# NumPy

**Lecture 05** 

Dr. Colin Rundel

## What is NumPy?

NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

```
1 import numpy as np
2 np.__version__
```

<sup>&#</sup>x27;2.2.2'

### **Arrays**

In general NumPy arrays are constructed from sequences (e.g. lists), nesting as necessary for the number of desired dimensions.

```
1 np.array([1,2,3])
array([1, 2, 3])
 1 np.array([[1,2],[3,4]])
array([[1, 2],
       [3, 4]])
 1 np.array([[[1,2],[3,4]], [[5,6],[7,8]]
array([[[1, 2],
        [3, 4]],
       [[5, 6],
        [7, 8]]])
```

### Some properties of NumPy arrays:

- Arrays have a fixed size at creation
- All data must be homogeneous (i.e. consistent type)
- Built to support vectorized operations
- Avoids copying whenever possible (inplace operations)

## dtype

NumPy arrays all have the type() numpy.ndarray - specific type stored in the arrray is recorded as the array's dtype. This is accessible via the .dtype attribute and can be set at creation using the dtype argument.

## dtypes and overflow

Some types have a maximum and or minimum value that can be stored in them. If you try to create an array with a value outside of this range you will get an overflow error. If you are instead coercing values using astype() you will *not* get this error.

```
1 np.array([-1, 1,2], dtype = np.uint8)

OverflowError: Python integer -1 out of bounds for uint8

1 np.array([1,2,1000], dtype = np.uint8)

OverflowError: Python integer 1000 out of bounds for uint8

1 np.array([-1, 1,2,1000]).astype(np.uint8)

array([255, 1, 2, 232], dtype=uint8)
```

### **Creating 1d arrays**

Some common functions and methods for creating useful 1d arrays:

```
1 np.arange(10)
                                            1 np.ones(4)
                                          array([1., 1., 1., 1.])
array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
                                            1 np.zeros(6)
 1 np.arange(3, 5, 0.25)
array([3. , 3.25, 3.5 , 3.75, 4. ,
                                          array([0., 0., 0., 0., 0., 0.])
4.25, 4.5, 4.75])
                                            1 np.full(3, False)
 1 np.linspace(0, 1, 11)
                                          array([False, False, False])
array([0., 0.1, 0.2, 0.3, 0.4, 0.5, 0.6,
                                            1 np.empty(4)
0.7, 0.8, 0.9, 1. ])
                                          array([1., 1., 1., 1.])
 1 np.logspace(0, 2, 4)
array([ 1. , 4.64158883,
21.5443469 , 100. ])
```

# **Creating 2d arrays (matrices)**

Many of the same functions exist with some additional useful tools for common matrices,

```
1 np_eye(3)
                                              1 np.diag([3,2,1])
array([[1., 0., 0.],
                                            array([[3, 0, 0],
       [0., 1., 0.],
                                                    [0, 2, 0],
       [0., 0., 1.]]
                                                    [0, 0, 1]
 1 np.identity(2)
                                              1 np.tri(3)
                                            array([[1., 0., 0.],
array([[1., 0.],
                                                    [1., 1., 0.],
       [0., 1.]]
                                                    [1., 1., 1.]])
 1 np.zeros((2,2))
                                              1 np.triu(np.full((3,3),3))
array([[0., 0.],
       [0., 0.]]
                                            array([[3, 3, 3],
                                                    [0, 3, 3],
                                                    [0, 0, 3]]
```

### Creating nd arrays

For higher dimensional arrays just add dimensions when constructing,

```
1 np.zeros((2,3,2))
                                                 1 np.ones((2,3,2,2))
array([[[0., 0.],
                                               array([[[[1., 1.],
        [0., 0.],
                                                        [1., 1.]],
        [0., 0.]],
                                                        [[1., 1.],
       [[0., 0.],
                                                        [1., 1.]],
        [0., 0.],
        [0., 0.]]])
                                                        [[1., 1.],
                                                        [1., 1.]]],
                                                       [[[1., 1.],
                                                        [1., 1.]],
                                                        [[1., 1.],
                                                        [1., 1.]],
                                                        Γ [ 1
                                                              1 1
```

