dsb-bulk

17 2024 .

1 Differential Spectrum Balance - Bulk experiments

To find out where it works or not i need to check a lot of possible combinations between data and approximation model, as well as the sizes of the dataset. This notebook is for that mass testing hence it won't feature a lot of diagrams, just tables with data.

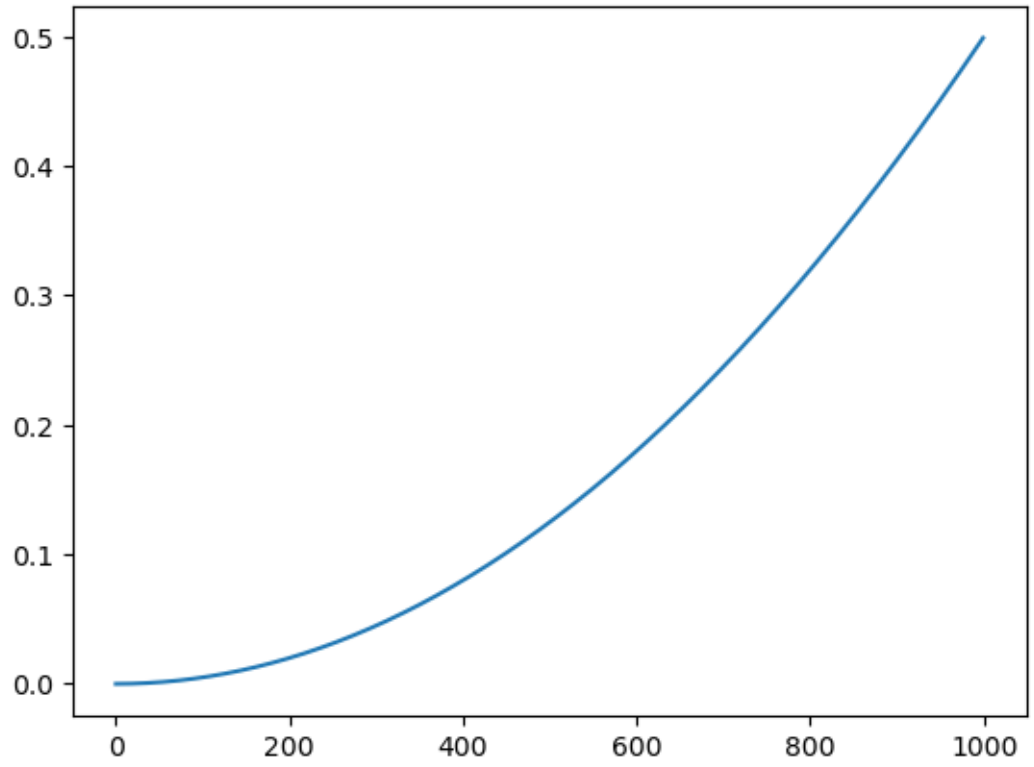
[ ]:
```python
# Main imports
import sympy as sp
import numpy as np
import matplotlib.pyplot as plt

from modules.extra import dsb_fit
from modules.utils import statistics
from modules.models import apply_noise


# Test function, approximates polys of different ranks
# against given data of varied size.
def mass_test(
    expr: str,
    data: list[tuple[np.ndarray, int]],
    ideal: list[np.ndarray],
    ranks: list[int],
    numeric: bool = False,
):
    for i, d in enumerate(data):
        n = d[1]
        data_model = d[0]
        ideal_model = ideal[i]
        results = []
        failed = []
        for r in ranks:
            try:
                results.append(
                    (
                        dsb_fit(expr, "t", data_model, rank=r, numeric=numeric),
                        f"rank:{r}",
                    )
```

```
                )
            except Exception as e:
                failed.append(f"rank:{r} failed: {e}")
        display(f"Data model size: {n}")
        statistics(ideal_model, *results)
        for f in failed:
            print(f)
```

1.1 Exponential testing

- Data model: $a_0 + e^{a_1 * t}$
- Slow growing

[ ]:
```
from modules.models import exponential1

dd = exponential1(1000)
plt.plot(dd)
plt.show()
```



