# 02471 Machine Learning for Signal Processing

# Problem set 3

This problem set is based on the teaching material and exercises from weeks 7–13.

You are free to use any code that has been handed out during the course. Try to work on these problems using your own code though, as the code you write in this homework will make your life easier down the road.

Hand in **one pdf** with your answers, **including your code**. Your solution will be graded and count 20% towards the final course grade. Late submission will be considered as 0% completed. All problems are weighted equally.

It is fine to work in small groups (2–3 people), in that case, when you hand, clearly state the group composition. All answers must be submitted individually and all people are accountable for the entire solution.

## Loading data

The data is stored in matlab files, and can be loaded in python using scipy loadmat, see here: https://docs.scipy.org/doc/scipy/reference/tutorial/io.html

## Problem 3.1 Sparse signal representations

In this problem we consider the compressed sensing scenario. Assume that you are informed that the signal you are measuring follows the form:

$$x_n = \sum_{j=1}^{K} a_j \cos\left(\frac{\pi}{2l}(2m_j - 1)n\right), \quad n = 0, ..., l-1.$$

You have the following information available:

- The total length of the original signal was $l = 2^9$, but you only obtained $2^5$ samples.
- The signal is sparse for the DCT of $x_n$ ($K \ll l$).
- You are informed of the index number of the recovered samples. Both the recovered samples indices and sample values are found as $n$ and $x$ in the `problem3_1.mat` file, where $n$ is the sample indices and $x$ is the sample values $x_n$.

### Problem 3.1.1

Describe how you can recover the original signal using a sparse analysis model. Explicitly back up your reasoning by referring to signal representation formulas.

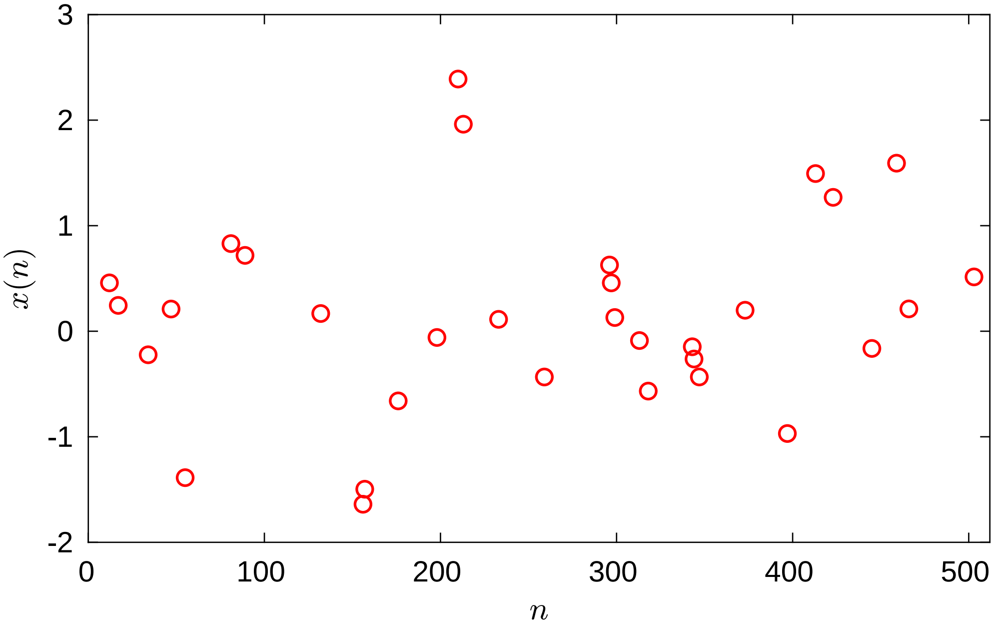


Figure 1: Problem 3.2.

**Problem 3.1.2**

This exercise is mostly a programming exercise.

Load the data from the given file (`problem3_1.mat`). The data contains two vectors $n$ and $x$, which are also plotted on Figure 1, page 2.

Use an appropriate sparsity promoting algorithm (argue why it is appropriate) and estimate the following parameters: $K$, $a_j$ and $m_j$. Submit your code as well, and be sure to write comments on what operations the lines carry out. Write out the found $K$, $a_j$ and $m_j$ explicitly with 2 digits.

