032CM - 2025

PROGRAMMING FOR COMPUTATIONAL CHEMISTRY

Working with NumPy arrays

Gianluca Levi

gianluca.levi@units.it, giale@hi.is

Office: Building C11, 3rd floor, Room 329

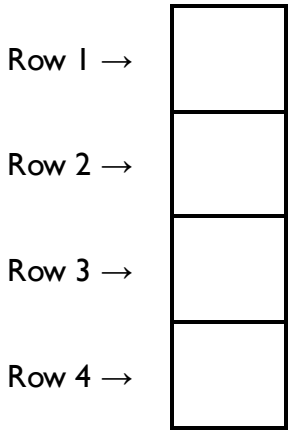
Fall 2025

What is a NumPy array?

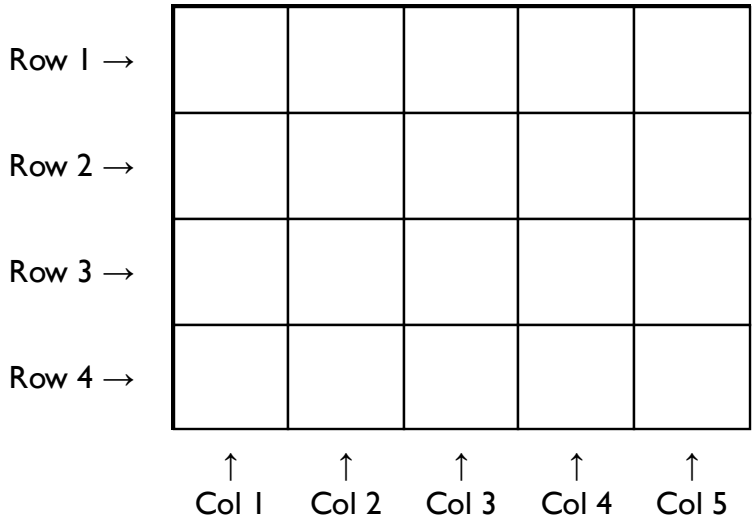
Multidimensional collection of values of the same type

- Naturally represents **scalars** (0-D), **vectors** (1-D), **matrices** (2-D), and higher-order "grids" of value.
- **Core attributes**
  - dtype — single data type of all elements (e.g., int, float).
  - shape — dimensions (e.g., (n,), (m, n), (m, n, l))
  - ndim — number of dimensions
  - size — total number of elements

1-D array (vector)



2-D array (matrix)



How to create NumPy arrays

np.array(seq) — Directly from sequence

```
np.array([1,2,3])
np.array([[1,2],[3,4]])
```

np.zeros(shape, dtype) — all zeros

```
np.zeros((2,3))
```

np.ones(shape, dtype) — all ones

```
np.ones((3,3))
```

np.full(shape, val, dtype) — constant fill

```
np.full((2,2), 99)
```

np.eye(n, dtype) — identity matrix

```
np.eye(3)
```

How to create NumPy arrays continued

np.arange([start], stop, [step]) — evenly spaced values (number of points)

```
np.arange(0, 10, 1)
```

np.linspace(start, stop, num) — evenly spaced values (step)

```
np.linspace(0, 1, 10)
```

np.random.random(shape) — random floats in [0, 1)

```
np.random.random((2, 3))
```

np.empty(shape, dtype) — uninitialized (arbitrary memory)

```
np.empty((4, 2))
```

np.zeros_like(a) / np.ones_like(a) — match shape/dtype of array

```
np.zeros_like(A)
np.ones_like(A)
```

Why using Numpy arrays instead of lists?

Lists

- Convenient for small, heterogeneous data
- Each element handled separately within loops (**slow** math)
- **Not well-suited to larger-scale numerical calculations**

Arrays

- Operations apply to **every element simultaneously**
- **No Python loops** needed

} **Vectorization**

Why using Numpy arrays instead of lists?

Lists

- Convenient for small, heterogeneous data
- Each element handled separately within loops (**slow** math)
- **Not well-suited to larger-scale numerical calculations**

Arrays

- Operations apply to **every element simultaneously**
- **No Python loops** needed

} **Vectorization**

Try it!

