Aristotle University of Thessaloniki
Department of Electrical and Computer Engineering

Advanced Signal Processing

3rd Assignment – Summer Semester 2020/2021

Kavelidis Frantzis Dimitrios – 9351

3/5/2021

Exercise 3:

VALIDITY CHECK OF GIANNAKIS' FORMULA!

Construct a real discrete signal $x[k], k = 1,2, \dots, N = 2048$, which is derived as the output of a MA-q process with coefficients of $[1.0, 0.93, 0.85, 0.72, 0.59, -0.10]$, driven by white non-Gaussian noise $v[k]$, which is derived from an exponential distribution with mean value of 1 (in Matlab, $v =$ `exprnd(1, [1,2048]);`). When you construct the signals $x[k]$ and $v[k]$ save them to use them throughout.

1. Justify the non-Gaussian character of input $v[k]$ by calculating its skewness γ_3^v using the following equation:

$$\gamma_3^v = \frac{\sum_{i=1}^{N}(v(i) - \widehat{m}_v)^3}{(N-1)\hat{\sigma}_v^3},$$

   where $\widehat{m}_v$ and $\hat{\sigma}_v$ denote the estimated from the data mean and standard deviation, respectively.

2. Estimate and plot the 3rd-order cumulants of $x[k]$, $c_3^x(\tau_1, \tau_2)$ using the indirect method with $K = 32$, $M = 64$, $L_3 = 20$ [that is $(-\tau_1{:}\,0{:}\,\tau_1) = (-20{:}\,0{:}\,20)$, $(-\tau_2{:}\,0{:}\,\tau_2) = (-20{:}\,0{:}\,20)$].

3. Use the estimated $c_3^x(\tau_1, \tau_2)$ to estimate the impulse response $\hat{h}[k]$ of the MA system using the Giannakis' formula, i.e.,

$$\hat{h}[k] = \frac{c_3^x(q,k)}{c_3^x(q,0)}, k = 0,1,2, \dots, q,$$
$$\hat{h}[k] = 0, k > q.$$

4. Estimate the impulse response of the MA system using the Giannakis' formula, yet considering:
   a. Sub-estimation of the order q, that is, $\hat{h}_{sub}[k]$: MA-q_{sub}, where $q_{sub} = q - 2$.
   b. Sup-estimation of the order q, that is, $\hat{h}_{sup}[k]$: MA-q_{sup}, where $q_{sup} = q + 3$.

5. Estimate the MA-q system output $x_{est}[k]$, using the convolution between the input $v[k]$ and the estimated impulse response from Step 3, i.e., $x_{est}[k] = v[k] * \hat{h}[k]$ and plot in the same figure the original $x[k]$ (blue color) and the estimated $x_{est}[k]$ (red color; keep only the first N samples). Find the normalized root mean square error (NRMSE) using the formula

$$NRMSE = \frac{RMSE}{\max(x[k]) - \min(x[k])}, \text{with}$$

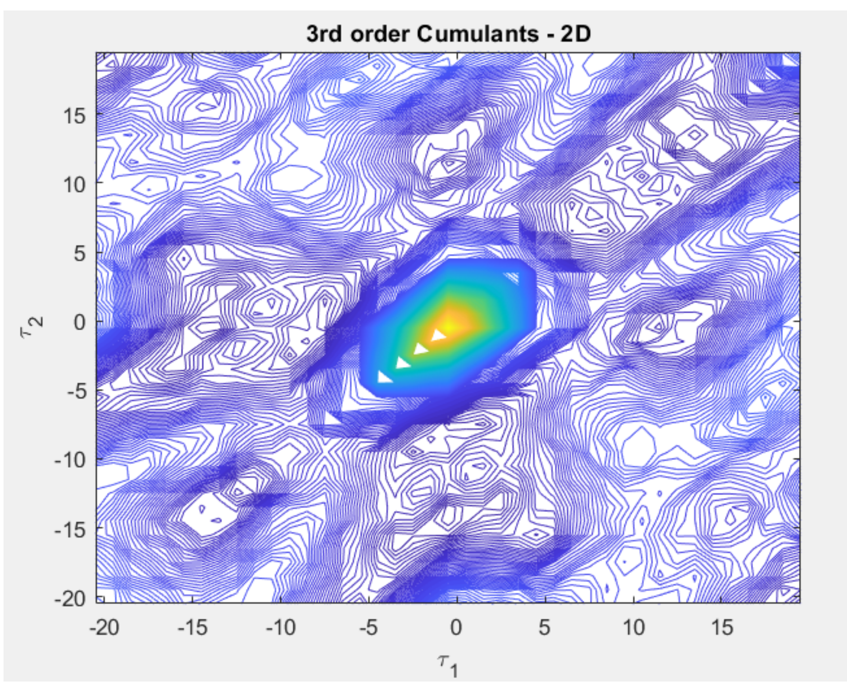
$$RMSE = \sqrt{\frac{\sum_{k=1}^{N}(x_{est}[k] - x[k])^2}{N}}$$