[ ]:
```
sizes = [10, 100, 1000, 10000, 20000, 30000]
test_data = [(apply_noise(exponential1(s)), s) for s in sizes]
ideal_data = [exponential1(s) for s in sizes]
```

```
display("No numeric extension:")
mass_test("a0 + exp(a1*t)", test_data, ideal_data, range(2, 7))
display("With numeric extension:")
mass_test("a0 + exp(a1*t)", test_data, ideal_data, range(2, 7), numeric=True)
```

```
'No numeric extension:'
```

```
'Data model size: 10'
```

```
         median        varian  std. div.   lin. div.         covar       concord
rank:2  8.93705  7.984653e+03  89.356884  578.000593  1.114957e-03  2.765090e-07
data    0.00001  1.802625e-10   0.000013    0.000000  1.802625e-10  1.000000e+00
rank:3 failed: No nonlinear solutions found for balance.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
```

```
'Data model size: 100'
```

```
          median    varian  std. div.   lin. div.     covar   concord
rank:2 -2.586759  0.000412   0.020293  258.818695 -0.000029 -0.000009
data    0.001225  0.000002   0.001477    0.000000  0.000002  1.000000
rank:3 failed: No nonlinear solutions found for balance.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
```

```
'Data model size: 1000'
```

```
          median    varian  std. div.   lin. div.     covar   concord
rank:2  0.099311  1.811898   1.346068  998.776474  0.200249  0.218288
data    0.124750  0.022181   0.148931    0.000000  0.022181  1.000000
rank:3 failed: No nonlinear solutions found for balance.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
```

```
'Data model size: 10000'
```

```
              median        varian     std. div.     lin. div.         covar
concord
rank:2  6.353157e+10  1.579177e+41  3.973886e+20  8.132488e+23  2.550467e+21
3.230121e-20
data    1.249750e+01  2.221806e+02  1.490572e+01  0.000000e+00  2.221806e+02
1.000000e+00
rank:3 failed: No nonlinear solutions found for balance.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
```

```
'Data model size: 20000'
```

```
              median         varian     std. div.     lin. div.         covar
concord
rank:2  2.669433e+43  1.269327e+171  3.562762e+85  7.161523e+88  4.702912e+86
7.410087e-85
data    4.999500e+01   3.555222e+03  5.962568e+01  0.000000e+00  3.555222e+03
1.000000e+00
rank:3 failed: No nonlinear solutions found for balance.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 30000'

              median        varian   std. div.      lin. div.          covar
concord
rank:2  5.258772e+97           inf         inf  1.857254e+197  1.844854e+195
0.0
data    1.124925e+02  1.799887e+04  134.159886   0.000000e+00   1.799887e+04
1.0
rank:3 failed: No nonlinear solutions found for balance.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

c:\Users\ringa\miniconda3\envs\data-science\Lib\site-
packages\numpy\core\_methods.py:176: RuntimeWarning: overflow encountered in
multiply
  x = um.multiply(x, x, out=x)

'With numeric extension:'

'Data model size: 10'

          median        varian  std. div.  lin. div.         covar   concord
rank:2 -1.539105  2.285720e+00   1.511860  16.782851  2.026987e-05  0.000009
data    0.000010  1.802625e-10   0.000013   0.000000  1.802625e-10  1.000000
rank:3 failed: No nonlinear solutions found for balance.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 100'

          median    varian  std. div.   lin. div.     covar   concord
rank:2 -1.311918  0.000352   0.018755  131.337792 -0.000027 -0.000031
data    0.001225  0.000002   0.001477    0.000000  0.000002  1.000000
rank:3 failed: No nonlinear solutions found for balance.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 1000'
```

```
          median    varian  std. div.   lin. div.     covar   concord
rank:2  0.161574  0.215452   0.464168  273.813285  0.068543  0.573608
data    0.124750  0.022181   0.148931    0.000000  0.022181  1.000000
rank:3 failed: No nonlinear solutions found for balance.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
```

```
'Data model size: 10000'
```

```
           median      varian  std. div.     lin. div.       covar   concord
rank:2  10.549542  198.510782  14.089385  30605.514994  204.461361  0.963336
data    12.497500  222.180555  14.905722      0.000000  222.180555  1.000000
rank:3 failed: No nonlinear solutions found for balance.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
```

```
'Data model size: 20000'
```

