```
>> Lecture 7
>>
>>              -- Matrix Computation
>>
```

Vectors

- Let $\mathbb{R}$ be the set of all real numbers.
- $\mathbb{R}^n$ denotes the vector space of all m-by-1 column vectors:

$$u = (u_i) = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}. \tag{1}$$

  - You can simply use the colon (:) operator to reshape any array in a column major, say $u(:)$.
- Similarly, the row vector v is

$$v = (v_i) = \begin{bmatrix} v_1 \cdots v_n \end{bmatrix}. \tag{2}$$

- We consider column vectors unless stated.

Matrices

- $\mathbf{M}_{m\times n}(\mathbb{R})$ denotes the vector space of all m-by-n real matrices, for example,

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}.$$

- Complex vectors/matrices[1] follow similar definitions and operations introduced later, simply with some care.

[1] Matlab treats a complex number as a single value.

Transposition

```
1  >> A = [1 i];
2  >> A' % Hermitian operator; see any textbook for ...
       linear algebra
3
4  ans =
5
6     1.0000 + 0.0000i
7     0.0000 - 1.0000i
8
9  >> A.' % transposition of A
10
11 ans =
12
13    1.0000 + 0.0000i
14    0.0000 + 1.0000i
```

Arithmetic Operations

- Let a_{ij} and b_{ij} be the elements of the matrices A and $B \in \mathbf{M}^{m\times n}(\mathbb{R})$ for $1 \leq i \leq m$ and $1 \leq j \leq n$.
- Then $C = A \pm B$ can be calculated by $c_{ij} = a_{ij} \pm b_{ij}$. (Try.)

Inner Product[2]

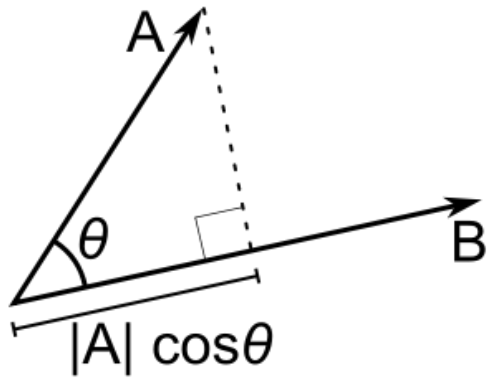
- Let $u, v \in \mathbb{R}^m$.
- Then the inner product, denoted by $u \cdot v$, is calculated by

$$u \cdot v = u'v = [u_1 \cdots u_m] \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}.$$

```
1 clear; clc;
2
3 u = [1; 2; 3];
4 v = [4; 5; 6];
5 u' * v     % normal way; orientation is important
6 dot(u, v) % using the built-in function
```

[2] Akaa dot product and scalar product.

- Inner product is also called projection for emphasizing its geometric significance.



- Recall that we know

$$u \cdot v = 0$$

if and only if these two are orthogonal to each other, denoted by

$$u \perp v.$$

Generalization of Inner Product

- Let $x \in \mathbb{R}$, $f(x)$ and $g(x)$ be real-valued functions.
- In particular, assume that $g(x)$ is a basis function.[3]
- Then we can define the inner product of f and g on $[a, b]$ by

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx.$$

[3] See https://en.wikipedia.org/wiki/Basis_function, https://en.wikipedia.org/wiki/Eigenfunction, and https://en.wikipedia.org/wiki/Approximation_theory.

- For example, Fourier transform is widely used in engineering and science.
  - Fourier integral[4] is defined as

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

    where $f(t)$ is a square-integrable function.
  - The Fast Fourier transform (FFT) algorithm computes the discrete Fourier transform (DFT) in $O(n \log n)$ time.[5,6]

[4] See `https://en.wikipedia.org/wiki/Fourier_transform`.
[5] Cooley and Tukey (1965).
[6] See `https://en.wikipedia.org/wiki/Fast_Fourier_transform`.

Matrix Multiplication

- Let $A \in \mathbf{M}_{m\times q}(\mathbb{R})$ and $B \in \mathbf{M}_{q\times n}(\mathbb{R})$.
- Then $C = AB$ is given by

$$c_{ij} = \sum_{k=1}^{q} a_{ik} \times b_{kj}. \tag{3}$$

- For example,

$$\underset{4\times 2 \text{ matrix}}{\begin{bmatrix} a_{11} & a_{12} \\ \cdot & \cdot \\ a_{31} & a_{32} \\ \cdot & \cdot \end{bmatrix}} \underset{2\times 3 \text{ matrix}}{\begin{bmatrix} \cdot & b_{12} & b_{13} \\ \cdot & b_{22} & b_{23} \end{bmatrix}} = \underset{4\times 3 \text{ matrix}}{\begin{bmatrix} \cdot & x_{12} & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & x_{33} \\ \cdot & \cdot & \cdot \end{bmatrix}}$$

Example

```
1  clear; clc;
2
3  A = randi(10, 5, 4); % 5-by-4
4  B = randi(10, 4, 3); % 4-by-3
5  C = zeros(size(A, 1), size(B, 2));
6  for i = 1 :  size(A, 1)
7      for j = 1 : size(B, 2)
8          for k = 1 : size(A, 2)
9              C(i, j) = C(i, j) + A(i, k) * B(k, j);
10         end
11     end
12 end
13 C % display C
```

- Time complexity: $O(n^3)$.
- Strassen (1969): $O(n^{log_2 7})$.

Matrix Exponentiation

- Raising a matrix to a power is equivalent to repeatedly multiplying the matrix by itself.
  - For example, $A^2 = AA$.
- The matrix exponential[7] is a matrix function on square matrices analogous to the ordinary exponential function, more explicitly,

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}.$$

- However, it is not allowed to perform A^B.

[7] See matrix exponentials and Pauli matrices.

Determinants

- Consider the matrix
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$
- Then $\det(A) = ad - bc$ is called the determinant of A.
  - The method of determinant calculation in high school is a wrong way but produces correct answers for all 3×3 matrices.
- Let's try the minor expansion formula for det(A).[8]

[8] See `http://en.wikipedia.org/wiki/Determinant`.

Recursive Algorithm for Minor Expansion Formula

```
1  function y = myDet(A)
2
3      [r, ~] = size(A);
4
5      if r == 1
6          y = A;
7      elseif r == 2
8          y = A(1, 1) * A(2, 2) - A(1, 2) * A(2, 1);
9      else
10         y = 0;
11         for i = 1 : r
12             B = A(2 : r, [1 : i - 1, i + 1 : r]);
13             cofactor = (-1) ^ (i + 1) * myDet(B);
14             y = y + A(1, i) * cofactor;
15         end
16     end
17 end
```

- It needs $n!$ terms in the sum of products, so this algorithm runs in $O(n!)$ time!
- Use **det** for determinants, which can be done in $O(n^3)$ time by using LU decomposition or alike.[9]

[9] See `https://en.wikipedia.org/wiki/LU_decomposition`. Moreover, various decompositions are used to implement efficient matrix algorithms in numerical analysis. See `https://en.wikipedia.org/wiki/Matrix_decomposition`.