   Comment upon the findings.

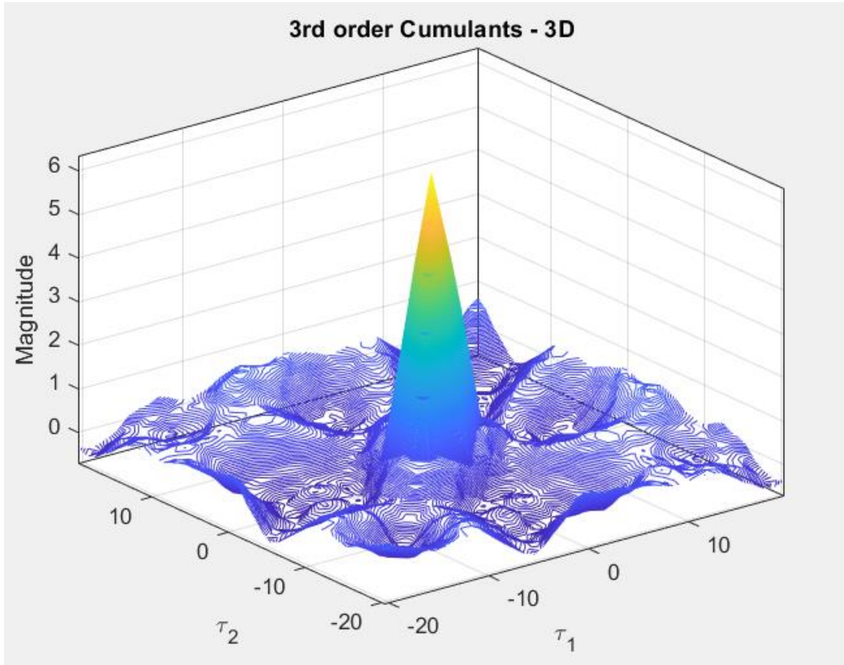
6. Repeat Step 5, for the case of $x_{sub}[k] = v[k] * \hat{h}_{sub}[k]$ and $x_{sup}[k] = v[k] * \hat{h}_{sup}[k]$.
   Comment upon the findings and compare with the results of Step 5.
7. Consider that we add a noise source of white Gaussian noise at the output of the system, producing a variation in the signal-to-noise-ratio (SNR) of $[30:-5:-5]$dB, i.e., $y_i[k] = x[k] + n_i[k], i = 1:8$. Repeat Steps 2, 3 and 5, but instead of $x[k]$ use the noise contaminated output $y_i[k]$ for each ith given level of SNR (you can easily create the contaminate signal by using the awgn.m function of Matlab (`y = awgn(x,snr,'measured')`) and simply changing each time the given SNR). Make a plot of the NRMSE error in the estimation of $y_i[k]$ versus the SNR range. Comment upon your results.
8. (optional) Instead of using just one realization of the input and output data of MA-q system, you could repeat the whole process 50-100 times and work with the mean values of your results (that is mean NRMSE) to increase the viability and generalization of your conclusions about the Giannakis' formula.

