Advanced signal processing techniques
Third assignment

Theodoros Katzalis
AEM: 9282
katzalis@auth.gr

Aristotle University of Thessaloniki

May 7, 2021

Description

VALIDITY CHECK OF GIANNAKIS' FORMULA!

Construct a real discrete signal $x[k], k = 1,2,\ldots, N = 2048$, 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:
$$\gamma_3^v = \frac{\sum_{i=1}^{N}(v(i) - \hat{m}_v)^3}{(N-1)\hat{\sigma}_v^3},$$
where $\hat{m}_v$ 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.,
$$\hat{h}[k] = \frac{c_3^x(q,k)}{c_3^x(q,0)}, k = 0,1,2,\ldots,q,$$
$$\hat{h}[k] = 0, k > q.$$
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.

Solutions

The Matlab code for the following can be found here.

2

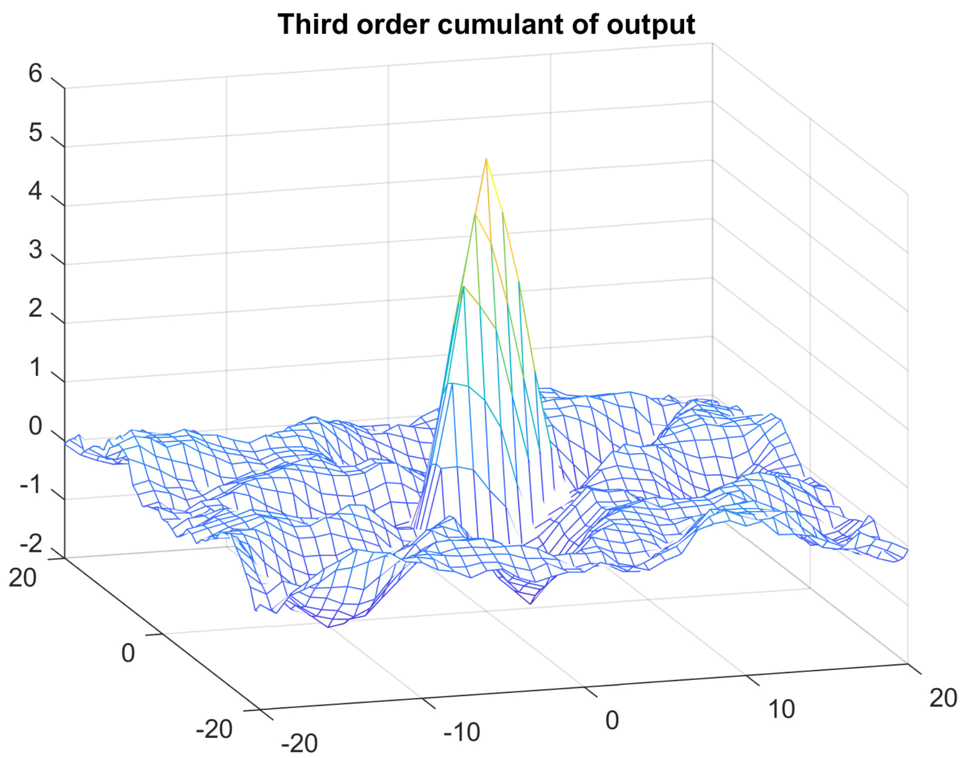


Figure 1

3, 4, 5, 6

The impulse response and the NRMSE estimations for the accurate order of the model (q=5) and the sub (q=3) and sup (q=7) variants.

$$h = \begin{bmatrix} 1.0000 & 0.9300 & 0.8500 & 0.7200 & 0.5900 & -0.1000] \end{bmatrix} \tag{1}$$
$$\hat{h}_5 = \begin{bmatrix} 1.0000 & -0.4317 & -0.7550 & 0.1845 & 0.5083 & 0.3519 \end{bmatrix} \tag{2}$$
$$\hat{h}_3 = \begin{bmatrix} 1.0000 & 0.4310 & 0.8275 & 2.2743 \end{bmatrix} \tag{3}$$
$$\hat{h}_8 = \begin{bmatrix} 1.0000 & 0.5724 & 0.1242 & -0.2649 & -0.4322 & -0.3553 & -0.2517 & 0.0135 & 0.0074 \end{bmatrix} \tag{4}$$

$$NRMSE_5 = 0.1901 \tag{5}$$
$$NRMSE_3 = 0.1913 \tag{6}$$
$$NRMSE_8 = 0.2142 \tag{7}$$