Linear Systems (Transformation/Mapping)[10]

- A linear system is a mathematical model of a system based on linear operators satisfying the property of superposition.
  - For simplicity, $Ax = y$ for any input x associated with the output y.
  - Then A is a linear operator if and only if

$$A(ax_1 + bx_2) = aAx_1 + bAx_2 = ay_1 + by_2$$

    for $a, b \in \mathbb{R}$.
  - For example, $\frac{d(x^2+3x)}{dx} = \frac{dx^2}{dx} + 3\frac{dx}{dx} = 2x + 3$.
- Linear systems typically exhibit features and properties that are much simpler than the nonlinear case.
  - What about nonlinear cases?

[10] See `https://en.wikipedia.org/wiki/Linear_system`.

First-Order Approximation: Local Linearization

- Let $f(x)$ be any nonlinear function.
- Assume that $f(x)$ is infinitely differentiable at x_0.
- By Taylor's expansion[11], we have

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + O\left((x - x_0)^2\right),$$

where $O\left((x - x_0)^2\right)$ is the collection of higher-order terms, which can be neglected as $x - x_0 \to 0$.

- Then we have a first-order approximation

$$f(x) \approx f'(x_0)x + k,$$

with $k = f(x_0) - x_0 f'(x_0)$, a constant.

[11] See `https://en.wikipedia.org/wiki/Taylor_series`.

Two Observations

- We barely feel like the curvature of the ground; however, we look at Earth on the moon and agree that Earth is a sphere.
- Newton's kinetic energy is a low-speed approximation (classical limit) to Einstein's total energy.
  - Let m be the rest mass and v be the velocity relative to the inertial coordinate.
  - The resulting total energy is

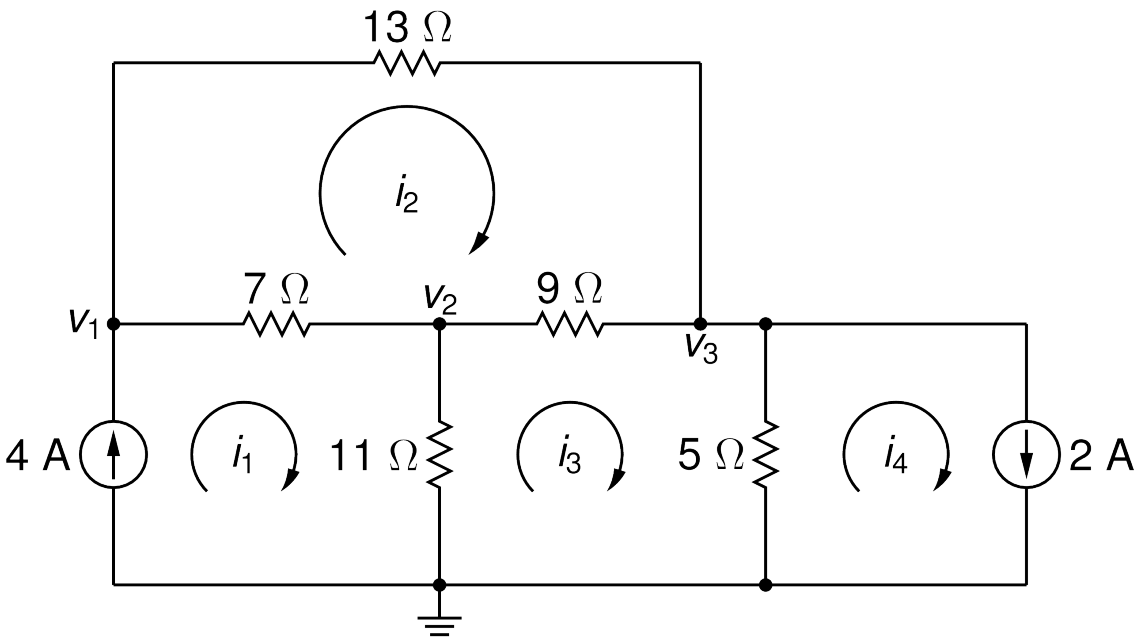
$$E = \frac{mc^2}{\sqrt{1-(v/c)^2}}.$$

  - By applying the first-order approximation,

$$E \approx mc^2 + \frac{1}{2}mv^2.$$

Example: Kirchhoff's Laws[12]

- The algebraic sum of currents in a network of conductors meeting at a point is zero.
- The directed sum of the potential differences (voltages) around any closed loop is zero.



[12] See https://en.wikipedia.org/wiki/Kirchhoff's_circuit_laws.

General Form of Linear Equations[13]

- Let n be the number of unknowns and m be the number of constraints.
- A general system of m linear equations with n unknowns is

$$\begin{cases} a_{11}x_1 & +a_{12}x_2 & \cdots & +a_{1n}x_n & = & y_1 \\ a_{21}x_1 & +a_{22}x_2 & \cdots & +a_{2n}x_n & = & y_2 \\ \vdots & \vdots & \ddots & \vdots & = & \vdots \\ a_{m1}x_1 & +a_{m2}x_2 & \cdots & +a_{mn}x_n & = & y_m \end{cases}$$

where $x_1, \ldots, x_n$ are unknowns, $a_{11}, \ldots, a_{mn}$ are the coefficients of the system, and $y_1, \ldots, y_m$ are the constant terms.

[13] See `https://en.wikipedia.org/wiki/System_of_linear_equations`.

Matrix Equation

- Hence we can rewrite the aforesaid equations as follows:

$$Ax = y.$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \text{ and } y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

- Finally, x can be done by $x = A^{-1}y$, where A^{-1} is called the inverse of A.

Inverse Matrices[14]

- For simplicity, let $A \in \mathbf{M}_{n\times n}(\mathbb{R})$ and $x, y \in \mathbb{R}^n$.
- Then A is called invertible if there exists $B \in \mathbf{M}_{n\times n}(\mathbb{R})$ such that

$$AB = BA = I_n,$$

  where I_n denotes a $n \times n$ identity matrix.
  - We use A^{-1} to denote the inverse of A.
  - You can use **eye**(n) to generate an identity matrix I_n.
- Use **inv**(A) to calculate the inverse of A.

[14] See `https://en.wikipedia.org/wiki/Invertible_matrix#The_invertible_matrix_theorem`.

- However, **inv**(A) may return a weird result even if A is ill-conditioned, indicates how much the output value of the function can change for a small change in the input argument.[15]
- For example, calculate the inverse of the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

  - Recall the Cramer's rule[16]: A is invertible iff $\mathbf{det}(A) \neq 0$. (Try.)
- If these constraints cannot be eliminated by row reduction, they are linearly independent.

[15] You may refer to the condition number of a function with respect to an argument. Also try **rcond**.

[16] See `https://en.wikipedia.org/wiki/Cramer's_rule`.

Linear Independence

- Let $K = \{a_1, a_2, \ldots, a_n\}$ for each $a_i \in \mathbb{R}^m$.
- Now consider this linear superposition

$$x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = 0,$$

  where $x_1, x_2, \ldots, x_n \in \mathbb{R}$ are the weights.
- Then K is linearly independent iff

$$x_1 = x_2 = \cdots = x_n = 0.$$

Example: $\mathbb{R}^3$

- Let
$$K_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$
- It is clear that K_1 is linearly independent.
- Moreover, you can represent all vectors in $\mathbb{R}^3$ if you collect all linear superpositions from K_1.
- We call this new set a span of K_1, denoted by $\text{Span}(K_1)$.[17]
- Clearly, $\text{Span}(K_1) = \mathbb{R}^3$.