Analysis:

1. The skewness of the input $\boldsymbol{v[k]}$ calculated is different than zero, thus the non-Gaussian character of $\boldsymbol{v[k]}$ is justified.
2. Plots of the 3rd order cumulants of $\boldsymbol{x[k]}$:



Plot 1 - 3rd Order Cumulants of x[k] – 2D



Plot 2 - 3rd Order Cumulants of x[k] – 3D

3. Estimated $\boldsymbol{\hat{h}[k]}$ =

   1.0000 1.0560 1.2088 0.7936 0.4263 -0.0285

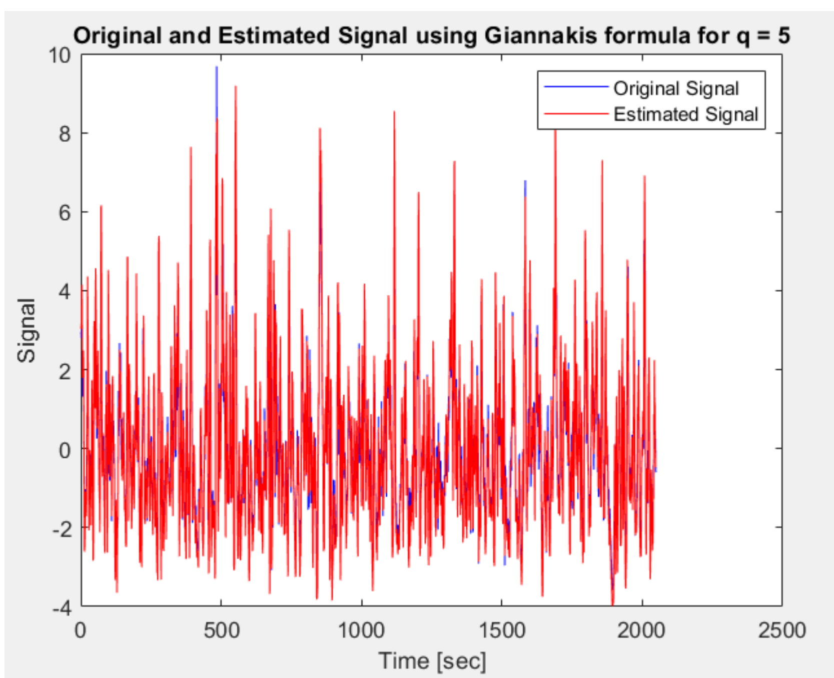
4. Estimated $\boldsymbol{\hat{h}_{sub}[k]}$ =

   1.0000 0.9228 0.8291 0.7360

   Estimated $\boldsymbol{\hat{h}_{sup}[k]}$ =

   1.0000 0.5529 0.7547 1.2591 1.0418 1.9877 2.7711 3.1355 2.8666

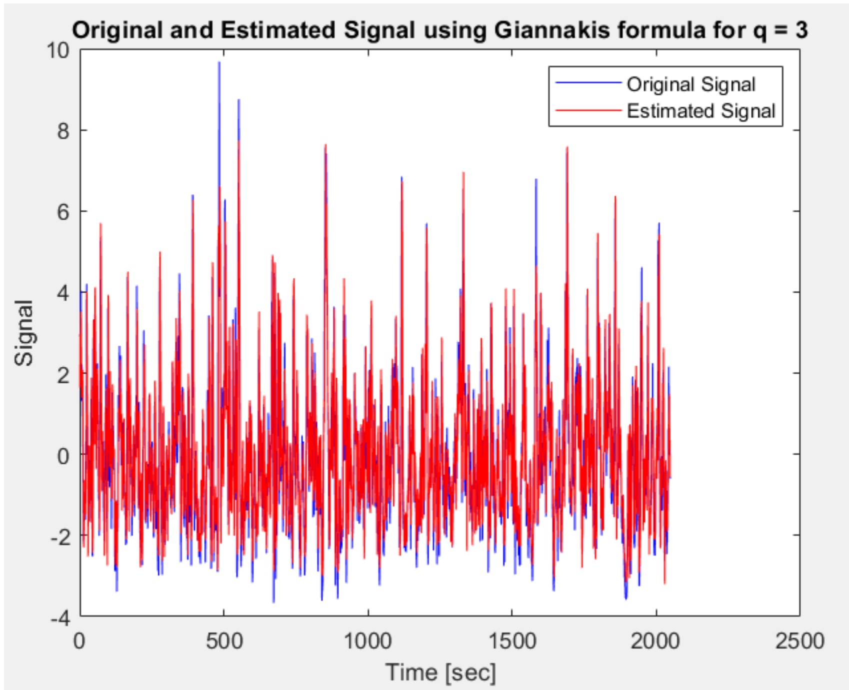
5. Plot of original signal and estimate of the signal using the estimated $\boldsymbol{\hat{h}[k]}$:



