

[FMAN45] - Assignment1

Filip Kalkan (fi231ka-s)

April 2022

1 Task 1

We begin by differentiating the expression in

$$\underset{\omega_i}{\text{minimize}} \frac{1}{2} \|r_i - x_i \omega_i\|_2^2 + \lambda |\omega_i| \quad (1)$$

with respect to ω_i .

$$\begin{aligned} & \frac{d}{d\omega_i} \left(\frac{1}{2} \|r_i - x_i \omega_i\|_2^2 + \lambda |\omega_i| \right) \\ &= \frac{1}{2} \left(2(r_i - x_i \omega_i)(-x_i) \right) + \lambda \frac{\omega_i}{|\omega_i|} \\ &= (r_i - x_i \omega_i)(-x_i) + \lambda \frac{\omega_i}{|\omega_i|} \\ &= -x_i^T (r_i - x_i \omega_i) + \lambda \frac{\omega_i}{|\omega_i|} \end{aligned} \quad (2)$$

(1) is equivalent to setting (2) to 0.

$$\begin{aligned} 0 &= -x_i^T (r_i - x_i \hat{\omega}_i) + \lambda \frac{\hat{\omega}_i}{|\hat{\omega}_i|} \\ \iff 0 &= -x_i^T r_i + x_i^T x_i \hat{\omega}_i + \lambda \frac{\hat{\omega}_i}{|\hat{\omega}_i|} \\ \iff x_i^T r_i &= x_i^T x_i \hat{\omega}_i + \lambda \frac{\hat{\omega}_i}{|\hat{\omega}_i|} \end{aligned} \quad (3)$$

Now, we consider the equation by its absolute value.

$$|x_i^T r_i| = |\hat{\omega}_i| \left| x_i^T x_i + \lambda \frac{1}{|\hat{\omega}_i|} \right| \quad (4)$$

$x_i^T x_i$ and λ are positive, meaning that we can write the equation as

$$|x_i^T r_i| = |\hat{\omega}_i| \left(x_i^T x_i + \lambda \frac{1}{|\hat{\omega}_i|} \right) \quad (5)$$

Now, we can easily identify that the entire right hand side is positive. Thus, we can remove the absolute value of the left hand side.

$$x_i^T r_i = |\hat{\omega}_i| \left(x_i^T x_i + \lambda \frac{1}{|\hat{\omega}_i|} \right) \quad (6)$$

$$\iff |\hat{\omega}_i| = \frac{1}{x_i^T x_i} \left(|x_i^T r_i| - \lambda \right) \quad (7)$$

In order to determine $\text{sgn}(\hat{\omega}_i)$ we need to consider the last equation in (3), where $x_i^T x_i$ and λ are positive. Thus,

$$\text{sgn}(\hat{\omega}_i) = \text{sgn}(x_i^T r_i) = \frac{x_i^T r_i}{|x_i^T r_i|} \quad (8)$$

Multiplying (7) with (8) yields

$$\begin{aligned} \hat{\omega}_i &= \text{sgn}(\hat{\omega}_i) |\hat{\omega}_i| = \frac{x_i^T r_i}{|x_i^T r_i|} \frac{1}{x_i^T x_i} \left(|x_i^T r_i| - \lambda \right) \\ &\iff \hat{\omega}_i = \frac{x_i^T r_i}{x_i^T x_i |x_i^T r_i|} \left(|x_i^T r_i| - \lambda \right) \end{aligned} \quad (9)$$

which is what we wanted to show.