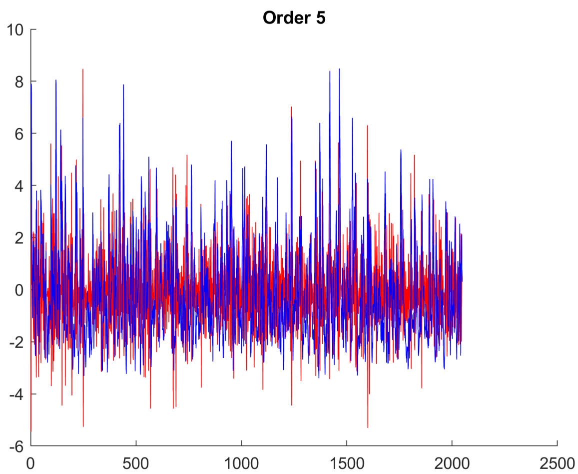
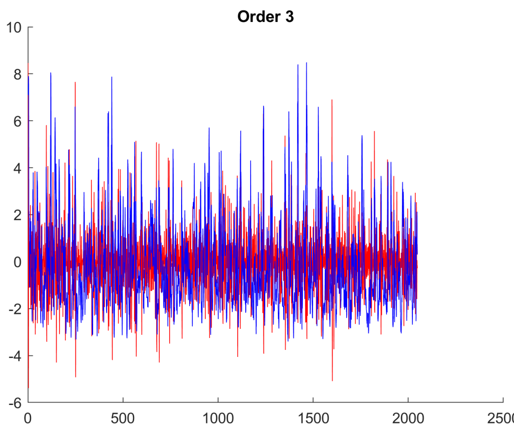
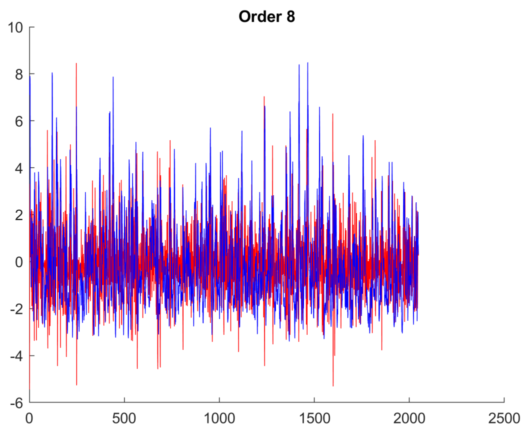


Figure 2: Correct order



(a) Sub order



(b) Sup order

Figure 3

7

The following error measurements are based on the accurate order of the MA model (q=5).

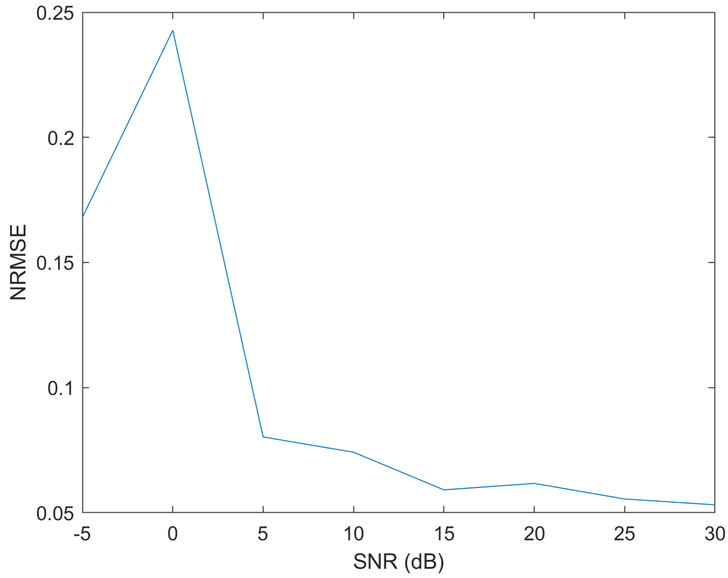
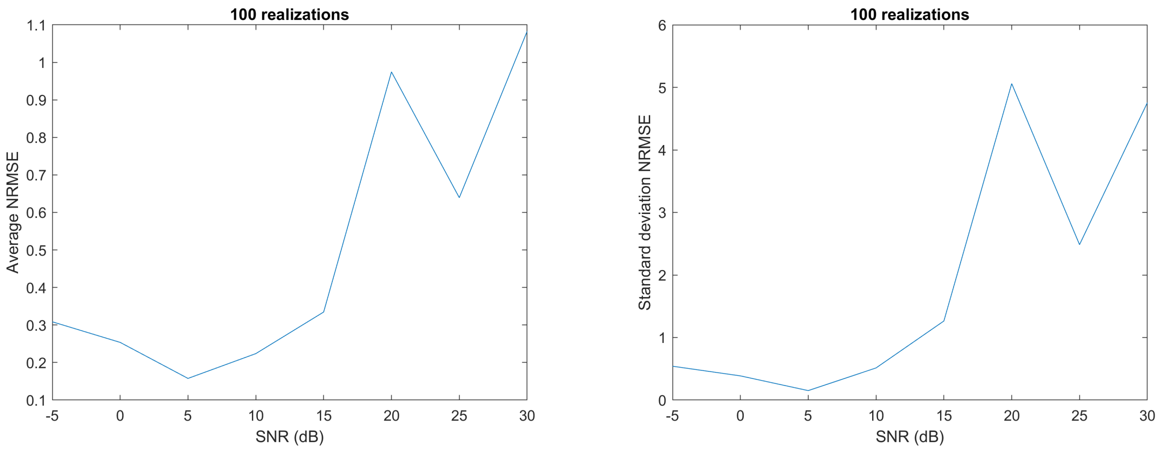