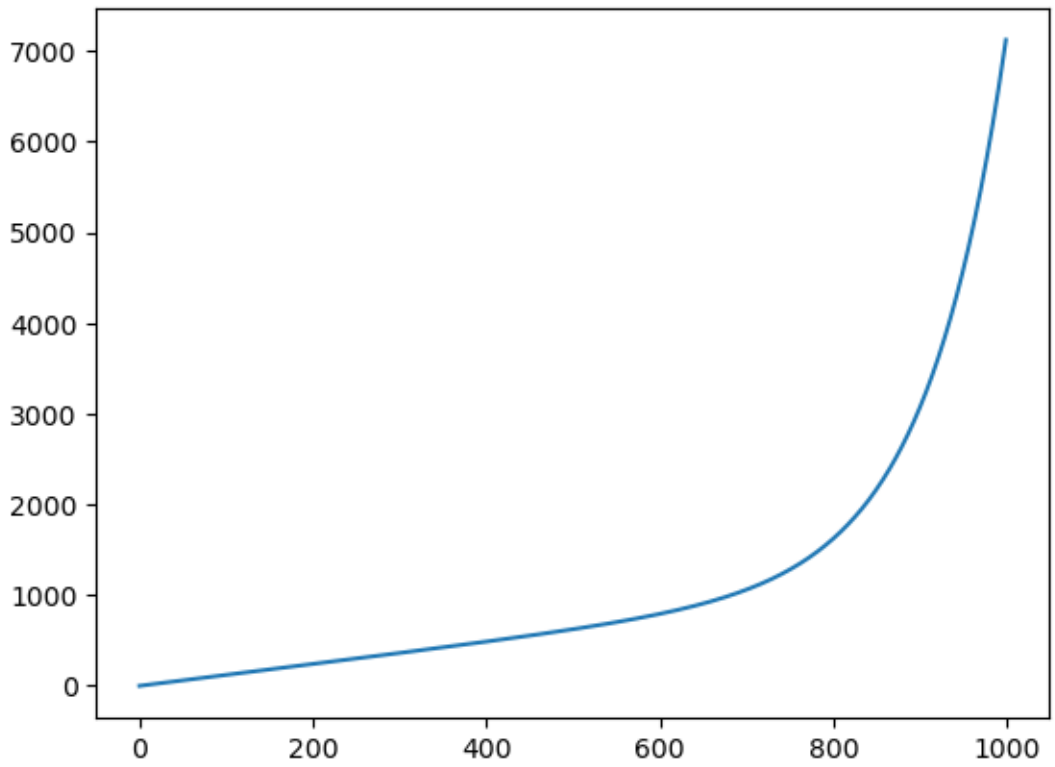
```
           median       varian  std. div.     lin. div.        covar   concord
rank:2  40.350227  2950.625285  54.319658  367512.19337  3040.900876  0.921643
data    49.995000  3555.222219  59.625684       0.00000  3555.222219  1.000000
rank:3 failed: No nonlinear solutions found for balance.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
```

```
'Data model size: 30000'
```

```
            median        varian   std. div.     lin. div.         covar
concord
rank:2   90.965567  14166.559038  119.023355  1.443113e+06  14632.209901
0.896888
data    112.492500  17998.874994  134.159886  0.000000e+00  17998.874994
1.000000
rank:3 failed: No nonlinear solutions found for balance.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
```

- Data model: $a_0 + a_1 * t + a_2 * e^{a_3 * t}$
- Fast growing

[ ]:
```
from modules.models import exponential2

dd = exponential2(1000)
plt.plot(dd)
plt.show()
```



[ ]:
```
sizes = [10, 100, 200, 500, 1000, 2000]
test_data = [(apply_noise(exponential2(s)), s) for s in sizes]
ideal_data = [exponential2(s) for s in sizes]

# display("No numeric extension:")
# mass_test("a0 + a1*t + exp(a2*t)", test_data, ideal_data, range(3, 7))
display("With numeric extension:")
mass_test("a0 + a1*t + exp(a2*t)", test_data, ideal_data, range(3, 7),␣
 ↪numeric=True)
```

```
'With numeric extension:'

'Data model size: 10'

        median     varian  std. div.  lin. div.      covar  concord
data  5.505076  11.902903   3.450058        0.0  11.902903      1.0
rank:3 failed: Cannot convert complex to float
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 100'

           median      varian  std. div.  lin. div.        covar   concord
```