2 Task 2

We begin by rewriting

$$\hat{\omega}_i^{(j)} = \frac{x_i^T r_i^{(j-1)}}{x_i^T x_i |x_i^T r_i^{(j-1)}|} \left(|x_i^T r_i^{(j-1)}| - \lambda \right) \quad (10)$$

by using that $x_i^T x_i = I$. This results in

$$\begin{aligned} \hat{\omega}_i^{(j)} &= \frac{x_i^T r_i^{(j-1)}}{|x_i^T r_i^{(j-1)}|} \left(|x_i^T r_i^{(j-1)}| - \lambda \right) \\ &= x_i^T r_i^{(j-1)} - \lambda \text{sgn}(x_i^T r_i^{(j-1)}) \end{aligned} \quad (11)$$

in order to further simplify the equation, we need to rewrite $x_i^T r_i^{(j-1)}$, where

$$r_i^{(j-1)} = t - \sum_{l < i} x_l \hat{\omega}_l^j - \sum_{l < i} x_l \hat{\omega}_l^{j-1} \quad (12)$$

as follows.

$$\begin{aligned} x_i^T r_i^{(j-1)} &= x_i^T \left(t - \sum_{l < i} x_l \hat{\omega}_l^j - \sum_{l < i} x_l \hat{\omega}_l^{j-1} \right) \\ &= \left[x_i^T x_l = 0 \ \forall \ l \neq i, \ x_i^T x_i = 1 \right] \\ &= x_i^T t \end{aligned} \quad (13)$$

Inserting (13) into (11) yields

$$\hat{\omega}_i^{(j)} = x_i^T t - \lambda \text{sgn}(x_i^T t) \quad (14)$$

which shows that $\hat{\omega}_i^{(j)}$ does not depend on previous any previous estimations. Finally, we can show that the coordinate descent will converge after at most one full pass over the coordinates in ω as follows.

$$\hat{\omega}_i^{(2)} - \hat{\omega}_i^{(1)} = (x_i^T t - \lambda \text{sgn}(x_i^T t)) - (x_i^T t - \lambda \text{sgn}(x_i^T t)) = 0 \quad (15)$$

These properties also apply for the case $|x_i^T r_i^{(j-1)}| < \lambda$ as $\hat{\omega}_i^{(j)} = 0$.

3 Task 3

Firstly, we note that

$$\hat{\omega}_i^{(j)} = x_i^T t - \lambda \text{sgn}(x_i^T t) \quad (16)$$

as shown in (14) and investigate what happens to $x_i^T t$ if we let $\sigma \rightarrow 0$.

$$\lim_{\sigma \rightarrow 0} x_i^T t = x_i^T X \omega^* = \omega^* \quad (17)$$

The three cases we want to investigate are

$$x_i^T r_i^{(j-1)} > \lambda, \quad (18)$$

$$x_i^T r_i^{(j-1)} < -\lambda, \quad (19)$$

$$|x_i^T r_i^{(j-1)}| \leq \lambda \quad (20)$$

In the case given by (18):

$$\begin{aligned} & \lim_{\sigma \rightarrow 0} E(\hat{\omega}_i^{(1)} - \omega_i^*) \\ &= \lim_{\sigma \rightarrow 0} E(x_i^T t - \lambda \text{sgn}(x_i^T t) - \omega_i^*) \\ &= E(\hat{\omega}_i^* - \text{sgn}(\hat{\omega}_i^*) - \omega_i^*) \\ &= -\lambda E(\text{sgn}(\omega_i^*)) = -\lambda \end{aligned} \quad (21)$$

In the case given by (19), we re-use the calculations from (21) but with a negative $\hat{\omega}_i^*$:

$$\begin{aligned} & \lim_{\sigma \rightarrow 0} E(\hat{\omega}_i^{(1)} - \omega_i^*) \\ &= E(\omega_i^* - \text{sgn}(\omega_i^*) - \omega_i^*) \\ &= -\lambda E(\text{sgn}(\omega_i^*)) = \lambda \end{aligned} \quad (22)$$

In the final case given by (20), we note that $\hat{\omega}_i^{(1)} = 0$. Meaning that

$$\begin{aligned} & \lim_{\sigma \rightarrow 0} E(\hat{\omega}_i^{(1)} - \omega_i^*) \\ &= E(0 - \omega_i^*) \\ &= -\omega_i^* \end{aligned} \quad (23)$$

LASSO (or Least Absolute Shrinkage and Selection Operator) appears to be a great method for avoiding high variance in a model. However, as is shown in the results above, strong regularization (large λ) is likely to result in high bias and therefore an inaccurate model.