## Problem 3.2 Bayesian inference and the EM algorithm

For this problem, we consider the regression problem,

$$y_n = \boldsymbol{\theta}^T \boldsymbol{x}_n + \eta_n,$$

where $\eta_n$ is independently and identically distributed (i.d.d) with a Gaussian distribution $\eta_n \sim \mathcal{N}(0, \sigma_\eta^2)$. We additionally assume that the elements in the $\boldsymbol{\theta}$ vector are i.i.d. with a Gaussian distribution $\theta_i \sim \mathcal{N}(0, \sigma_\theta^2)$.

The parameters $\boldsymbol{\theta}$ can be found by solving the optimization problem,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_\eta^2} \|\boldsymbol{y} - X\boldsymbol{\theta}\|_2^2 + \frac{1}{2\sigma_\theta^2} \|\boldsymbol{\theta}\|_2^2,$$

where $\boldsymbol{y} = [y_1, y_2, \cdots, y_N]^T$ and $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N]$.

**Problem 3.2.1**

Identify what type of problem the optimization problem solves. Describe the role of the two terms $\frac{1}{2\sigma_\eta^2}\|\boldsymbol{y} - X\boldsymbol{\theta}\|_2^2$ and $\frac{1}{2\sigma_\theta^2}\|\boldsymbol{\theta}\|_2^2$ in terms of likelihood, prior, regularization and mean squared error. How will the values of the parameters $\sigma_\eta$ and $\sigma_\theta$ influence the optimization problem?

**Problem 3.2.2**

In the question above, we assume that $\sigma_\eta^2$ is known. We now wish to estimate it in a Bayesian manner. To this end, we introduce an inverse gamma prior on $\sigma_\eta^2$. The density function of an inverse gamma distributed random variable z is,

$$p(\mathrm{z}|a, b) = \frac{b^a}{\Gamma(a)}\mathrm{z}^{-a-1}\exp\left(-\frac{b}{\mathrm{z}}\right),$$

where $a > 0, b > 0$ are hyperparameters for the distribution, and $\Gamma(\cdot)$ is the gamma-function, defined as an integral,

$$\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}dy$$

Find an expression for the log of the density function $\ln p(\mathrm{z}|a, b)$. Omit the additive terms that are not dependent on z in your solution (the constants).

**Problem 3.2.3**

We will now introduce the prior on $\sigma_\eta^2$ in the loss function. Using Bayes theorem, show that ($X$ is considered fixed and is therefore omitted) the following posterior is true.

$$p(\boldsymbol{\theta}, \sigma_\eta^2|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\sigma_\eta^2, \boldsymbol{\theta})p(\sigma_\eta^2)p(\boldsymbol{\theta})}{p(\boldsymbol{y})}.$$

Write out the log-posterior $\ln p(\boldsymbol{\theta}, \sigma_\eta^2|\boldsymbol{y})$ by inserting the corresponding distributions for $p(\boldsymbol{y}|\sigma_\eta^2, \boldsymbol{\theta})$, $p(\sigma_\eta^2)$ and $p(\boldsymbol{\theta})$.

## Problem 3.3 Estimation of ICA solution

We will work with the ICA model, written as

$$\mathbf{x} = A\mathbf{s}$$

where we have observed $N$ realizations of $\mathbf{x}$ ($\mathbf{x}$ is a random vector). We want to recover the matrix A used to mix the sources $\mathbf{s}$. Assume that $A$ is a square invertible matrix so that we can write $\mathbf{z} = W\mathbf{x} = WA\mathbf{s}$, where $W = A^{-1}$ when perfect recovery is obtained. For the entirety of this problem, have the matrix $A = [\boldsymbol{a}_1 \;\; \boldsymbol{a}_2]$, $\boldsymbol{a}_1 = [3 \;\; 1]^T$ and $\boldsymbol{a}_2 = [1 \;\; 1]^T$.

### Problem 3.3.1

Consider the case where $\mathbf{s}$ is drawn from a uniform distribution $U(0, 1)$.

Create a program that estimates $A$ (denoted $\hat{A}$) based on ICA (mutual information as we did in the exercise).

Create a function (in code) that calculates the error between $\hat{A}$ and $A$. The function should take into account that ICA can neither estimate the magnitude of the vectors in $A$, nor the ordering of the basis vectors (e.g. $\boldsymbol{a}_1$ may be estimated as $\hat{\boldsymbol{a}}_2$). Validate that you get a "low error" in this setup.

Make plots of the following: the sources, the observations, the estimated sources. Also plot/compute the error of $\hat{A}$ based on 100 repeated experiments.