Figure 4: NRMSE - SNR for 1 realization

8



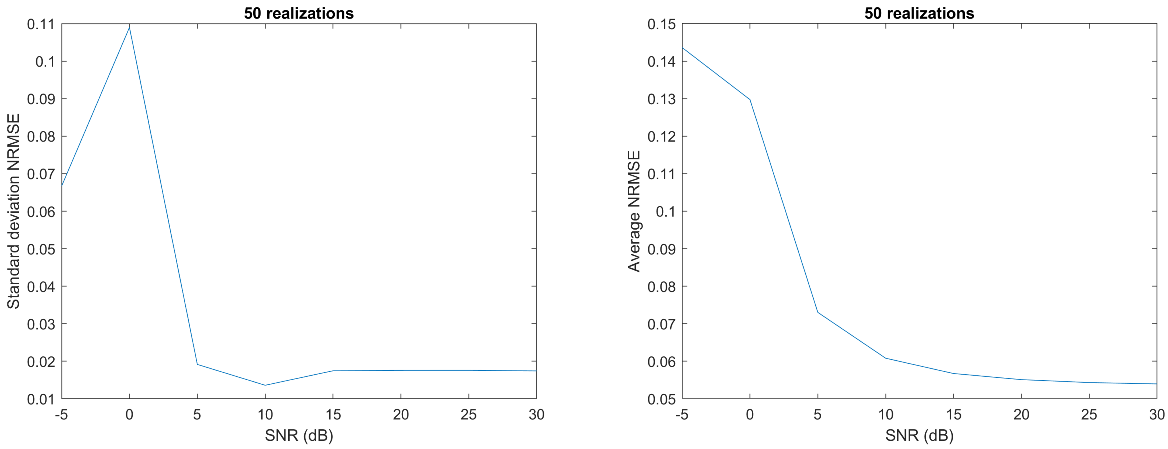
(a) NRMSE - SNR for 100 realizations

(b) Standard deviation

Figure 5

Conclusions

- There is significant variability across realizations for the output estimation and the NRMSE. We could claim that the culprit is the number of samples. Increasing the number of them, we can claim a much lower variability and more accurate estimations. For proof of concept we represent the following two plots, increasing the number of samples from $2^{11} \rightarrow 2^{15}$:



(a) NRMSE - SNR for 10 realizations

(b) Standard deviation

Figure 6

- Theoretically about NRMSE and SNR, since third order cumulants are unaffected with the addition of gaussian noise, we should observe a straight line. However, this is not true and we can see some spikes. Of course, a zoom out version of these plots will represent a straight line. The reason, why practically a straight line isn't observed is due to the limitations of the cumulant calculation. Increasing the number of samples or averaging cumulant of higher orders could give also better estimations and noise canceling.

- Averaging the estimation didn't improve the NRMSE-SNR plot. In fact, it gets worse. The reason is the significant variability within estimations. In other words, for the 7) question, the NRMS-SNR plot just happened to have a low variability but with averaging the values across multiple iterations we have a broad overview of the process. Thus, our estimator with the current configuration exhibits high variability between realizations.

- Sub and sup estimations should perform worse than the accurate order choice. Again this could be spotted clearly for a larger number of samples.

- Estimation of impulse response compared to the original one isn't quite accurate and NRMSE isn't small enough. Overall, if our estimations are based on one realization with a small number of samples, then they won't have acceptable accuracy.