Plot 3 – Comparison of original and estimated signal / q = 5

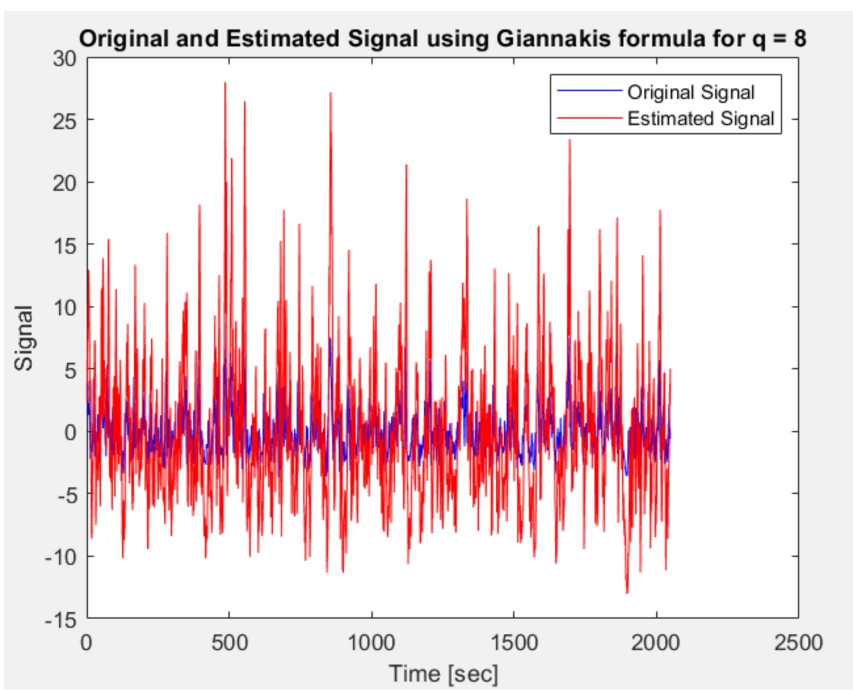
This estimation led to $\boldsymbol{NRMSE} = \mathbf{0.0313}$. We can observe that using the correct order for the MA-q system is giving a low $\boldsymbol{NRMSE}$ even though the $\widehat{\boldsymbol{h}}[\boldsymbol{k}]$ seems to differ from the given coefficients of the generated process.

6. By repeating step 5 for sub estimation and sup estimation, we get:



Plot 4 – Comparison of original and estimated signal / q = 3

This estimation led to $\boldsymbol{NRMSE} = \mathbf{0.0740}$.