Calculate the values of the function $y = \sin(x)$ for **10000** evenly spaced values of x between 0 and 2π using (1) Python lists and loops, and (2) NumPy arrays with vectorized operations. Measure the time of execution in each case using %%timeit.

Operations with arrays

Mathematical operations are **vectorized**

- Executed element by element **without writing loops**
- Use fast C/Fortran code under the hood

Addition (elementwise)

```
c = a + b
d = np.add(a, b)     # equivalent
```

Multiplication vs. Matrix Product

```
a * b             # elementwise multiplication
np.multiply(a, b)   # equivalent

a @ b             # matrix multiplication
np.dot(a, b)        # equivalent
```

Operations with arrays

Aggregation functions

- Summarize array values to provide quick **statistical summaries**
- **Fast and vectorized**

np.sum(arr) → sum of all elements

np.mean(arr) → average value

np.var(arr) → variance

np.std(arr) → standard deviation

np.min(arr) → minimum value

np.max(arr) → maximum value

np.argmin(arr) → index of min

np.argmax(arr) → index of max

Indexing and slicing arrays

Select a **single element**, **row**, or **column**

Extract **ranges or subarrays** using slice notation

Use **boolean conditions** to select elements that meet a criterion

```
a[1]          → second row
a[::2]        → every other row
a[:, 1]       → second column
a[2:, 1]      → all elements after the third in second column
idx = (a > 5) → boolean mask
a[idx]        → elements greater than 5
a[a > 5]      → equivalent
```

Exercise

In a Jupyter notebook, write a Python code to compute the **matrix product** of two arbitrary 2D NumPy arrays **using only for loops** and compare the result to NumPy built-in np.dot(). Before the calculation, check that the inner dimensions of the two arrays match.

Assignment 3

Problem 3

In a `Jupyter notebook`, create a Python code to write an `XYZ` file for a water molecule with the following atomic Cartesian coordinates using Python lists (or dictionaries) and simple loops.

Atom	x (bohr)	y (bohr)	z (bohr)
O	0.00000000	0.00000000	0.00000000
H	1.80941647	0.00000000	0.00000000
H	-0.45334744	1.75170319	0.00000000

Use an **if**-statement to handle units: If the units are bohr, convert all coordinates to ångström. Otherwise (for any other value), assume the numbers are already in ångström. Write the XYZ file `water.xyz` with the standard format:

1. First line: number of atoms.
2. Second line: a short comment, e.g. `H2O coordinates in Å`.
3. Then one line per atom: `symbol x y z` with 6 decimals.

Hint: Below is an example of Python code to write to a text file. *Note* that `\n` is used to break a line.

```
with open('example.txt', 'w') as f:
    f.write('This is an example, line 1' + '\n')
    f.write('This is an example, line 2' + '\n')
```

Assignment 3

Problem 4

Using the Cartesian coordinates of the atoms of a water molecule from the previous problem, compute:

(a) The distance (in bohr) between the oxygen atom and each hydrogen atom.

(b) The H–O–H bond angle (in degrees).

Hints:

1. The distance between two points in Cartesian coordinates, $\mathbf{r}_1 = (x_1, y_1, z_1)$ and $\mathbf{r}_2 = (x_2, y_2, z_2)$, is given by:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

2. To compute the bond angle, define two vectors $\mathbf{a}$ and $\mathbf{b}$ along the O–H bonds, both pointing away from the oxygen atom:

$$\mathbf{a} = \mathbf{r}_{\mathrm{H1}} - \mathbf{r}_{\mathrm{O}}, \qquad \mathbf{b} = \mathbf{r}_{\mathrm{H2}} - \mathbf{r}_{\mathrm{O}}$$

   The bond angle θ can then be obtained from the dot product:

$$\cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}|\,|\mathbf{b}|}$$

   where $|\mathbf{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2}$ is the norm (or magnitude) of vector $\mathbf{a}$.

3. In the calculations, you can use the `np.dot`, `np.linalg.norm`, and `np.arccos` functions.