[17] See `https://en.wikipedia.org/wiki/Linear_span`.

- Now let

$$K_2 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\}.$$

- Then K_2 is not a linearly independent set. (Why?)
- If you take one or more vectors out of K_2, then K_2 becomes linearly independent.

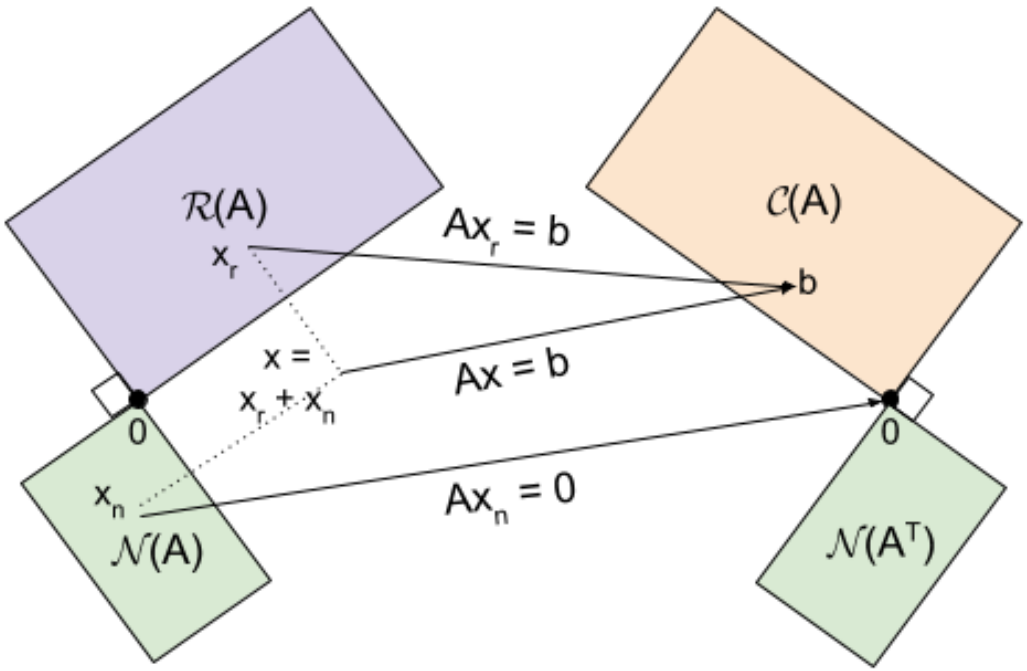
Basis of Vector Space & Its Dimension[18]

- However, you can take only one vector out of K_2 if you want to represent all vectors in $\mathbb{R}^3$. (Why?)
  - The dimension of $\mathbb{R}^3$ is exactly the size (element number) of K_2.
- We say that the basis of $\mathbb{R}^n$ is a maximally linearly independent set of size n.
- Note that the basis of $\mathbb{R}^3$ is not unique.
  - For example, K_1 could be also a basis of $\mathbb{R}^3$.

[18] See https://en.wikipedia.org/wiki/Basis_(linear_algebra), https://en.wikipedia.org/wiki/Vector_space, and https://en.wikipedia.org/wiki/Dimension_(vector_space).

Linear Transformation (Revisited)[19]

Matrix A converts n-tuples into m-tuples $\mathbb{R}^n \rightarrow \mathbb{R}^m$.
That is, linear transformation T_A is a map between rows and columns



Fundamental Subspaces

$\mathcal{C}(A)$: Column space (image)
$\mathcal{R}(A)$: Row space (coimage)
$\mathcal{N}(A)$: Null space (kernel)
$\mathcal{N}(A^T)$: Left null space (cokernel)

Identities

$\dim(\mathcal{C}) \equiv \text{rank}(A)$
$\dim(\mathcal{N}) \equiv \text{nullity}(A)$

Theorems

$\dim(\mathcal{C}) + \dim(\mathcal{N}) = n$
$\dim(\mathcal{R}) = \dim(\mathcal{C})$

[19] See https://en.wikipedia.org/wiki/Linear_map; also see https://kevinbinz.com/2017/02/20/linear-algebra/.

Example: Vector Projection ($\mathbb{R}^3 \to \mathbb{R}^2$)

- Let $u \in \mathbb{R}^3$ and $v \in \mathbb{R}^2$.
- We consider the projection matrix (operator),

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

  so that $Au = v$.
- For example,

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

Solution Set to System of Linear Equations[20]

- Recall that m is the number of constraints and n is the number of unknowns.
- Now consider the following cases.
- If $m = n$, then there exists a unique solution.
- If $m > n$, then it is called an overdetermined system and there is no solution.
  - Fortunately, we can find a least-squares error solution such that $\| Ax - y \|^2$ is minimal, shown later.
- If $m < n$, then it is called a underdetermined system which has infinitely many solutions.
  - Become an optimization problem?
- For all cases,

$$x = A \setminus y.$$

[20] See `https://www.mathworks.com/help/matlab/ref/mldivide.html`.

Case 1: $m = n$

- For example,

$$\begin{cases} 3x + 2y - z = 1 \\ x - y + 2z = -1 \\ -2x + y - 2z = 0 \end{cases}$$

```
1 >> A = [3 2 -1; 1 -1 2; -2 1 -2];
2 >> b = [1; -1; 0];
3 >> x = A \ b
4
5  1
6  -2
7  -2
```

Case 2: $m > n$

- For example,

$$\begin{cases} 2x - y = 2 \\ x - 2y = -2 \\ x + y = 1 \end{cases}$$

```
>> A = [2 -1; 1 -2; 1 1];
>> b = [2; -2; 1];
>> x = A \ b

 1
 1
```

Case 3: $m < n$

- For example,

$$\begin{cases} x + 2y + 3z = 7 \\ 4x + 5y + 6z = 8 \end{cases}$$

```
>> A = [1 2 3; 4 5 6];
>> b = [7; 8];
>> x = A \ b

-3
0
3.333
```

- Note that this solution is a basic solution, one of infinitely many.
- How to find the directional vector? (Try **cross**.)

Gaussian Elimination Algorithm[21]

- First we consider the linear system is represented as an augmented matrix $[\,A\,|\,y\,]$.
- We then transform A into an upper triangular matrix

$$[\,\bar{A}\,|\,y\,] = \left[\begin{array}{cccc|c} 1 & \bar{a}_{12} & \cdots & \bar{a}_{1n} & \bar{y}_1 \\ 0 & 1 & \cdots & \bar{a}_{2n} & \bar{y}_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \bar{y}_n \end{array}\right].$$

  where $\bar{a}_{ij}$'s and $\bar{y}_i$'s are the resulting values after elementary row operations.
- This matrix is said to be in reduced row echelon form.

[21] See `https://en.wikipedia.org/wiki/Gaussian_elimination`.

- The solution can be done by backward substitution:

$$x_i = \bar{y}_i - \sum_{j=i+1}^{n} \bar{a}_{ij} x_j,$$

  where $i = 1, 2, \cdots, n$.
- Time complexity: $O(n^3)$.