4 Task 4

In this task, a signal was reconstructed using LASSO regularization with 3 different λ . The original signal is a linear combination of two sinusoids with different frequencies.

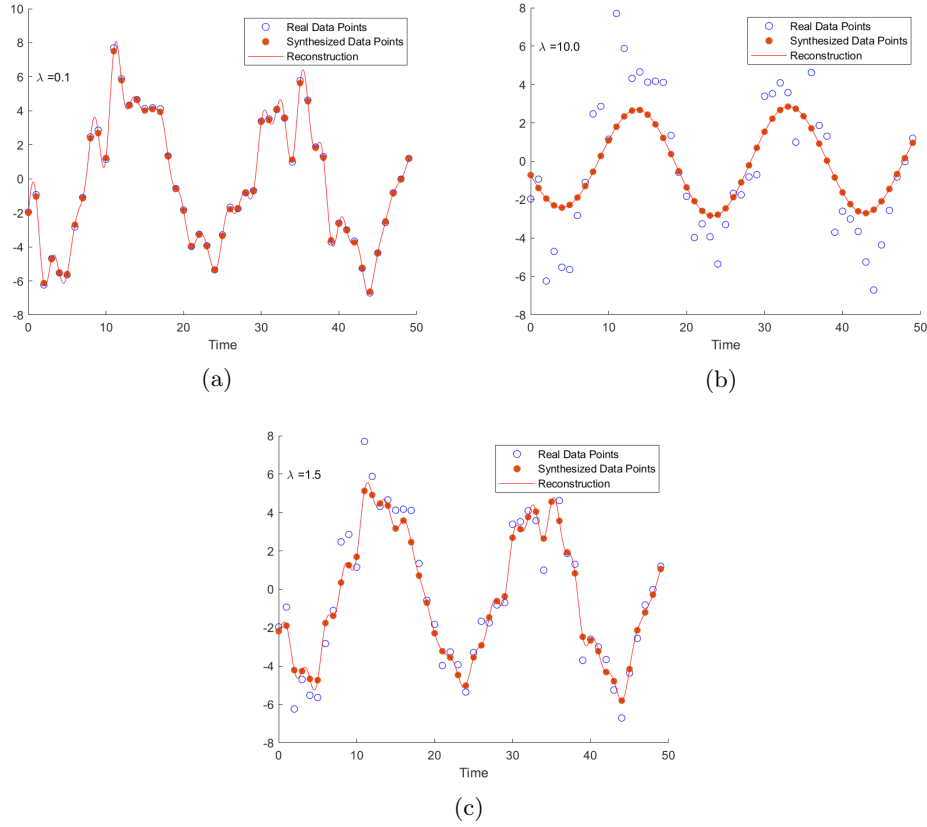


Figure 1: (a) Reconstruction plot using $\lambda = 0.1$. (b) Reconstruction plot using $\lambda = 10$. (c) Reconstruction plot using $\lambda = 1.5$.

In figure 1 (a) it is clearly visible that the model is overfitted as the real and synthesized data points overlap to a great extent. This means that the model is fitting to the noise in the signal as opposed to the actual original signal that we are trying to reproduce. This can be seen as the reconstruction line crosses all real data points.

Figure 1 (b) does not show any signs of overfitting, instead it appears that the model is underfitted. The reason being that the reconstruction line can be considered far away from most real data points.

In contrast, figure 1 (c) shows a reasonable reconstruction of the signal. The line appears to be a linear combination of two sinusoids with different frequencies, while keeping a reasonable distance to the real data points. This indicates that the noise in the signal has been filtered out.

λ	Subfigure	Number of non-zero coordinates
0.1	(a)	253
10	(b)	8
1.5	(c)	61

Table 1: Number of non-zero coordinates for each model presented in Figure 1.

As indicated in table 1, there is an inversely proportional relation between λ and the number of non-zero coordinates for the models presented. The table along with the figures also show that a reasonable reconstruction needs more non-zero coordinates than the actual explanatory variable consisting of 4 non-zero coordinates.