# Subsetting

Arrays are subsetted using the standard python syntax with either indexes or slices, dimensions are separated by commas.

```
1 x = np.array([[1,2,3],[4,5,6],[7,8,9]])
 2 x
array([[1, 2, 3],
        [4, 5, 6],
        [7, 8, 9]])
 1 \times [0]
                                                    1 \times [0:3:2, :]
array([1, 2, 3])
                                                  array([[1, 2, 3],
                                                          [7, 8, 9]])
 1 \times [0, 0]
                                                    1 \times [0:3:2,]
np.int64(1)
                                                  array([[1, 2, 3],
  1 \times [0][0]
                                                          [7, 8, 9]]
np.int64(1)
                                                    1 x[1:, ::-1]
 1 \times [0:3:2, :]
                                                  array([[6, 5, 4],
                                                          [9, 8, 7]])
array([[1, 2, 3],
        [7, 8, 9]])
```

### Views and copies

Basic subsetting of ndarray objects does not result in a new object, but instead a "view" of the original object. There are a couple of ways that we can investigate this behavior,

```
1 \times = np.arange(10)
 2 y = x[2:5]
 3 z = x[2:5].copy()
  1 f"{x=}, {x.base=}"
                                                     1 np.shares memory(x,y)
'x=array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]),
                                                   True
x.base=None'
                                                     1 np.shares memory(x,z)
 1 f"{y=}, {y.base=}"
                                                   False
'y=array([2, 3, 4]), y.base=array([0, 1, 2, 3,
                                                     1 np.shares memory(y,z)
4, 5, 6, 7, 8, 9])'
                                                   False
 1 f"{z=}, {z.base=}"
                                                     1 v.flags
'z=array([2, 3, 4]), z.base=None'
                                                     C CONTIGUOUS : True
  1 type(x), type(y), type(z)
                                                     F CONTIGUOUS : True
                                                     OWNDATA : False
(<class 'numpy.ndarray'>, <class</pre>
                                                     WRITEABLE: True
'numpy.ndarray'>, <class 'numpy.ndarray'>)
                                                     ALIGNED : True
                                                     WRITEBACKIFCOPY : False
```

# Subsetting with ...

Unlike R, it is not possible to leave an argument blank - to select all elements with numpy we use :. To avoid having to type excess : you can use ... which expands to the number of : needed to account for all dimensions,

```
1 \times = (np.arange(16))
                                     1 \times [0, 1, :, :]
                                                                1 x[:, :, :, 1]
 2 .reshape(2,2,2,2))
                                   array([[4, 5],
                                                              array([[[ 1, 3],
 3 x
                                          [6, 7]])
                                                                      [5, 7]
array([[[ 0, 1],
                                     1 \times [0, 1, ...]
                                                                     [[ 9, 11],
         [ 2, 3]],
                                                                      [13, 15]])
                                   array([[4, 5],
        [[4, 5],
                                          [6, 7]])
                                                               1 x[..., 1]
         [ 6, 7]]],
                                                              array([[[ 1, 3],
       [[[8, 9],
                                                                      [5, 7]],
         [10, 11]],
                                                                     [[9, 11],
        [[12, 13],
                                                                      [13, 15]])
         [14, 15]]])
```

# **Subsetting with tuples**

Unlike lists, an idarray can be subset by a tuple containing integers,

```
1 x = np.arange(6); x

array([0, 1, 2, 3, 4, 5])

1 x[(0,1,3),]

array([0, 1, 3])

1 x[(3,5,1,0),]

array([3, 5, 1, 0])

1 x[(0,1,3)]
```

IndexError: too many indices for array: array
is 1-dimensional, but 3 were indexed

```
1 x = np.arange(16).reshape((4,4))
 2 x
array([[ 0, 1, 2, 3],
       [4, 5, 6, 7],
       [8, 9, 10, 11],
       [12, 13, 14, 15]])
 1 \times [(0,1,3), :]
array([[ 0, 1, 2, 3],
       [4, 5, 6, 7],
       [12, 13, 14, 15]])
 1 \times [:, (0,1,3)]
array([[ 0, 1, 3],
       [4, 5, 7],
       [8, 9, 11],
       [12, 13, 15]])
 1 \times [(0,1,3), (0,1,3)]
array([ 0, 5, 15])
```

# **Subsetting assignment**

Most of the subsetting approaches we've just seen can also be used for assignment, just keep in mind that we cannot change the *size* or *type* of the ndarray,

```
1 x = np.arange(9).reshape((3,3)); x
array([[0, 1, 2],
       [3, 4, 5].
       [6, 7, 8]]
 1 \times [0,0] = -1; \times
                                                 1 \times [0:2,1:3] = -3; \times
                                               array([-2, -3, -3],
array([-1, 1, 2],
       [ 3, 4, 5],
                                                      [3, -3, -3]
       [ 6, 7, 8]])
                                                       [6, 7, 8]])
 1 \times [0, :] = -2; \times
                                                 1 \times [(0,1,2), (0,1,2)] = -4; \times
array([-2, -2, -2],
                                               array([-4, -3, -3],
       [3, 4, 5],
                                                      [3, -4, -3],
                                                       [6, 7, -4]
       [6, 7, 8]])
 1 \times [0,0] = "A"
```