Exercise

```
1  clear; clc;
2
3  A = [3 2 -1; 1 -1 2; -2 1 -2];
4  b = [1; -1; 0];
5  A \ b % check the answer
6
7  for i = 1 : 3
8      for j = i : 3
9          b(j) = b(j) / A(j, i); % why first?
10         A(j, :) = A(j, :) / A(j, i);
11     end
12     for j = i + 1 : 3
13         A(j, :) = A(j, :) - A(i, :);
14         b(j) = b(j) - b(i);
15     end
16 end
17 x = zeros(3, 1);
```

```
18  for i = 3 : -1 : 1
19      x(i) = b(i);
20      for j = i + 1 : 1 : 3
21          x(i) = x(i) - A(i, j) * x(j);
22      end
23  end
24  x
```

Selected Functions of Linear Algebra[22]

- Matrix properties: **norm**, **null**, **orth**, **rank**, **rref**, **trace**, **subspace**.
- Matrix factorizations: **lu**, **chol**, **qr**.

[22] See `https://www.mathworks.com/help/matlab/linear-algebra.html`.

Numerical Example: 2D Laplace's Equation

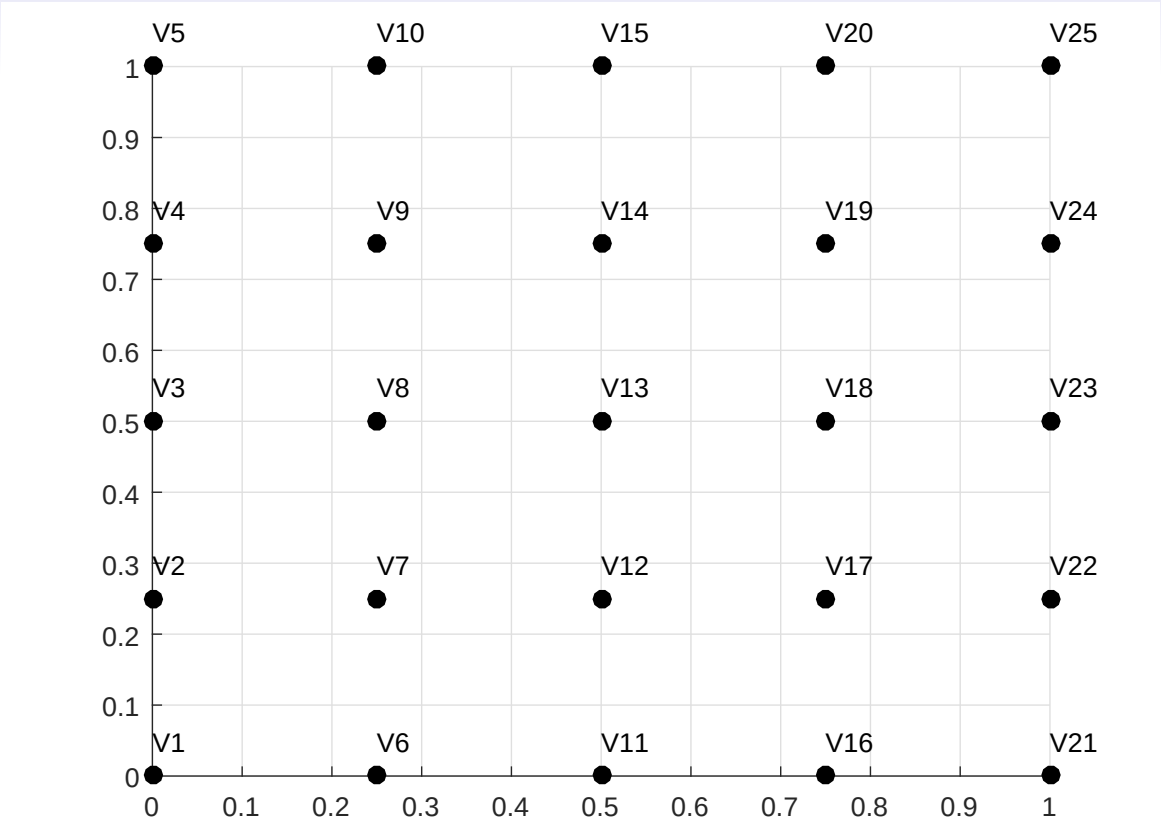
- A partial differential equation (PDE) is a differential equation that contains unknown multivariable functions and their partial derivatives.[23]
- Let $\Phi(x, y)$ be a scalar field on $\mathbb{R}^2$.
- Consider Laplace's equation[24] as follows:

$$\nabla^2 \Phi(x, y) = 0,$$

  where $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplace operator.
- Consider the system shown in the next page.

[23] See `https://en.wikipedia.org/wiki/Partial_differential_equation`.

[24] Pierre-Simon Laplace (1749–1827).



- Consider the boundary condition:
  - $V_1 = V_2 = \cdots = V_4 = 0.$
  - $V_{21} = V_{22} = \cdots = V_{24} = 0.$
  - $V_1 = V_6 = \cdots = V_{16} = 0.$
  - $V_5 = V_{10} = \cdots = V_{25} = 1.$

An Simple Approximation[25]

- As you can see, we partition the region into many subregions by applying a proper mesh generation.
- Then $\Phi(x, y)$ can be approximated by

$$\Phi(x, y) \approx \frac{\Phi(x+h, y) + \Phi(x-h, y) + \Phi(x, y+h) + \Phi(x, y-h)}{4},$$

where h is small enough.

[25] See https://en.wikipedia.org/wiki/Finite_difference_method#Example:_The_Laplace_operator.

Matrix Formation

- By collecting all constraints, we have $Ax = b$ where

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

and

$$b = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}^T.$$

Dimension Reduction by Symmetry

- As you can see, $V_7 = V_{17}$, $V_8 = V_{18}$ and $V_9 = V_{19}$.
- So we can reduce A to A'

