2023-24 Artificial Evolution Mini-Project

Problem to be solved : Sines peeling

The standard way to decompose a sum of sines is to use a Fourier Transform. However, it is impossible to ask an FT to find one sine out of a sum of n sines, but this is what is needed when a signal is made of a large number of sines, where scientists are only interested by 4 or 5 of them. Concerning FT, once more, it tries to find them all at once, but this is very difficult if there are 10000 sines in the signal (case for a molecule of oil, for instance).

The objective of this mini-project is to see if we could “peel” sines from a signal containing many sines.

For this, we need a difficult signal that we can create artificially. Rather than creating a sum of 3 sines, please create a sum of 100 sines.

In order to have amplitudes, frequencies and phases that are not multiples of each other (harmonics), here is a nice list of prime numbers: https://prime-numbers.info/list/primes

Out of this list, I suggest to use the 100 prime numbers starting with 31 as our amplitudes **IN DECREASING ORDER (this is important)**.

11 - 20	31	37	41	43	47	53	59	61	67	71
21 - 30	73	79	83	89	97	101	103	107	109	113
31 - 40	127	131	137	139	149	151	157	163	167	173
41 - 50	179	181	191	193	197	199	211	223	227	229
51 - 60	233	239	241	251	257	263	269	271	277	281
61 - 70	283	293	307	311	313	317	331	337	347	349
71 - 80	353	359	367	373	379	383	389	397	401	409
81 - 90	419	421	431	433	439	443	449	457	461	463
91 - 100	467	479	487	491	499	503	509	521	523	541
101 - 110	547	557	563	569	571	577	587	593	599	601

Then, for frequencies, we could use the first 100 sines above 1000:
1009, 1013, 1019, 1021, 1031, 1033, 1039, 1049, 1051, 1061, 1063, 1069, 1087, 1091, 1093, 1097, 1103, 1109, 1117, 1123, 1129, 1151, 1153, 1163, 1171, 1181, 1187, 1193, 1201, 1213, 1217, 1223, 1229, 1231, 1237, 1249, 1259, 1277, 1279, 1283, 1289, 1291, 1297, 1301, 1303, 1307, 1319, 1321, 1327, 1361, 1367, 1373, 1381, 1399, 1409, 1423, 1427, 1429, 1433, 1439, 1447, 1451, 1453, 1459, 1471, 1481, 1483, 1487, 1489, 1493, 1499, 1511, 1523, 1531, 1543, 1549, 1553, 1559, 1567, 1571, 1579, 1583, 1597, 1601, 1607, 1609, 1613, 1619, 1621, 1627, 1637, 1657, 1663, 1667, 1669, 1693, 1697, 1699, 1709, 1721

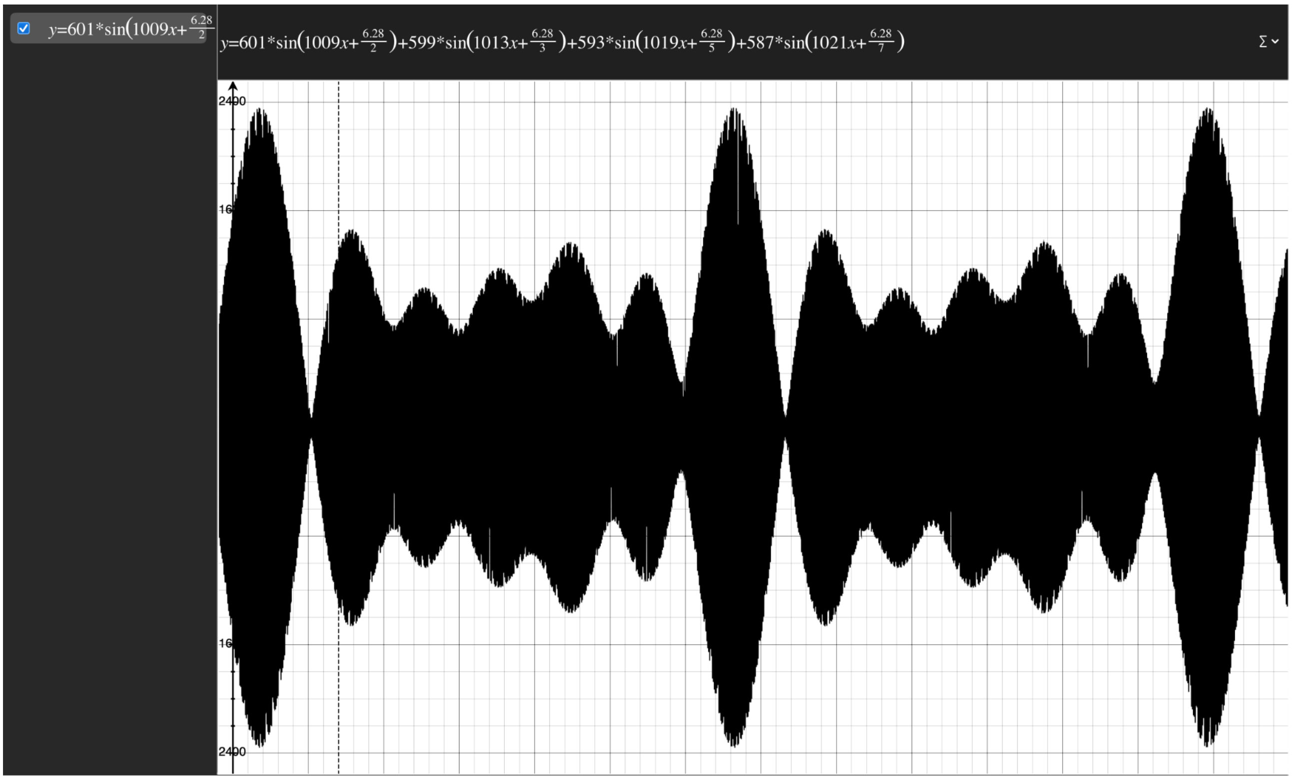
And finally, for the phases of each of these sines, we could take the exact value 6.28 divided by the first 100 primes:
2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 521, 523, 541