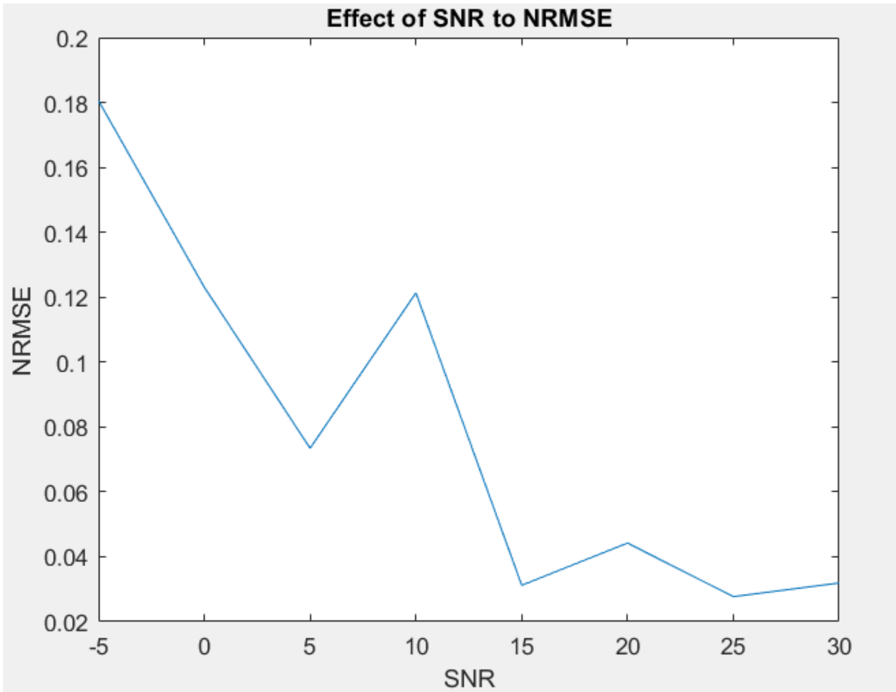


Plot 5 – Comparison of original and estimated signal / q = 8

This estimation led to $\boldsymbol{NRMSE} = \mathbf{0.4064}$.

As we expect, using different order for q than the one that generated the process, results to a significant difference in the NRMSE, especially in the sup estimation.

7. In this step, the process is repeated 8 times for the same signal with white Gaussian noise added on the output with different SNR each time. Then, the plot of NRMSE vs SNR range is generated to show how estimation with Giannakis' formula is robust to white Gaussian noise:



Plot 6 – NRMSE vs SNR / Effect of different SNR to the estimation

Assuming that Giannakis' formula is always true, the expected behavior of Plot 6 would be to get a low constant NRMSE value, meaning that we expect that the NRMSE does not change when we add more noise to the output. However, as we can see, this is not the behavior we get from the system. The estimator of the 3rd order cumulants

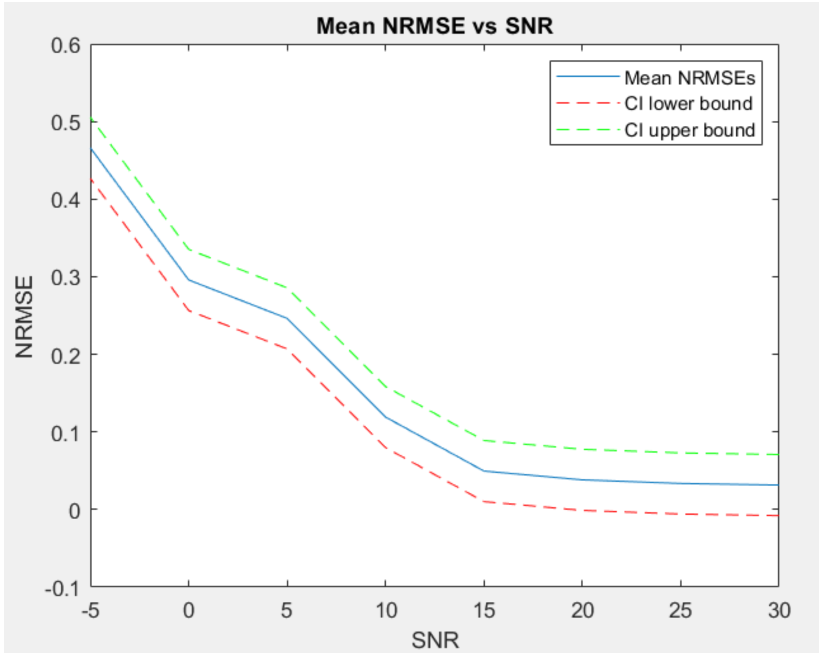
$$\hat{c}_3(q,k) = \frac{1}{N}\sum_{i=0}^{N} y(i)\, y(i+q) y(i+k), \qquad (1)$$

and thus, the estimator of IR process, are consistent [1],[2]. Also, when symmetrically distributed noise like the white Gaussian noise is added to the output, the cumulants remain unaffected [3].