ValueError: invalid literal for int() with base 10: 'A'

### Reshaping arrays

The dimensions of an array can be retrieved via the shape attribute, these values can changed via the reshape() method or updating shape

```
1 \times = np.arange(6); \times
array([0, 1, 2, 3, 4, 5])
 1 y = x.reshape((2,3)); y
                                                1 z = x
                                                2 z.shape = (2,3)
array([[0, 1, 2],
       [3, 4, 5]])
                                              array([[0, 1, 2],
 1 np.shares_memory(x,y)
                                                      [3, 4, 5]])
True
                                                1 x
                                              array([[0, 1, 2],
                                                      [3, 4, 5]])
                                                1 np.shares memory(x,z)
                                              True
```

## Implicit dimensions

When reshaping an array, the value –1 can be used to automatically calculate a dimension,

```
1 x = np.arange(6); x
array([0, 1, 2, 3, 4, 5])
 1 x.reshape((2,-1))
                                       1 x_reshape(-1)
array([[0, 1, 2],
                                     array([0, 1, 2, 3, 4, 5])
       [3, 4, 5]])
                                       1 x.reshape((-1,4))
 1 x.reshape((-1,3,2))
                                     ValueError: cannot reshape array
array([[[0, 1],
                                     of size 6 into shape (4)
        [2, 3],
        [4, 5]])
```

## Flattening arrays

We just saw one of the more common approaches to creating a flat *view* of an array (reshape(-1)), there are two other common methods / functions:

- ravel creates a flattened view of the array and
- flatten creates a flattened *copy* of the array.

```
1 w = np.arange(6).reshape((2,3)); w
array([[0, 1, 2],
       [3, 4, 5]]
 1 \times = w_{\bullet} \operatorname{reshape}(-1)
                                    y = w.ravel()
                                                                  1 z = w.flatten()
 2 x
                                 2 y
                                                                  2 z
array([0, 1, 2, 3, 4, 5])
                                array([0, 1, 2, 3, 4, 5])
                                                                array([0, 1, 2, 3, 4, 5])
 1 np.shares_memory(w,x)
                                  1 np.shares_memory(w,y)
                                                                  1 np.shares_memory(w,z)
True
                                True
                                                                False
```

# Resizing

The size of an array cannot be changed but a new array with a different size can be created from an existing array via the resize function and method. Note these have different behaviors around what values the new entries will have.

```
1 x = np.resize(
2 np.ones((2,2)),
3 (3,3)
4 )
5 x

1 y = np.ones(
2 (2,2)
3 ).resize(
4 (3,3)
5 )
6 y
```

Why didn't this work?

```
1 y = np.ones(
2 (2,2)
3 )
4 y.resize((3,3))
5 y
```

```
array([[1., 1., 1.],
[1., 0., 0.],
[0., 0., 0.]])
```

[1., 1., 1.],

[1., 1., 1.]])

## Joining arrays

concatenate() is a general purpose function for joining arrays, with specialized versions hstack(), vstack(), and dstack() for rows, columns, and slices respectively.

```
1 x = np.arange(4).reshape((2,2)); x
                                              1 y = np.arange(4,8).reshape((2,2)); y
                                            array([[4, 5],
array([[0, 1],
       [2, 3]])
                                                    [6, 7]
 1 np.concatenate((x,y), axis=0)
                                              1 np.vstack((x,y))
array([[0, 1],
                                            array([[0, 1],
       [2, 3],
                                                    [2, 3],
       [4, 5],
                                                    [4, 5]
       [6, 7]])
                                                    [6, 7]]
 1 np.concatenate((x,y), axis=1)
                                              1 np.hstack((x,y))
                                            array([[0, 1, 4, 5],
array([[0, 1, 4, 5],
       [2, 3, 6, 7]])
                                                    [2, 3, 6, 7]])
```

1 np.concatenate((x,y), axis=2)

1 np.dstack((x,y))

numpy.exceptions.AxisError: axis 2 is out
of bounds for array of dimension 2

array([[[0, 4], [1, 5]],

1 np.concatenate((x,y), axis=None)