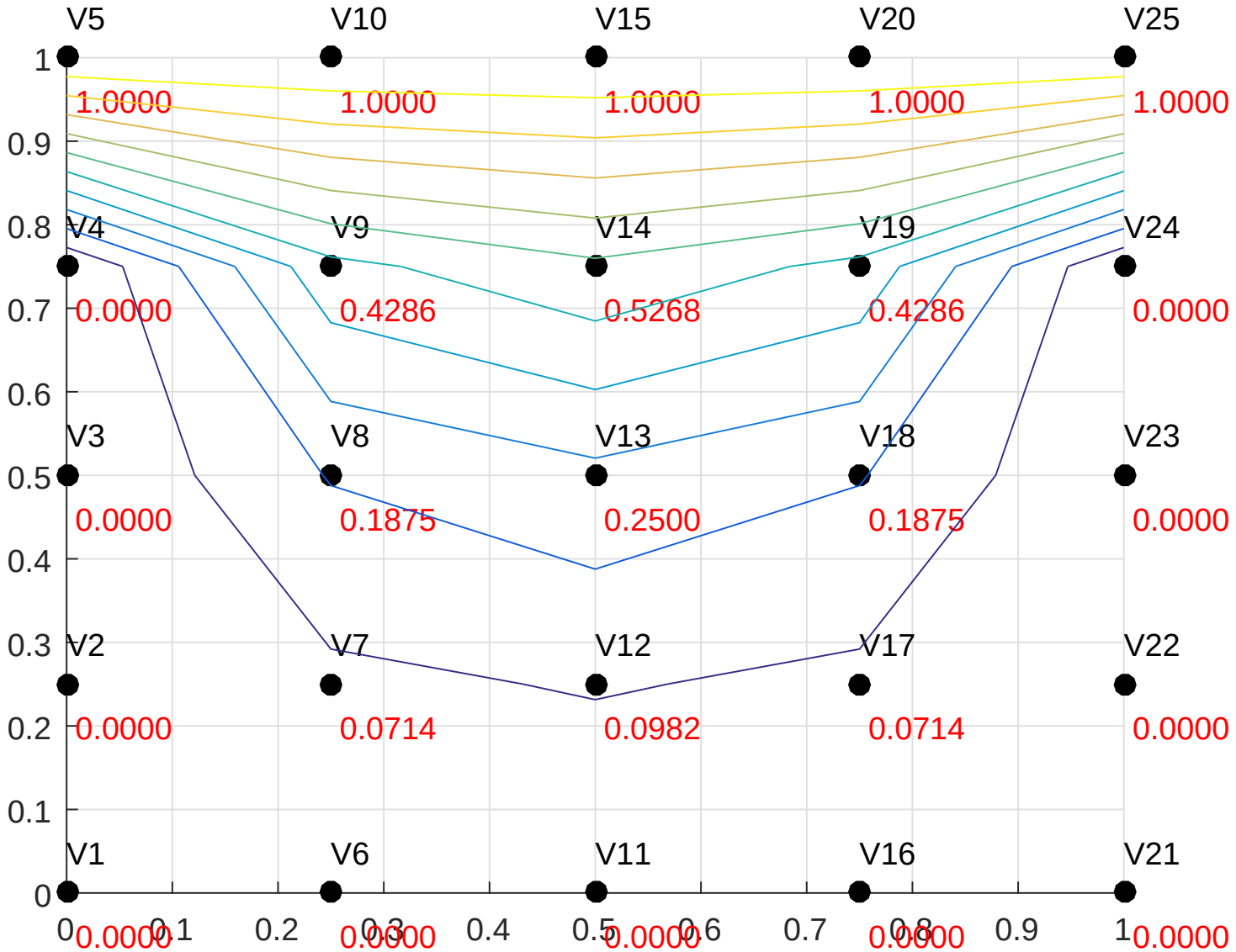
$$A' = \begin{bmatrix} 4 & -1 & 0 & -1 & 0 & 0 \\ -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & -1 & 4 & 0 & 0 & -1 \\ -2 & 0 & 0 & 4 & -1 & 0 \\ 0 & -2 & 0 & -1 & 4 & -1 \\ 0 & 0 & -2 & 0 & -1 & 4 \end{bmatrix}$$

and

$$b' = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}^T.$$

- The dimensions of this problem are cut to 6 from 9.
- This trick helps to alleviate the curse of dimensionality.[26]

[26] See `https://en.wikipedia.org/wiki/Curse_of_dimensionality`.


V5
V10
V15
V20
V25
1.0000
1.0000
1.0000
1.0000
1.0000
V4
V9
V14
V19
V24
0.0000
0.4286
0.5268
0.4286
0.0000
V3
V8
V13
V18
V23
0.0000
0.1875
0.2500
0.1875
0.0000
V2
V7
V12
V17
V22
0.0000
0.0714
0.0982
0.0714
0.0000
V1
V6
V11
V16
V21
0.0000
0.0000
0.0000
0.0000
0.0000
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Remarks

- This is a toy example for numerical methods of PDEs.
- We can use the PDE toolbox for this case. (Try.)
  - You may consider the finite element method (FEM).[27]
  - The mesh generation is also crucial for numerical methods.[28]
  - You can use the Computational Geometry toolbox for triangular mesh.[29]

[27] See https://en.wikipedia.org/wiki/Finite_element_method.

[28] See https://en.wikipedia.org/wiki/Mesh_generation.

[29] See https://www.mathworks.com/help/matlab/computational-geometry.html.

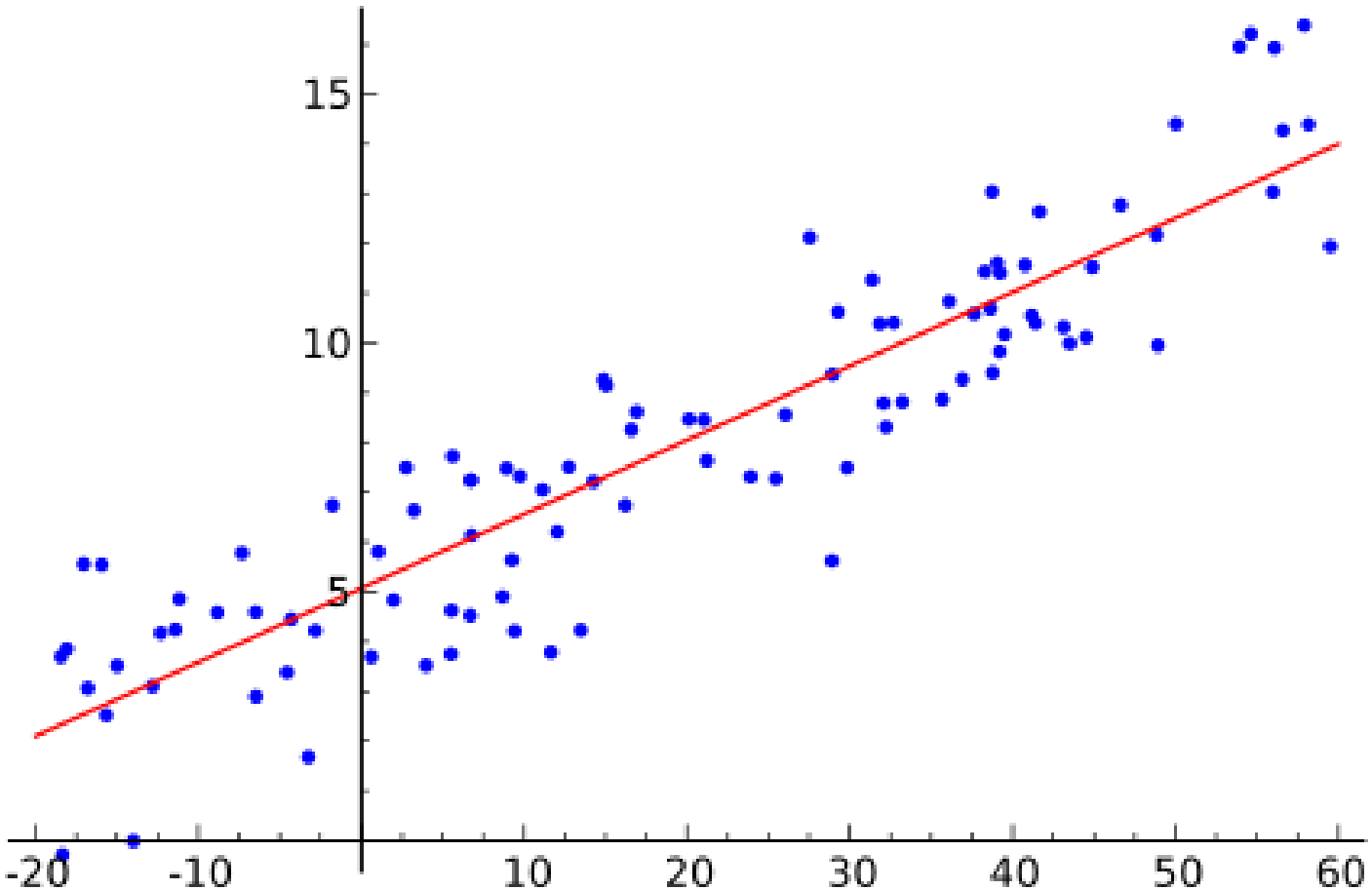
Numerical Example: Method of Least Squares[30]

- The method of least squares is a standard approach to the approximate solution of overdetermined systems ($m > n$).
- Let $\{\hat{y}_i\}_{i=1}^n$ be the observed response values and $\{y_i\}_{i=1}^n$ be the fitted response values.
- Let $\varepsilon_i = \hat{y}_i - y_i$ be the residual for $i = 1, \ldots, n$.
- Then the sum of square residuals estimates associated with the data is given by

$$S = \sum_{i=1}^{n} \epsilon_i^2.$$

- The best fit in the least-squares sense minimizes the sum of squared residuals.