### Problem 3.3.2

Repeat the above exercise for the following cases

- $\mathrm{s}_1$ is drawn from $U(0, 1)$, and $\mathrm{s}_2$ is drawn from a beta distribution $B(0.1, 0.1)$.
- $\mathrm{s}_1$ is drawn from $U(0, 1)$, and $\mathrm{s}_2$ is drawn from a normal distribution $N(0, 1)$.
- $\mathbf{s}$ is drawn from a multivariate normal distribution with $\mu = (0, 1)$, $\Sigma = [2 \;\; 0.25; 0.25 \;\; 1]$.

In which cases are you able to successfully estimate $A$ ? In the cases where you cannot estimate $A$ reliably, provide mathematical arguments on why that is the case.

## Problem 3.4 Hidden Markov Models

Consider the following problem description. We have a system with two internal states, denoted $s_1$ and $s_2$ and three possible activities, denoted $a_1$, $a_2$ and $a_3$. We are not able to measure the internal states, but only estimate the internal state through observing the ongoing activity. The system is modeled using a Hidden Markov Model.

### Problem 3.4.1

We observe the activities at a fixed time interval. We are informed of the following

- The probabilities for the system to start in state $s_2$ is 40%.
- During time-intervals, the probability of the system to change from state $s_1$ to $s_2$ is 10%, and the probability of the system staying in state $s_2$ is 65%.
- The probability of activity $a_1$ happening when the system is in state $s_1$ is 60% and in state $s_2$ 10% respectively. Additionally, the probability of activity $a_3$ occurring when the system is in state $s_2$ is 30%, and, finally, the probability of activity $a_3$ happening when the system is in state $s_1$ is 10%.

Determine the appropriate model parameters.

**Problem 3.4.2**

We observe the sequence $\boldsymbol{y} = \{a_1, a_2, a_1\}$. Apply the filtering recursion and write out the formulas for $\alpha(\boldsymbol{x}_1)$ and $\alpha(\boldsymbol{x}_2)$ (the filtering recursion coefficients), and calculate them numerically. Give all numerical results with 3 digits. Calculate $P(x_2 = s_1 | y_{1:2})$, where $y_{1:2}$ means $\{y_1, y_2\}$. Include your calculations and intermediate steps.

**Problem 3.4.3**

Assume the same conditions as the previous question. Derive an expression for $P(y_3|x_2)$ using marginalization and calculate the numerical values for $P(y_3 = a_1|x_2 = s_1)$ and $P(y_3 = a_1|x_2 = s_2)$ and give the results with 3 digits. Include your intermediate steps.

**Problem 3.4.4**

Use the previous calculations to calculate $P(x_2 = s_1|y_{1:3})$. Give the numerical results with 3 digits. Include your intermediate steps.

## Problem 3.5 Kalman Filter

In this problem we consider the Kalman filtering equations and the state-space model:

$$\begin{aligned}\mathbf{x}_n &= F_n \mathbf{x}_{n-1} + \boldsymbol{\eta}_n \\ \mathbf{y}_n &= H_n \mathbf{x}_n + \mathbf{v}_n\end{aligned}$$

**Problem 3.5.1**

Consider the movement of an object in one-dimensional space, e.g a train moving. Setup a Kalman filtering model that uses position and velocity of the object as states but only observes the location. Create 10 data-points and assess your model. Show plots of the tracking performance, i.e the estimated states, versus the true states and observed states.

## Problem 3.6 Kernel methods

In this problem we will consider the smoothing problem where we estimate the underlying signal. The signal of interest is a chirp signal based on the sinc function, and we need to determine the SNR of the observed signal, and the time-delay of the chirp.

### Problem 3.6.1

Load the signal in `problem3_6.mat`, where `t` is the time-indices and `y` is the sampled signal. Plot the signal, and determine the center location of the chirp (the center of the sinc function).

### Problem 3.6.2

Use Kernel ridge regression with a Gaussian kernel, and choose appropriate kernel parameters.

Calculate the SNR ratio, assuming that the measurement noise is white noise. Determine the center location of the chirp as estimated by Kernel ridge regression.

### Problem 3.6.3

Use support vector regression with a Gaussian kernel, and choose appropriate kernel parameters. Consider again the signal from the previous problem. Use support vector regression (SVR) to determine the underlying signal. Choose appropriate parameters for the kernel, and choose appropriate values for $C$ and $\epsilon$ for SVR.

Describe how the outliers can be identified based on the output from SVR (create a simple rule that can be implemented in code)

Calculate the SNR ratio, assuming that the measurement noise is white noise, and discount the outliers in the SNR calculation. Determine the center location of the chirp as estimated by Kernel ridge regression.