B_{mat}) \neq 0$, unique solution $\lambda = B_{mat}^{-1}\mathbf{0} = \mathbf{0}$
- $k < n$: Bring B_{mat} to general upper triangular form
  - Equivalent system: $B_{mat}\lambda = \mathbf{0} \Leftrightarrow EB_{mat}\lambda = E\mathbf{0} = \mathbf{0}$
  - Drop zero rows $\rightarrow$ square system, apply determinant criterion
  - Example? Self-study exercises of Ch. 2!

- Eigenvalues and definiteness are linked to inversion but also important in their own right
- **Only square matrices have eigenvalues and are definite!**
- Def. Eigenvalue $\lambda \in \mathbb{R}$ of $A \in \mathbb{R}^{n \times n}$: $\exists x \in \mathbb{R}^n \backslash \{0\} : Ax = \lambda x$
  - $x \in \mathbb{R}^n \backslash \{0\}$ with $Ax = \lambda x$ is called an eigenvector of the eigenvalue λ
  - Eigenvectors are not unique: $Ax = \lambda x \Rightarrow A(cx) = \lambda(cx) \; \forall c \in \mathbb{R}$
  - Eigenspace of λ: $Span(\{x \in \mathbb{R}^n : Ax = \lambda x\})$
- How to find eigenvalues?

$$Ax = \lambda x = \lambda \cdot (\mathbf{I_n} x) \quad \Leftrightarrow \quad \mathbf{0} = Ax - \lambda \mathbf{I_n} x = (A - \lambda \mathbf{I_n})x.$$

$\Rightarrow$ Eigenvectors solve this equation system!

- Thus far: interested in unique solutions and invertible matrices
- Here: candidate unique solution $x = (A - \lambda \mathbf{I_n})^{-1}\mathbf{0} = \mathbf{0}$ is *not* an eigenvector!

$\rightarrow$ Interested in scenarios where $A - \lambda \mathbf{I_n}$ is *not* invertible:

$$\mathcal{P}(\lambda) = \det(A - \lambda \mathbf{I_n}) = 0$$

$\rightarrow$ Infinitely many solutions/eigenvectors ($\neq \mathbf{0}$), and λ is eigenvalue!

- $\mathcal{P}(\lambda)$: characteristic polynomial of A, we want to know its roots:

$$Eigenvalues(A) = \{\lambda \in \mathbb{R} : \mathcal{P}(\lambda) = 0\}$$

- Example: Self-study exercises, online exercises

Proposition (Eigenvalues and Invertability)

Let $A \in \mathbb{R}^{n\times n}$. Then, A is invertable if and only if all eigenvalues of A are non-zero.

Proof. A is invertible if and only if $0 \neq \det(A) = \det(A - 0 \cdot \mathbf{I_n})$, which is the case if and only if 0 is not an eigenvalue of A. □

- If you know the eigenvalues, you can skip the det-check
- If you know the determinant, you know whether 0 is an eigenvalue

Definition (Definiteness of a Matrix)

A **symmetric** *square matrix* $A \in \mathbb{R}^{n \times n}$ *is called*

- *positive semi-definite if* $\forall x \in \mathbb{R}^n : x'Ax \geq 0$
- *negative semi-definite if* $\forall x \in \mathbb{R}^n : x'Ax \leq 0$
- *positive definite if* $\forall x \in \mathbb{R}^n \backslash \{\mathbf{0}\} : x'Ax > 0$
- *negative definite if* $\forall x \in \mathbb{R}^n \backslash \{\mathbf{0}\} : x'Ax < 0$

Otherwise, it is called indefinite.

- Why exclude the zero vector for definiteness?
- Intuition: “smaller/greater than zero” equivalent for matrices
- Fundamentally important for optimization: sign of second derivative

Proposition (Definiteness and Eigenvalues)

A symmetric square matrix $A \in \mathbb{R}^{n \times n}$ is

(i) *positive (negative) definite if and only if all eigenvalues of A are strictly positive (negative).*

(ii) *positive (negative) semi-definite if and only if all eigenvalues of A are strictly non-negative (non-positive).*

Corollary (Definiteness and Invertability)

If $A \in \mathbb{R}^{n \times n}$ is symmetric and positive definite or negative definite, it is invertible.

Positive/negative definiteness is sufficient for invertability!

Context: OLS

- Earlier: if there is an identity transformation of A using EOs, i.e. $EA = \mathbf{I_n}$, then $A^{-1} = E = E\mathbf{I_n}$ ("Gauß-Jordan method")
  - This means that "$EA = \mathbf{I_n}$ for an EO matrix $E \Rightarrow A$ invertible"
  - Now: is it also true that "A invertible $\Rightarrow EA = \mathbf{I_n}$ for an EO matrix E"?
  - *That is, does Gauß-Jordan always find the inverse?*
- Yes, and you know the arguments already!

- Gauß-Jordan Validity:
  - If $A \in \mathbb{R}^{n\times n}$ is invertible, then rk $A = n$
  - Because A is square, we can triangularize it to $\tilde{A} = EA$
  - The rank is preserved under EOs, so that rk $\tilde{A} = n$
  - Thus, $\tilde{A}$ has only non-zero diagonal elements (cf. summary theorem)

$$\tilde{A} = \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} \\ 0 & \tilde{a}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \tilde{a}_{2,n-1} \\ 0 & \cdots & 0 & \tilde{a}_{nn} \end{pmatrix} \quad \rightarrow \quad \hat{A} = \begin{pmatrix} 1 & \hat{a}_{12} & \cdots & \hat{a}_{1n} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{a}_{2,n-1} \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

$\rightarrow$ eliminate the upper triangle using the ones iteratively (right to left) $\checkmark$

This gives ...

Theorem (Gauß-Jordan Algorithm Validity)

Suppose that $A \in \mathbb{R}^{n\times n}$ is an invertible matrix. Then, we can apply elementary operations $E_1, \dots, E_k$ in ascending order of the index to A to arrive at the identity matrix $\mathbf{I_n}$, and the inverse can be determined as $A^{-1} = E_k \cdot \ldots \cdot E_1$.

Example: Define

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

Can you invert A? If so, what is its inverse?

Proposition (Inverse of a 2×2 Matrix)

Consider the matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *where* $ad \neq bc$. *Then,*

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Proof: Self-study exercise (can use Gauß-Jordan, easier: verify $A^{-1}A = \mathbf{I_n}$).