Miniproject 3

Kai Ye ye000064@umn.edu

December/18/2020

Task 1

Write down an expression for the joint density of the elements of the observed mixtures. Specifically, use analogous shorthand notation to the above, letting $\chi := \{x_m(n)_{1\leq m\leq M, 1\leq n\leq N}\}$ denote the MN random variables comprising your observations, and find an expression for $p(\chi)$.

Solution: The data model can be expressed as $\boldsymbol{x}(n) = \boldsymbol{A}\boldsymbol{s}(n)$.

Based on our knowledge from the course materials, the density of each linear transform vector is given by

$$p_x(\boldsymbol{x}(n)) = |det(A^{-1})|p_s(\boldsymbol{s}(n)) = |det(A^{-1})|\prod_{m=1}^{M} p_{s_m}(\boldsymbol{s}_m(n)) = |det(A^{-1})|\prod_{m=1}^{M} p_{s_m}(\boldsymbol{s}_m(n))$$

The total joint density is a product of over the n values:

$$\begin{aligned} p(\chi) = \prod_{n=1}^{N} p_x(\boldsymbol{x}(n)) &= \prod_{n=1}^{N} |det(A^{-1})| \prod_{m=1}^{M} p_{s_m}(\boldsymbol{s}_m(n)) \\ &= |det(A^{-1})|^N \prod_{n=1}^{N}\prod_{m=1}^{M} p_m(A_{m1}^{-1}x_1(n) + A_{m2}^{-1}x_2(n) + ... + A_{mM}^{-1}x_M(n)) \end{aligned}$$

Task2

Calculate an explicit expression for the gradient of the log-likelihood function with respect to the unknown matrix W.

Solution: Given that $W = A^{-1}$, the expression above can be rewritten as:

$$p(\boldsymbol{\chi}; \boldsymbol{W}) = |det(\boldsymbol{W})|^N \prod_{n=1}^{N} \prod_{m=1}^{M} p_m(\boldsymbol{W}_{m1}x_1(n) + \boldsymbol{W}_{m2}x_2(n) + ... + \boldsymbol{W}_{mM}x_M(n))$$

$$\begin{aligned}\log p(\boldsymbol{\chi}; \boldsymbol{W}) &= N \log |det(\boldsymbol{W})| + \log \prod_{n=1}^{N} \prod_{m=1}^{M} p_m(\boldsymbol{W}_{m1}x_1(n) + \boldsymbol{W}_{m2}x_2(n) + ... + \boldsymbol{W}_{mM}x_M(n)) \\ &= N \log |det(\boldsymbol{W})| + \sum_{n=1}^{N} \sum_{m=1}^{M} \log p_m(\boldsymbol{W}_{m1}x_1(n) + \boldsymbol{W}_{m2}x_2(n) + ... + \boldsymbol{W}_{mM}x_M(n))\end{aligned}$$

With the help from the Matrix Cookbook:

$$\nabla \log p(\boldsymbol{\chi}; \boldsymbol{W}) = \frac{\partial \log p(\boldsymbol{\chi}; \boldsymbol{W})}{\partial \boldsymbol{W}} = N(\boldsymbol{W}^{-1})^T + \frac{\partial}{\partial \boldsymbol{W}} \sum_{n=1}^{N} \sum_{m=1}^{M} \log p_m(\boldsymbol{W}_{m1}x_1(n) + \boldsymbol{W}_{m2}x_2(n) + ... + \boldsymbol{W}_{mM}x_M(n))$$

whose elements are given by:

$$[\nabla \log p(\boldsymbol{\chi}; \boldsymbol{W})]_{ij} = N[(\boldsymbol{W}^{-1})^T]_{ij} + \frac{\partial}{\partial \boldsymbol{W}_{ij}} \sum_{n=1}^{N} \sum_{m=1}^{M} \log p_m(\boldsymbol{W}_{m1}x_1(n) + \boldsymbol{W}_{m2}x_2(n) + ... + \boldsymbol{W}_{mM}x_M(n)$$

Task 3

Your task is to code the gradient ascent method outlined above in order to obtain an estimate of the (invertible) mixing matrix, from which you can ultimately estimate the original sources themselves. Ultimately, you should submit your code as well as plots of your estimates of the original sources.