5 Task 5

In order to find the optimal λ for reconstruction of the signal, the RMSE of the reconstruction using the validation data set was minimized. This was done using 10-fold cross validation with 100 different λ values, where $\lambda \in [0.01, \max(|X^T t|)]$. The results are shown in figure 2. Using the optimal $\lambda = 1.8$

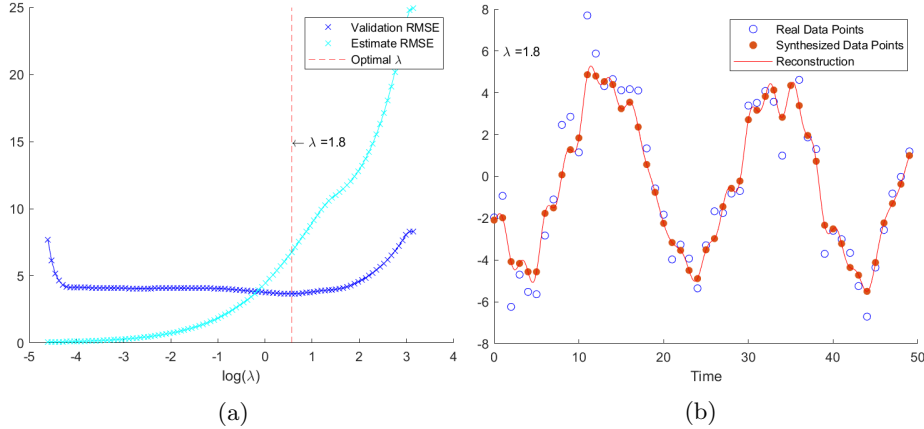


Figure 2: (a) Validation and estimation RMSE using 10-fold cross validation with 100 different λ . (b) Reconstruction using optimal $\lambda = 1.8$.

to construct our model, we get a fairly reasonable reconstruction of the signal (see Figure 2 (b)) as per the description of Figure 1 (c) under the section *Task 4*.

6 Task 6

In order to find the optimal lambda across the frames of the audio excerpt, multiframe 3-fold cross validation was used. The RMSE for one λ was calculated as the mean RMSE for that lambda across frames. The optimal λ was obtained by identifying the λ corresponding to the smallest validation RMSE.

In this task, $\lambda \in [0.00001, \max(|X^T t_i|) \forall i]$ where i is the frame index.

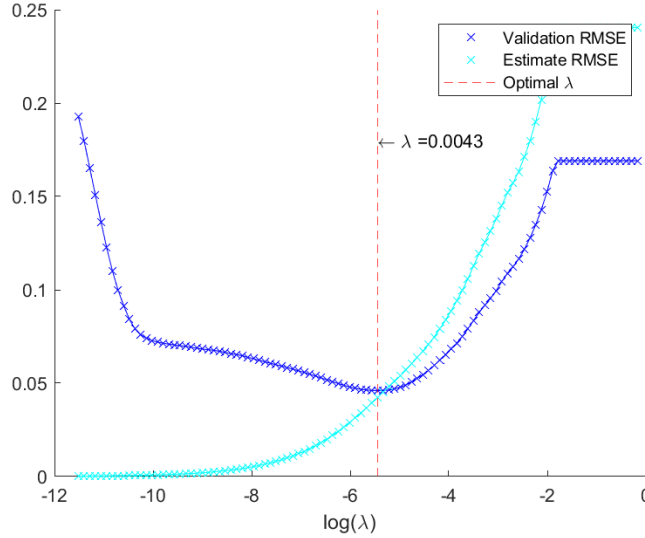


Figure 3: Validation and estimate RMSE for different λ .

As shown in figure 3, the regularization is needed in order to produce a clear reconstruction of the audio. No, or little regularization strength is shown to result in a noisy reconstruction as the validation RMSE is high for small λ . This also applies to strong regularization. However, there appears to be an optimum at $\lambda = 0.0043$.

7 Task 7

When listening to the original unfiltered audio, there was a significant amount of noise present. However, after filtering the noise using our model the sound was much more clear. Adjusting the *lambda* for the model to a greater or lesser value both resulted in more distorted sound, indicating that optimal λ proposed by the multiframe cross validation was very close to the actual optimal value.