[[2, 6],

array([0, 1, 2, 3, 4, 5, 6, 7])

[3, 7]]])

# NumPy numerics

### **Basic operators**

All of the basic mathematical operators in Python are implemented for arrays, they are applied element-wise to the array values.

```
1 np.arange(3) + np.arange(3)
                                               1 np.arange(3) * np.arange(3)
array([0, 2, 4])
                                             array([0, 1, 4])
 1 np.arange(3) - np.arange(3)
                                               1 np.arange(1,4) / np.arange(1,4)
array([0, 0, 0])
                                             array([1., 1., 1.])
 1 \text{ np.arange}(3) + 2
                                               1 np.arange(3) * 3
array([2, 3, 4])
                                             array([0, 3, 6])
 1 np.full((2,2), 2) ** np.arange(4).reshape((2,2))
array([[1, 2],
       [4, 8]])
 1 np.full((2,2), 2) ** np.arange(4)
```

ValueError: operands could not be broadcast together with shapes (2,2) (4,)

#### **Mathematical functions**

NumPy provides a wide variety of basic mathematical functions that are vectorized, in general they will be faster than their base equivalents (e.g. np. sum() vs sum()),

```
1 np.sum(np.arange(1000))
np.int64(499500)
 1 np.cumsum(np.arange(10))
array([ 0, 1, 3, 6, 10, 15, 21, 28, 36, 45])
 1 np.log10(np.arange(1,4))
array([0.
                 , 0.30103 , 0.47712125])
 1 np.median(np.arange(10))
np.float64(4.5)
```

### **Matrix multiplication**

is supported using the matmul() function or the @ operator,

```
1 \times = np.arange(6).reshape(3,2)
 2 y = np.tri(2,2)
                                     1 np.matmul(x.T, x)
 1 x @ y
                                   array([[20, 26],
array([[1., 1.],
       [5., 3.],
                                           [26, 35]])
       [9., 5.]]
                                     1 y @ x
 1 y.T @ y
                                   ValueError: matmul: Input
array([[2., 1.],
                                   operand 1 has a mismatch in its
       [1., 1.]])
                                   core dimension 0, with gufunc
                                   signature (n?,k),(k,m?) \rightarrow
                                   (n?,m?) (size 3 is different
                                   from 2)
```

## Other linear algebra functions

All of the other common linear algebra functions are (mostly) implemented in the linalg submodule.

```
1 np.linalq.det(y)
np.float64(1.0)
 1 np.linalg.eig(x.T @ x)
EigResult(eigenvalues=array([ 0.43988174, 54.56011826]),
eigenvectors=array([-0.79911221, -0.6011819],
       [0.6011819, -0.79911221]))
 1 np.linalq.inv(x.T @ x)
array([[ 1.45833333, -1.08333333],
      [-1.083333333]
 1 np.linalg.cholesky(x.T @ x)
array([[4.47213595, 0.
       [5.81377674, 1.09544512]])
```

#### Random values

NumPy has another submodule called random for functions used to generate random values.

In order to use this, you construct a generator via default\_rng(), with or without a seed, and then use the generator's methods to obtain your desired random values.

```
1 rng = np.random.default_rng(seed = 1234)
 1 rng.random(3) \# \sim Uniform [0,1)
array([0.97669977, 0.38019574, 0.92324623])
 1 rng.normal(loc=0, scale=2, size = (2,2))
array([[ 0.30523839, 1.72748778],
       [ 5.82619845, -2.95764672]])
 1 rng.binomial(n=5, p=0.5, size = 10)
array([2, 4, 2, 2, 3, 4, 4, 3, 3, 3])
```

# **Advanced Indexing**

# **Advanced Indexing**

Advanced indexing is triggered when the selection object, obj, is a non-tuple sequence object, an indexray (of data type integer or bool), or a tuple with at least one sequence object or indexray (of data type integer or bool).

- There are two types of advanced indexing: integer and Boolean.
- Advanced indexing always returns a *copy* of the data (contrast with basic slicing that returns a view).