[30] See `https://en.wikipedia.org/wiki/Least_squares`.



https://commons.wikimedia.org/wiki/File:Linear_regression.svg

Linear Least Squares

- The approach is called linear least squares since the assumed function is linear in the parameters to be estimated.
- For example, consider

$$y = ax + b,$$

where a and b are to be determined.

- Then we have $\epsilon_i = (ax_i + b) - \hat{y}_i$ so that

$$S = \sum_{i=1}^{n} ((ax_i + b) - \hat{y}_i)^2 .$$

- Now consider the partial derivatives of S with respective to a and b:

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{n} x_i(y_i - (ax_i + b)) = 0,$$

$$\frac{\partial S}{\partial b} = -2\sum_{i=1}^{n} (y_i - (ax_i + b)) = 0.$$

- We reorganize the above equations as follows:

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i,$$

$$a\sum_{i=1}^{n} x_i + nb = \sum_{i=1}^{n} y_i.$$

- It could be done by using normal equations.[31]

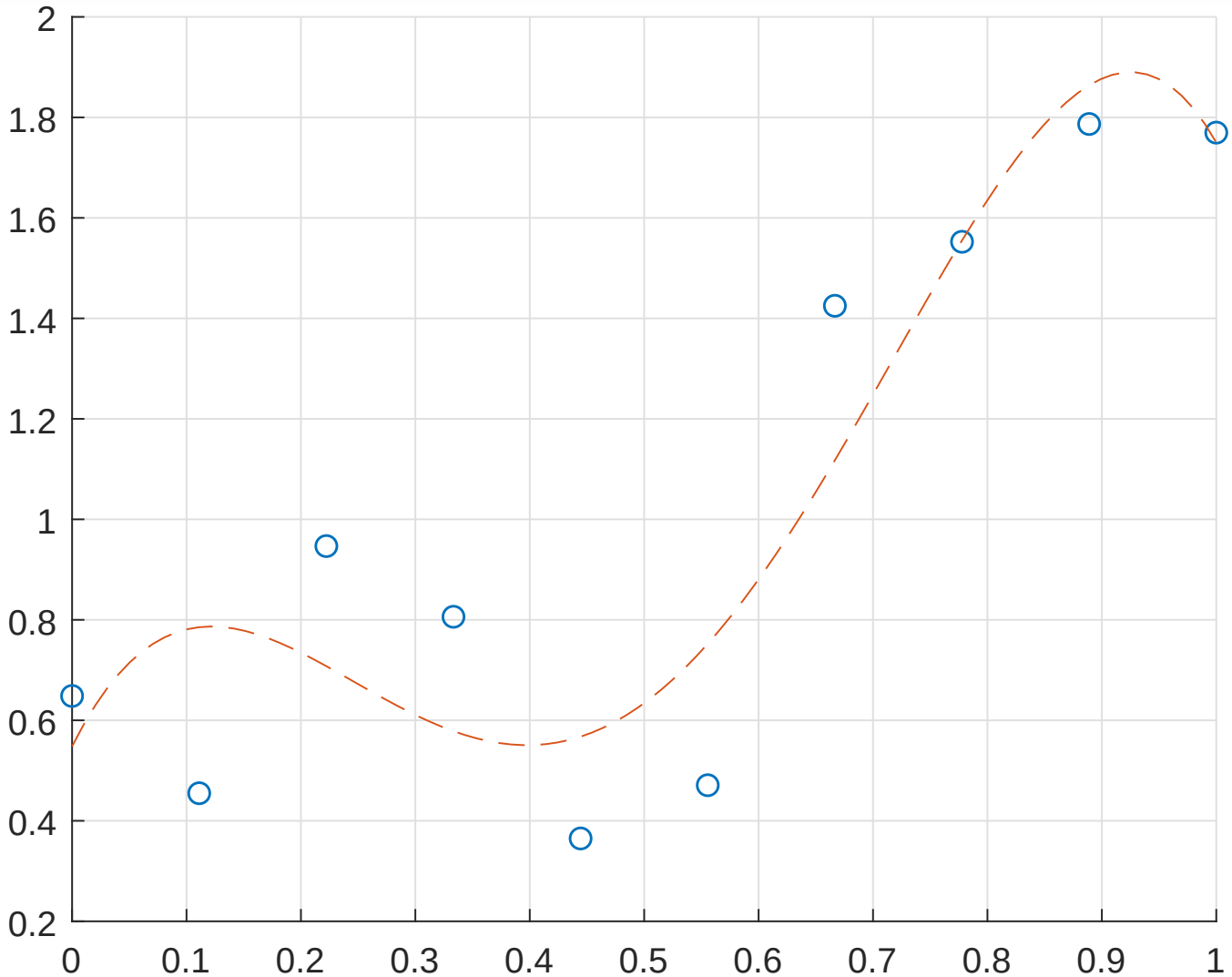
[31] See https://en.wikipedia.org/wiki/Linear_least_squares_(mathematics)#Derivation_of_the_normal_equations.

Example

```
clear; clc; close all;

rng(3); % fix the random seed
N = 10;
x = linspace(0, 1, N); x = x(:);
y = cos(rand(size(x)) * pi / 2) + x .^ 2;
figure; hold on; grid on; plot(x, y, 'o');
degree = 4;

M = @(x, degree) repmat(x, 1, degree + 1);
A = @(mat) bsxfun(@(x, i) x .^ i, mat, ...
                  size(mat, 2) - 1 : -1 : 0);
pp = A(M(x, degree)) \ y % show the coefficients
xq = linspace(min(x), max(x), 100); xq = xq(:);
yq = A(M(xq, degree)) * pp;
plot(xq, yq, '--');
```


2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Polynomial Regression

- **polyfit**(x, y, n) returns the coefficients for a polynomial of degree n that is a best fit for the set of sample data (x, y) (in a least-squares sense).