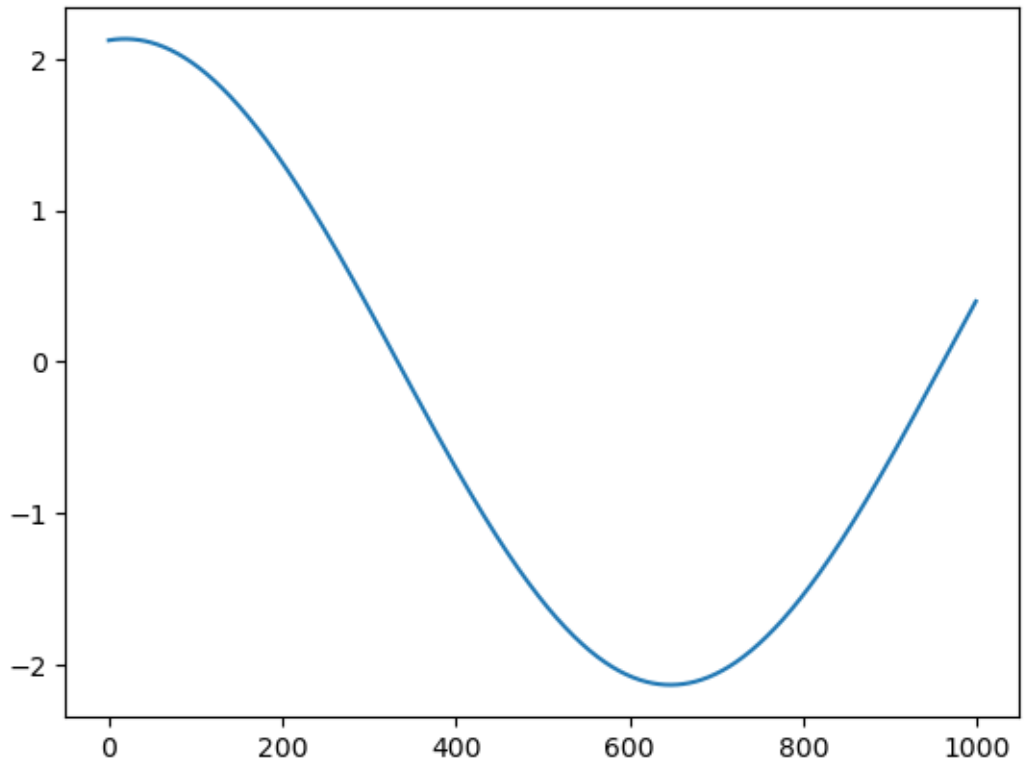
```
rank:3  59.151432  1214.77448  34.853615  42.473659  1209.270279  0.999916
data    59.572377  1203.79103  34.695692   0.000000  1203.791030  1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 200'

          median       varian  std. div.  lin. div.        covar  concord
data  119.698773  4829.541839  69.494905        0.0  4829.541839      1.0
rank:3 failed: Cannot convert complex to float
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 500'

            median        varian   std. div.   lin. div.         covar   concord
rank:3  300.266674  31688.717605  178.013251  240.321398  31713.418300  0.999987
data    300.955707  31738.443708  178.152866    0.000000  31738.443708  1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 1000'

            median        varian    std. div.     lin. div.         covar
concord
rank:3  535.456769  1.969719e+06  1403.466901  72411.409243  1.965929e+06
0.996004
data    623.735350  1.970119e+06  1403.609117      0.000000  1.970119e+06
1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

<lambdifygenerated-102>:2: RuntimeWarning: overflow encountered in exp
  return a0 + a1*t + exp(a2*t)

'Data model size: 2000'

           median        varian     std. div.  lin. div.         covar  concord
data  7154.063851  2.626221e+15  5.124667e+07        0.0  2.626221e+15      1.0
rank:3 failed: Optimal parameters not found: Number of calls to function has
reached maxfev = 100000.
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
```

1.2 Transcendental testing

- Data model: $a_1 * sin(a_2 * t) + a_3 * cos(a_2 * t)$

- Periodic

[ ]:
```
from modules.models import transcendental1

dd = transcendental1(1000)
plt.plot(dd)
plt.show()
```



[ ]:
```
sizes = [10, 100, 1000, 10000, 20000, 30000]
test_data = [(apply_noise(transcendental1(s)), s) for s in sizes]
ideal_data = [transcendental1(s) for s in sizes]

display("No numeric extension:")
mass_test("a0*sin(a1*t) + a2*cos(a1*t)", test_data, ideal_data, range(3, 7))
display("With numeric extension:")
mass_test("a0*sin(a1*t) + a2*cos(a1*t)", test_data, ideal_data, range(3, 7),␣
 ↪numeric=True)
```

```
'No numeric extension:'
```

```
'Data model size: 10'
```

```
          median     varian  std. div.   lin. div.     covar   concord
rank:3 -6.431228  93.876648   9.688996  104.674222 -0.021252 -0.000254
```