# Integer array subsetting (lists)

Lists of integers can be used to subset in the same way:

```
1 x = np.arange(16).reshape((4,4)); x
array([[ 0, 1, 2, 3],
       [4, 5, 6, 7],
       [8, 9, 10, 11],
       [12, 13, 14, 15]])
  1 \times [[1,3]]
                                                    1 \times [[0,1,3],]
array([[ 4, 5, 6, 7],
                                                  array([[ 0, 1, 2, 3],
      [12, 13, 14, 15]])
                                                          [4, 5, 6, 7],
                                                          [12, 13, 14, 15]])
  1 \times [[1,3],]
                                                    1 \times [[0,1,3]]
array([[ 4, 5, 6, 7],
       [12, 13, 14, 15]])
                                                  array([[ 0, 1, 2, 3],
                                                          [4, 5, 6, 7],
  1 x[:, [1,3]]
                                                          [12, 13, 14, 15]])
array([[ 1, 3],
                                                    1 \times [[1.,3]]
      [5, 7],
       [ 9, 11],
                                                  IndexError: only integers, slices (`:`),
       [13, 15]])
                                                  ellipsis (`...`), numpy.newaxis (`None`) and
                                                  integer or boolean arrays are valid indices
```

# Integer array subsetting (ndarrays)

Similarly we can also us integer ndarrays:

```
1 \times = np_arange(6)
                                                 1 x = np.arange(16).reshape((4,4))
 2 y = np.array([0,1,3])
                                                 2 y = np_array([1,3])
 3 z = np.array([1., 3.])
                                                 1 x[y]
 1 \times [y,]
                                               array([[ 4, 5, 6, 7],
array([0, 1, 3])
                                                       [12, 13, 14, 15]])
 1 \times [y]
                                                 1 \times [y,]
array([0, 1, 3])
                                               array([[ 4, 5, 6, 7],
                                                       [12, 13, 14, 15]])
 1 \times [z]
                                                 1 x[:, y]
IndexError: arrays used as indices must
be of integer (or boolean) type
                                               array([[ 1, 3],
                                                       [5, 7],
                                                       [ 9, 11],
                                                       [13, 15]])
                                                 1 \times [y, y]
                                               array([ 5, 15])
```

#### Exercise 1

Given the following matrix,

```
1 x = np.arange(16).reshape((4,4))
2 x

array([[ 0,  1,  2,  3],
       [ 4,  5,  6,  7],
       [ 8,  9,  10,  11],
       [12,  13,  14,  15]])
```

write an expression to obtain the center 2x2 values (i.e. 5, 6, 9, 10 as a matrix).

# **Boolean indexing**

Lists or ndarrays of boolean values can also be used to subset (positions with True are kept and False are discarded)

```
1 x = np.arange(6); x
array([0, 1, 2, 3, 4, 5])
 1 x[[True, False, True, False, True, False]]
array([0, 2, 4])
 1 x[[True]]
IndexError: boolean index did not match indexed array along axis 0; size of axis is 6
but size of corresponding boolean axis is 1
 1 x[np.array([True, False, True, False, True, False])]
array([0, 2, 4])
 1 x[np.array([True])]
```

IndexError: boolean index did not match indexed array along axis 0; size of axis is 6 but size of corresponding boolean axis is 1

### **Boolean expressions**

The primary utility of boolean subsetting comes from vectorized comparison operations,

```
1 \times = np.arange(6); \times
array([0, 1, 2, 3, 4, 5])
 1 x > 3
                                                1 y = np.arange(9).reshape((3,3))
                                                2 y % 2 == 0
array([False, False, False, True,
Truel)
                                               array([[ True, False, True],
                                                      [False, True, False],
 1 \times [x>3]
                                                      [ True, False, True]])
array([4, 5])
                                                1 \ v[v \% 2 == 0]
 1 x % 2 == 1
                                               array([0, 2, 4, 6, 8])
array([False, True, False, True, False,
Truel)
 1 \times [\times \% 2 == 1]
array([1, 3, 5])
```

### NumPy and Boolean operators

If we want to use a logical operators on an array we need to use &, |, and  $\sim$  instead of and, or, and not respectively.

```
1 \times = np.arange(6); \times
array([0, 1, 2, 3, 4, 5])
 1 y = (x \% 2 == 0); y
array([ True, False, True, False, True, False])
 1 ~y
array([False, True, False, True, False, True])
 1 y \& (x > 3)
array([False, False, False, True, False])
 1 y | (x > 3)
array([ True, False, True, False, True, True])
```

# meshgrid()

One other useful function in NumPy is meshgrid() which generates all possible combinations between the input vectors (as a tuple of ndarrays),

### **Exercise 2**

We will now use this to attempt a simple brute force approach to numerical optimization, define a grid of points using meshgrid() to approximate the minima the following function:

$$f(x,y) = (1-x)^2 + 100(y-x^2)^2$$

Considering values of  $x, y \in (-1, 3)$ , which value(s) of x, y minimize this function?