So, to make sure that everything is clear:
the 1st sine would be **601***sin(1009x+6.28/2)
the 2nd sine would be **599***sin(1013x+6.28/3)
the 3rd sine would be **593***sin(1019x+6.28/5)
the 4th sine would be **587***sin(1021x+6.28/7)
and so on for 100 sines.

Please note the decreasing prime order for the amplitudes

And the signal would be a sum of these exact 100 sines:
y=601*sin(1009x+6.28/2)+599*sin(1013x+6.28/3)+593*sin(1019x+6.28/5)+587*sin(1021x+6.28/7)+...

When plotting these 4 sines, we get this plot which shows that if we take x values in [0, 7] we get at least 2 periods of the sum of these 4 sines.



This means that to comply with the Nyquist theorem ***for these 4 sines***, it is enough to take points within the [0, 7] interval. This will not be enough for obtaining all the sines of the sum of 100 sines, but then, in reality, we get the same problem (the signal is not long enough to comply with the Nyquist theorem).

But the objective is to get AS MANY SINES AS POSSIBLE out of this signal (and not only 4)

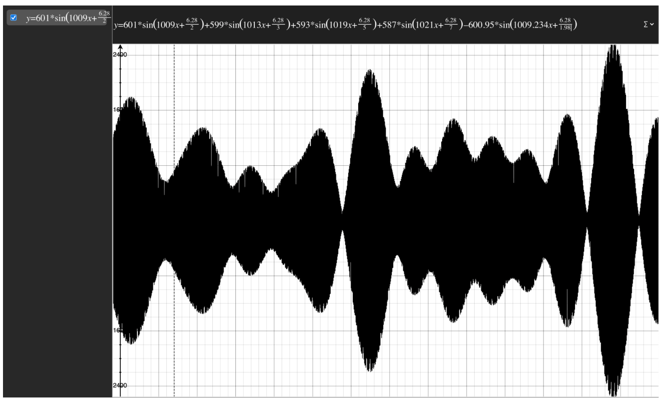
Let's say that it would be good to get the first 4 sines but we are interested in getting all of them.

So now, here is what is to be done:

A) create the signal (as already done in the program we did together) by taking 8192 points with [0, 7] over the sum of 100 sines.

B) run the program on ONE sine only for as long as necessary to find a result approaching the 1st sine only, i.e. 601*sin(1009x+6.28/2)

C) **REMOVE the sine that you obtained from the signal**.
This means that if you got 600.95*sin(1009.234x+6.28/1.98), **you must now generate a new signal** with:
601*sin(1009x+6.28/2)+599*sin(1013x+6.28/3)+593*sin(1019x+6.28/5)+587*sin(1021x+6.28/7)+...**–600.95*sin(1009.234x+6.28/1.98)**
After removing the 4 sines above, we get:



D) You now go back to step B, to find the next sine.

So if you understood well, the objective of this method is to determine if it is possible to find the first sine, then remove it, then find the second sine, then remove it, then find the 3rd sine, then remove it, and so on… and to see up to where we can go and with which precision (of course, the signal will degrade for each iteration, because the sines that will be found will not be EXACT, so by removing them inexactly, we will add some noise to the next signal.

Write a 10-20 pages report written in LaTeX, organised as a scientific paper, with your name and affiliation, an abstract, an introduction, some sections (including a state of the art section on Fourier-based harmonic analysis) and a conclusion. Don't forget to include a references section with references written correctly (name of the authors, title, journal or conference, year, page numbers, etc… so that the paper can be found again if people are interested in it).

Scientific papers are organised in the following way:

- There should be a (nice) title
- There should be an abstract (telling what the paper is about)
- An introduction to explain the context and explain what you will describe in the next sessions of the paper.
- then, using several sections, explain the methodology that you chose (protocol), **present the results, discuss the results and put them in perspective wrt the problem intially described** (has the problem been solved, and if not, what hypothesis can be proposed to explain why the protocol failed, and how the problem could have been addressed differently). If the problem is solved, point on possible improvements and explain what the results open to.
- And finally, write a conclusion.
- Don't forget a list of references (if you referenced some other papers in your paper)

Don't forget to provide the .ez file with the correct data values (including the seed) so that the results presented in your report/paper are reproducible.