Thus, the only error introduced to the system is because of the finite samples used when estimating the cumulants (1) [4].

Therefore, with finite samples, it is expected that the lower the SNR, the bigger the NRMSE that we are going to get, meaning a descending behavior.

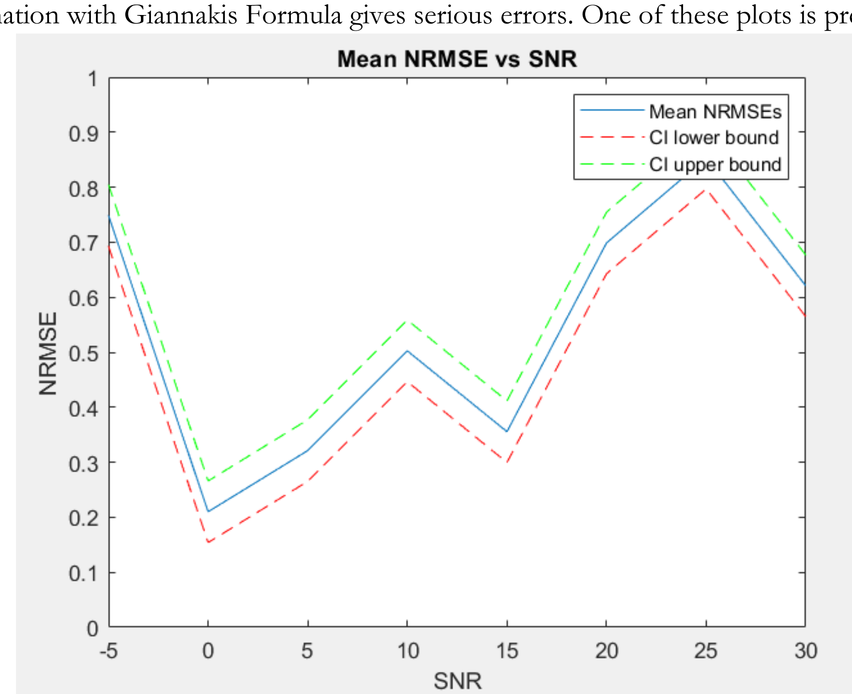
8. Apparently, even the plot 6 has some inconsistency because of the stochastic property of the noise. To have a more general and reliable conclusion, we repeat the process 50 times (50 realizations of $\boldsymbol{y}[\boldsymbol{k}]$) and estimate a "mean behavior" of the system's NRMSE. The plot generated accordingly with step 7:



Plot 7 – NRMSE vs SNR / Effect of different SNR to the estimation / Mean NRMSE taking 50 realization/ Simulation in 'Giannakis Favor'.

Here the expected behavior is clearly closer to the expected as mentioned in step 7, and additionally, the robustness of the estimation using Giannakis' formula seems to appear after SNR reaches a threshold on which the finite samples start to be enough for the estimation of cumulants to give a robust to white Gaussian noise system.

Note: The plots 6 and 7 are expected to differ when running the code each time, meaning that even the mean NRMSEs generated are not always consistent. The behavior presented is simply one generation of the 50-realizations implementation mentioned above with 2048 samples. However, as we run the simulation many times, we see that in a significant number of tries, the results show that the estimation with Giannakis Formula gives serious errors. One of these plots is presented below:

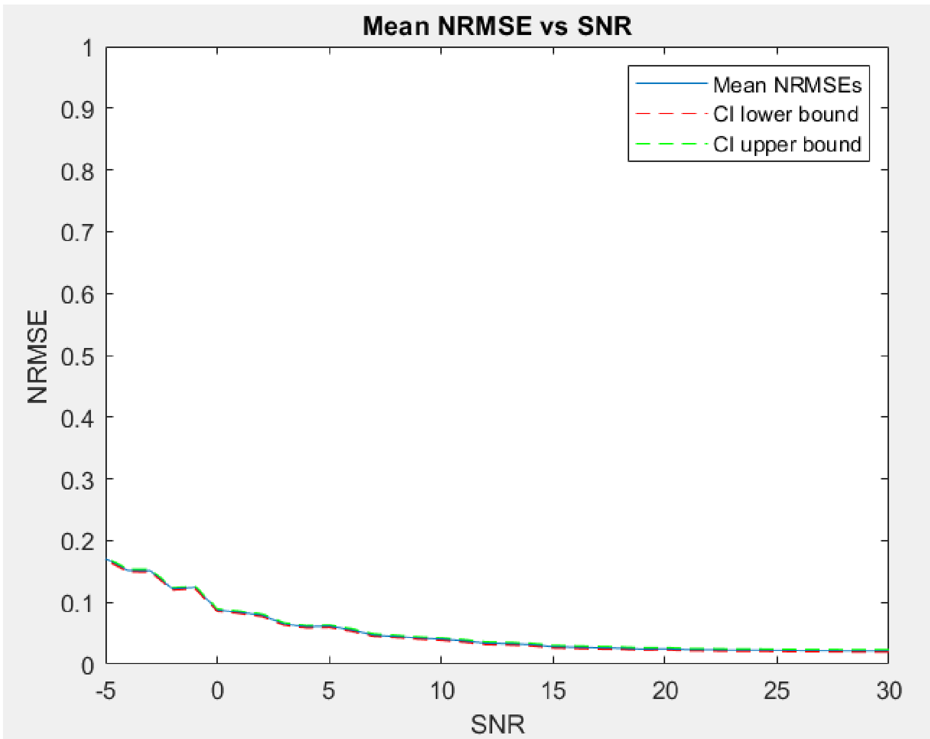


Plot 8 – NRMSE vs SNR / Effect of different SNR to the estimation / Mean NRMSE taking 50 realization/ Simulation in contrast with Giannakis Formula.

There have been even some extreme cases/results on some of the simulations, giving NRMSEs way higher than 1, or simulations that we took a better estimation in step 7 than on the same test on step 8.

Conclusion:

We have to consider the fact that the number of samples on this exercise (2048) as well as the range of SNR are significantly small and therefore the results produced can be easily misinterpreted. To get a better understanding and intuition behind Giannakis' Formula validity and effectiveness, a sufficient number of tests must be done with enough samples to give a more reliable estimation of the cumulants (Monte Carlo). Below we can see a plot for 50 realizations with 131072 samples each and a range of SNR = [30: -1: -5].



Plot 9 – NRMSE vs SNR / Effect of different SNR to the estimation / Mean NRMSE taking 50 realization/ SNR [30: -1: -5] / 131072 samples each time.

As it can be seen, with a better estimation of cumulants, the system is far more reliable and robust to noise. Although, we still have some small, unexpected peaks before 0, showing us that even now with a larger number of samples, some errors are still carried on from the cumulants estimation. Thus, Giannakis' Formula can be considered an effective theoretical formula in the optimal occasion of infinite samples. However, this makes his "Closed Form solution" to the problem of parameter estimation of an MA-system not particularly useful as practical estimation procedure, as it does not smooth out the effect of errors in the estimation of cumulants **[5]**.

References:

[1] K. S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes," Ann. Statist., vol. 10, pp. 1195-1208,1982

[2] G. Giannakis, "Signal processing via higher-order statistics," Ph.D. dissertation, University of Southern California, Los Angeles, CA, July 1986.

[3] A.W. Lohman and B. Wirnitzer, "Triple correlations" *Proc. IEEE,* vol. 72, pp.889-901, 1984.

[4] G. Giannakis, "Cumulants: a powerful tool in signal processing," *Proc. IEEE*, vol. 75, no. 9, pp. 1333-1334, 1987.

[5] A.G. Stogioglou, "MA Parameter Estimation Using Higher-Order Cumulant Statistics" (1996)