Solution: The gradient ascent procedure can be represented as: $\boldsymbol{W}^{i+1} = \boldsymbol{W}^i + \alpha(i) \nabla \log p(\boldsymbol{\chi}; \boldsymbol{W})$. This complete gradient ascent method is implemented in the MATLAB code attached. I found the original signal from course website and they are shown in the same figure with the estimation I got. The priori I chose is that the sources should be super-gaussian to keep the gradient ascending ($\nabla \log p(\boldsymbol{\chi}; \boldsymbol{W}) > 0$).

I used a constant step of 10^{-5}, and it takes 5922649 iterations to converge with the criteria that $|det(gradient)| < 10^{-5}$. Maybe we can use a large step size. The final results are plotted below in Figure 1, which match very well with the original ones.

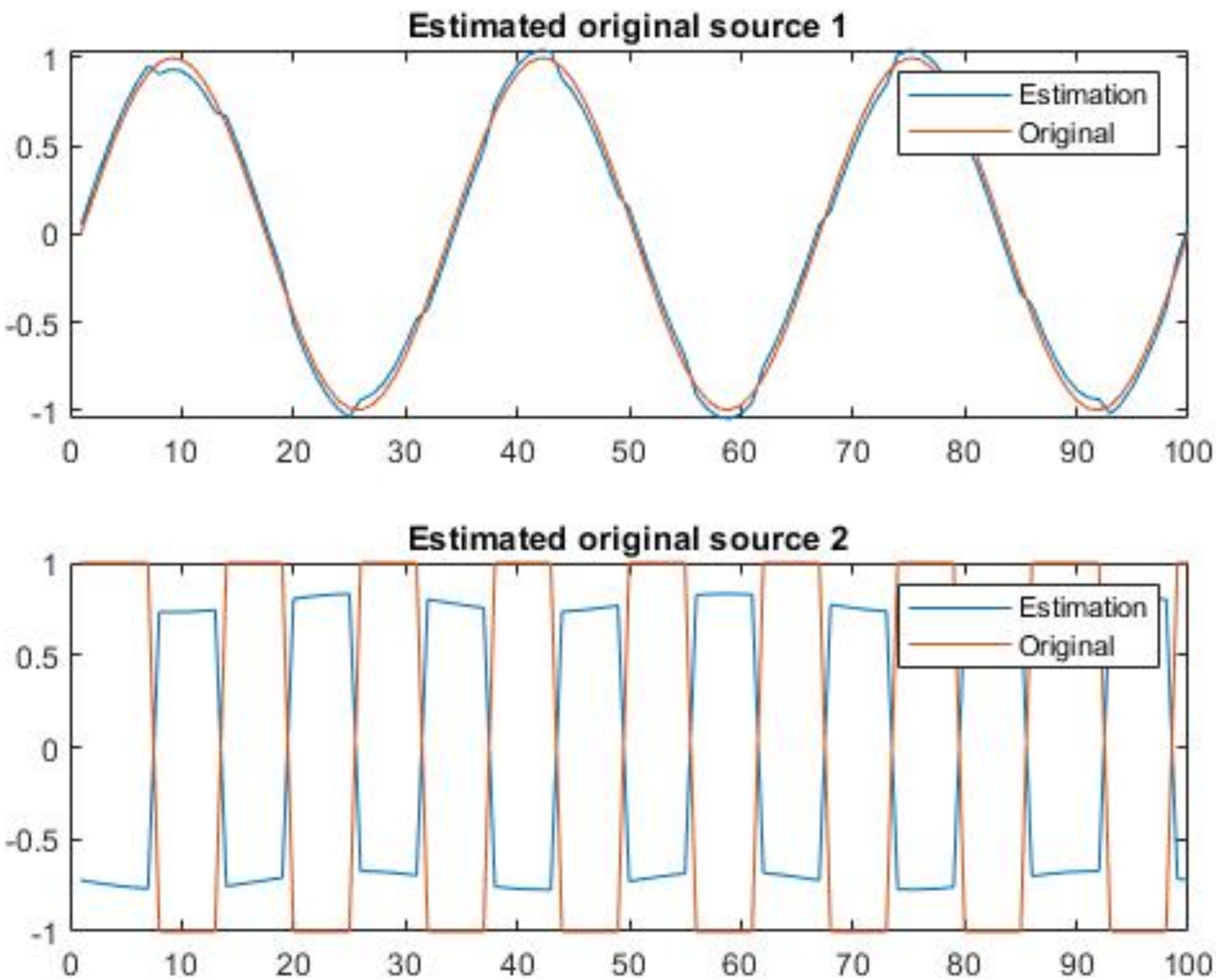

Figure 1: Plots of my estimation compared to the original sources