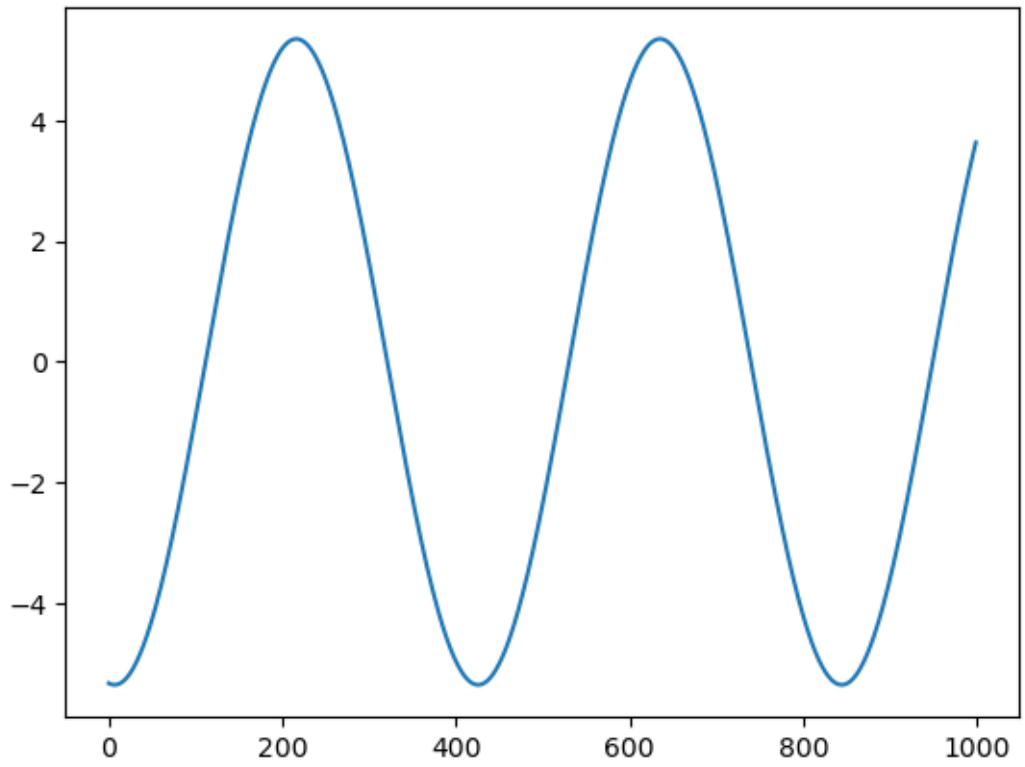
```
data    2.123956   0.000005   0.002195    0.000000  0.000005  1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 100'

          median    varian  std. div.  lin. div.     covar   concord
rank:3  2.621343  0.865383   0.930260  92.486557 -0.044756 -0.078852
data    2.104389  0.002560   0.050595   0.000000  0.002560  1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 1000'

          median    varian  std. div.    lin. div.     covar   concord
rank:3 -3.743459  6.654156   2.579565  2717.363226  3.167389  0.348536
data   -0.671451  2.083991   1.443603     0.000000  2.083991  1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 10000'

          median    varian  std. div.     lin. div.     covar   concord
rank:3  0.744385  1.952045   1.397156  19758.389071 -0.050248 -0.020984
data   -0.017666  2.256412   1.502136      0.000000  2.256412  1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 20000'

          median    varian  std. div.     lin. div.     covar   concord
rank:3  0.334896  0.065953   0.256814  27563.173054 -0.005002 -0.004086
data   -0.017666  2.258491   1.502828      0.000000  2.258491  1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 30000'

          median    varian  std. div.     lin. div.     covar   concord
rank:3  0.241747  0.016794   0.129592  40864.431865 -0.001109 -0.000946
data   -0.017666  2.261174   1.503720      0.000000  2.261174  1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'With numeric extension:'

'Data model size: 10'
```

```
          median    varian  std. div.  lin. div.     covar   concord
rank:3  2.795689  6.126353   2.475147  22.043391 -0.005412 -0.001646
data    2.123956  0.000005   0.002195   0.000000  0.000005  1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 100'

        median   varian  std. div.  lin. div.    covar  concord
data  2.104389  0.00256   0.050595        0.0  0.00256      1.0
rank:3 failed: Cannot convert complex to float
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 1000'

          median    varian  std. div.  lin. div.     covar   concord
rank:3 -0.712288  1.083844   1.041078  379.42681  1.493084  0.942157
data   -0.671451  2.083991   1.443603    0.00000  2.083991  1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 10000'

        median    varian  std. div.  lin. div.     covar  concord
data -0.017666  2.256412   1.502136        0.0  2.256412      1.0
rank:3 failed: Cannot convert complex to float
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 20000'

          median    varian  std. div.    lin. div.     covar   concord
rank:3 -0.021734  0.000648   0.025461  27054.95232 -0.000645 -0.000571
data   -0.017666  2.258491   1.502828      0.00000  2.258491  1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.

'Data model size: 30000'

          median    varian  std. div.     lin. div.     covar   concord
rank:3 -0.008905  0.006089   0.078031  40598.772114  0.001384  0.001221
data   -0.017666  2.261174   1.503720      0.000000  2.261174  1.000000
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
```