Example

```
clear; clc; close all;

rng(3);
N = 10;
x = linspace(0, 1, N); x = x(:);
y = cos(rand(size(x)) * pi / 2) + x .^ 2;
figure; hold on; grid on; plot(x, y, 'o');
degree = 4;

p = polyfit(x, y, degree)
xq = linspace(0, 1, 50);
yq = polyval(p, xq);
plot(xq, yq);
```

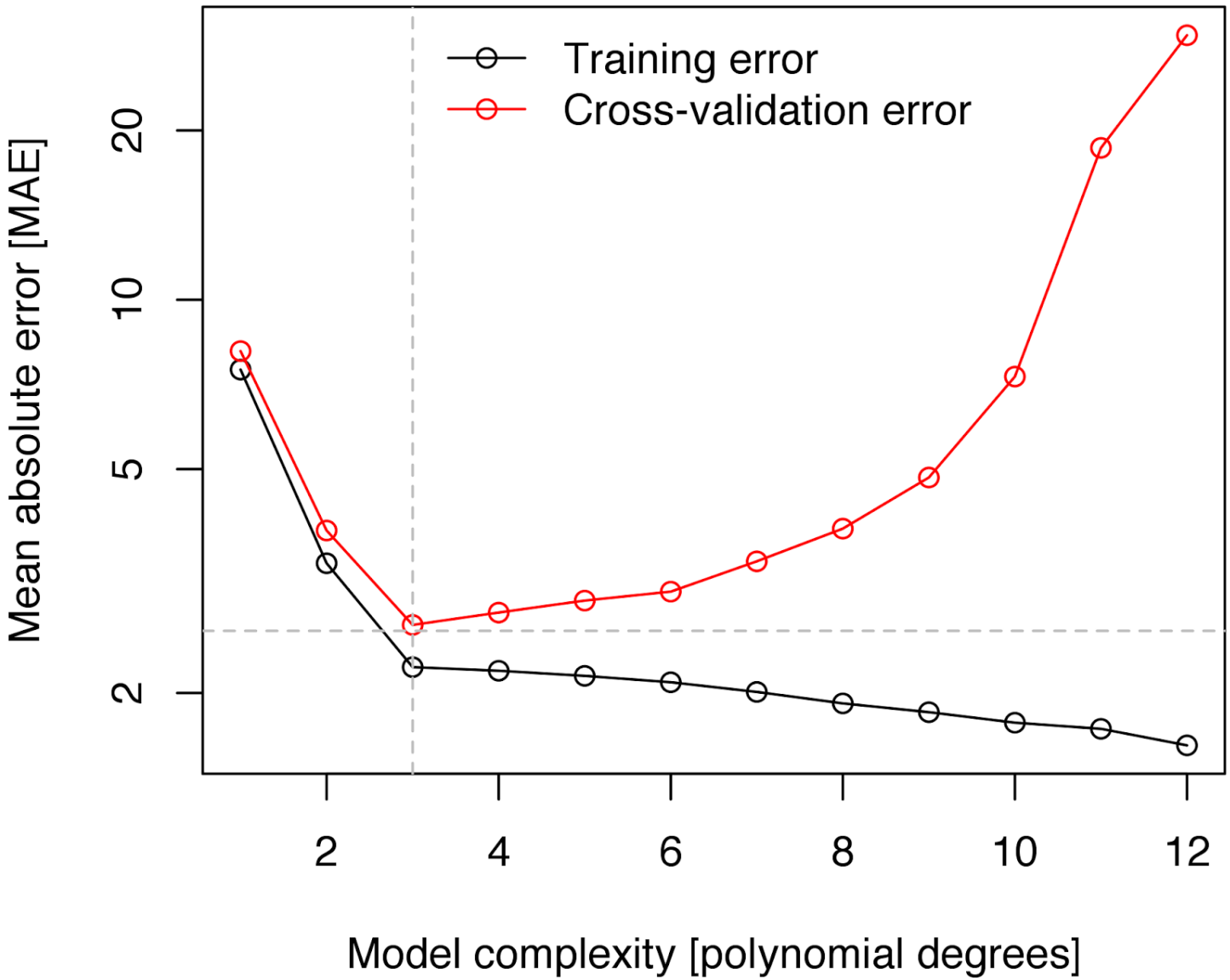
- The result is identical to the figure shown before.

Overfitting

- Overfitting is the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably.
  - In other words, the overfitted model is perfect to in-sample data but not robust in out-of-sample data.
  - For example, Runge's phenomenon.[32]
- Law of parsimony[33] states that simpler solutions are more likely to be correct than complex ones.

[32] See `https://en.wikipedia.org/wiki/Runge's_phenomenon`.

[33] Aka Occam's Razor.


Training error
Cross-validation error
Mean absolute error [MAE]
2
5
10
20
2
4
6
8
10
12
Model complexity [polynomial degrees]

Polynomials

- Let $n \in \mathbb{N} \cup \{0\}$, and $x, a_0, \ldots, a_n \in \mathbb{R}$.
- $f(x)$ is said to be a polynomial with degree n provided that

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0,$$

  where $a_n \neq 0$.
- We often express a polynomial by its coefficient vector $[a_n, a_{n-1}, \ldots, a_0]$.
- In fact, the set of polynomials with coefficients in $\mathbb{R}$ is a vector space over $\mathbb{R}$, denoted by $\mathbb{P}_n$.[34]

[34] See https://en.wikipedia.org/wiki/Examples_of_vector_spaces#Polynomial_vector_spaces.

Arithmetic Operations of Polynomials

- Let p_1 and p_2 be two coefficient vectors of polynomials of the same degree.
- Then we have the following operations:
  - addition and subtraction: $p_1 \pm p_2$.
  - multiplication: $\mathbf{conv}(p_1, p_2)$.[35]
  - division: $[q, r] = \mathbf{deconv}(p_1, p_2)$.[36]
- Use $\mathbf{polyval}(p_1, x)$ to calculate the function values of $p1$ on x.