- Data model: $a_1 * cos(a_2 * t) + a_3 * sin(a_2 * t) + a_4 * cos(a_2 * t)$

- Curved, Periodic

[ ]:
```
from modules.models import transcendental3

dd = transcendental3(1000)
plt.plot(dd)
plt.show()
```



[ ]:
```
sizes = [10, 100, 1000, 10000, 20000, 30000]
test_data = [(apply_noise(transcendental3(s)), s) for s in sizes]
ideal_data = [transcendental3(s) for s in sizes]

display("No numeric extension:")
mass_test("a0*cos(a1*t) + a2*sin(a1*t) + a3*cos(a1*t)", test_data, ideal_data,␣
 ↪range(4, 8))
display("With numeric extension:")
mass_test("a0*cos(a1*t) + a2*sin(a1*t) + a3*cos(a1*t)", test_data, ideal_data,␣
 ↪range(4, 8), numeric=True)
```

```
'No numeric extension:'

'Data model size: 10'

        median    varian  std. div.  lin. div.     covar  concord
```

```
data -5.343541  0.000079   0.008894        0.0  0.000079      1.0
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
rank:7 failed: No nonlinear solutions found for balance.

'Data model size: 100'

        median    varian  std. div.  lin. div.     covar  concord
data -4.284663  1.833231   1.353969        0.0  1.833231      1.0
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
rank:7 failed: No nonlinear solutions found for balance.

'Data model size: 1000'

        median     varian  std. div.  lin. div.      covar  concord
data -0.498163  13.835007   3.719544        0.0  13.835007      1.0
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
rank:7 failed: No nonlinear solutions found for balance.

'Data model size: 10000'

        median     varian  std. div.  lin. div.      covar  concord
data  0.044615  14.260392   3.776293        0.0  14.260392      1.0
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
rank:7 failed: No nonlinear solutions found for balance.

'Data model size: 20000'

        median     varian  std. div.  lin. div.      covar  concord
data  0.038324  14.308051   3.782598        0.0  14.308051      1.0
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
rank:7 failed: No nonlinear solutions found for balance.

'Data model size: 30000'

        median     varian  std. div.  lin. div.      covar  concord
data  0.009887  14.315643   3.783602        0.0  14.315643      1.0
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
rank:7 failed: No nonlinear solutions found for balance.

'With numeric extension:'

'Data model size: 10'
```

```
        median    varian  std. div.  lin. div.     covar  concord
data -5.343541  0.000079   0.008894        0.0  0.000079      1.0
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
rank:7 failed: No nonlinear solutions found for balance.

'Data model size: 100'

        median    varian  std. div.  lin. div.     covar  concord
data -4.284663  1.833231   1.353969        0.0  1.833231      1.0
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
rank:7 failed: No nonlinear solutions found for balance.

'Data model size: 1000'

        median     varian  std. div.  lin. div.      covar  concord
data -0.498163  13.835007   3.719544        0.0  13.835007      1.0
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
rank:7 failed: No nonlinear solutions found for balance.

'Data model size: 10000'

        median     varian  std. div.  lin. div.      covar  concord
data  0.044615  14.260392   3.776293        0.0  14.260392      1.0
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
rank:7 failed: No nonlinear solutions found for balance.

'Data model size: 20000'

        median     varian  std. div.  lin. div.      covar  concord
data  0.038324  14.308051   3.782598        0.0  14.308051      1.0
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
rank:7 failed: No nonlinear solutions found for balance.

'Data model size: 30000'

        median     varian  std. div.  lin. div.      covar  concord
data  0.009887  14.315643   3.783602        0.0  14.315643      1.0
rank:4 failed: No nonlinear solutions found for balance.
rank:5 failed: No nonlinear solutions found for balance.
rank:6 failed: No nonlinear solutions found for balance.
rank:7 failed: No nonlinear solutions found for balance.
```