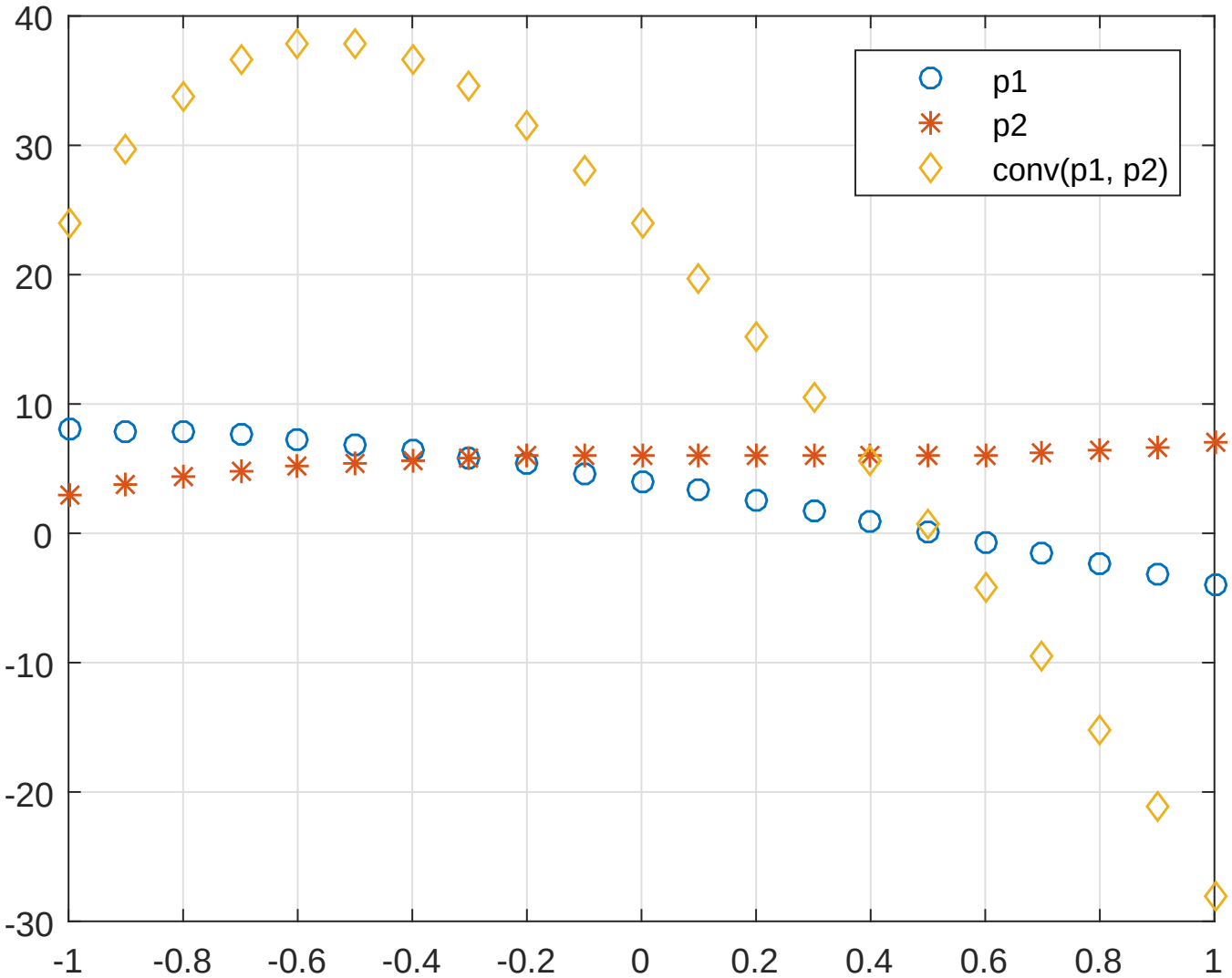
[35] See `http://en.wikipedia.org/wiki/Convolution`.

[36] Equivalently, $v = \text{conv}(u, q) + r$. Also see `http://en.wikipedia.org/wiki/Euclidean_division`.

```
clear; clc;

p1 = [1 -2 -7 4];
p2 = [2 -1 0 6];
p3 = p1 + p2
p4 = p1 - p2
p5 = conv(p1, p2)
[q, r] = deconv(p1, p2)

x = linspace(-1, 1, 20);
plot(x, polyval(p1, x), 'o', x, polyval(p2, x), ...
    '*',  x, polyval(p5, x), 'd');
grid on; legend('p1', 'p2', 'conv(p1, p2)');
```


40
30
20
10
0
-10
-20
-30
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
p1
p2
conv(p1, p2)

Finding Roots of Polynomial

- Use **roots**(p) for all roots of the polynomial p.[37]
- For example,

```
1 clear; clc; close all;
2
3 p = [1, 3, 1, 5, -1];
4 r = roots(p) % find all roots of p
5 polyval(p, r) % why not zeros?
```

[37] See https://en.wikipedia.org/wiki/Jenkins-Traub_algorithm.

Exercise: Internal Rate of Return (IRR)[38]

- Consider two assets.
- For asset A, you are promised to receive the cash flows as follows:

$$C_0 = -100, C_1 = 0, C_2 = 0, C_3 = 120.$$

- For asset B, the cash flows are

$$C_0 = -100, C_1 = 6, C_2 = 6, C_3 = 108.$$

- Which asset is more desirable?

[38] See `https://en.wikipedia.org/wiki/Internal_rate_of_return`.

- Given a collection of pairs (time, cash flow), the IRR is a rate of return when the net present value is zero.
- Explicitly, the IRR can be calculated by solving

$$\sum_{i=0}^{N} \frac{C_i}{(1+r)^i} = 0,$$

  where C_i is the cash flow at time i.
- So the IRR is 6.27% for A and 6.62% for B.

"Time is free, but its priceless.
You can't own it, but you can use it.
You can't keep it, but you can spend it.
Once you've lost it, you can never get it back."
– Harvey MacKay

"No man can achieve success
if he didn't first know the value of time."
– Sunday Adelaja

Integral and Derivative of Polynomials

- **polyder**(p) returns the derivative of the polynomial p.
- **polyint**(p, k) returns a polynomial representing the integral of polynomial p, using a scalar constant of integration k.

```
1 clear; clc;
2
3 p = [4 3 2 1];
4 p_der = polyder(p)
5 p_int = polyint(p, 0) % k = 0
```

Exercise

- Let p be the coefficient vector for any polynomial with degree 3.
- Write a program to calculate the coefficients of its derivative and integration.
- Also, you may write down the matrix representation of differentiation and integration of p.
- Do not use the built-in functions.

```
clear; clc;

p = randi(100, 1, 4)
q1 = [0, p(1 :  end - 1) .* [length(p) - 1 : -1 : 1]]
q2 = [p ./ [length(p)  : -1 : 1], 0]

T1 = [0 0 0 0;
      3 0 0 0;
      0 2 0 0;
      0 0 1 0];
T1 * p'
T2 = [0 1/4   0   0   0;
      0   0 1/3   0   0;
      0   0   0 1/2   0;
      0   0   0   0   1;
      0   0   0   0   0];
T2 * [0 p]'
```

Eigenvalues and Eigenvectors

- Let $A \in \mathbf{M}_{n\times n}(\mathbb{R})$, I be the identity, and $v \in \mathbb{R}^n$ be nontrivial.
- An eigenvalue problem[39] is a system which follows

$$Av = \lambda v.$$

- Then u is an eigenvector associated with the eigenvalue λ by solving $\det(A - \lambda I) = 0$, aka the characteristic polynomial.
  - Use $[\, V, D \,] = \mathbf{eig}(A)$ produces a diagonal matrix D of eigenvalues and a full matrix V whose columns are the corresponding eigenvectors so that $AV = VD$.

[39] See `https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors#Applications`.

Example: Google PageRank Algorithm[40]

- PageRank (PR) is an algorithm used by Google Search to rank web pages in their search engine results.
- PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is.
- The underlying assumption is that more important websites are likely to receive more links from other websites.

[40] Larry Page and Sergey Brin (1998).

Singular Value Decomposition (SVD)

- Let $A \in \mathbf{M}_{m\times n}(\mathbb{R})$, $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$, and $\sigma \in \mathbb{R}$.
- σ is called a singular value associated with the left singular vector u and the right singular vector v for A provided that

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^T.$$

- In matrix form,

$$A = U\Sigma V^T,$$

  where U and V consist of the left and right singular vectors, respectively, and Σ is a diagonal matrix whose diagonal entries are the singular values of A.
- Use $[\,U, S, V\,] = \mathbf{svd}(A)$ for SVD.[41]

[41] See `https://www.mathworks.com/help/matlab/math/singular-values.html`.

Example: Image Compression by Low-Rank Approximation

- We can have an image extremely similar to the original one, but with a smaller image size by keeping the vectors associated with a few number of first large principal components, aka Principal Component Analysis (PCA).[42]
- PCA can be done by **svd**.

[42] See `http://setosa.io/ev/principal-component-analysis/`.