

Assignment 4: CS 754, Advanced Image Processing

Group Details:

Vinit Awale (18D070067)

Piyush Bharambe (18D070019)

Question 1

- Consider a signal \mathbf{x} which is sparse in the canonical basis and contains n elements, which is compressively sensed in the form $\mathbf{y} = \Phi\mathbf{x} + \boldsymbol{\eta}$ where \mathbf{y} , the measurement vector, has m elements and Φ is the $m \times n$ sensing matrix. Here $\boldsymbol{\eta}$ is a vector of noise values that are distributed by $\mathcal{N}(0, \sigma^2)$. One way to recover \mathbf{x} from \mathbf{y} , Φ is to solve the LASSO problem, based on minimizing $J(\mathbf{x}) \triangleq \|\mathbf{y} - \Phi\mathbf{x}\|^2 + \lambda\|\mathbf{x}\|_1$. A crucial issue is to how to choose λ . One purely data-driven technique is called cross-validation. In this technique, out of the m measurements, a random subset of (say) 90 percent of the measurements is called the reconstruction set R , and the remaining measurements constitute the validation set V . Thus V and R are always disjoint sets. The signal \mathbf{x} is reconstructed using measurements only from R (and thus only the corresponding rows of Φ) using one out of many different values of λ chosen from a set Λ . Let the estimate using the g^{th} value from Λ be denoted \mathbf{x}_g . The corresponding validation error is computed using $VE(g) \triangleq \sum_{i \in V} (y_i - \Phi^i \mathbf{x}_g)^2 / |V|$. The value of λ for which the validation error is the least is chosen to be the optimal value of λ . Your job is to implement this technique for the case when $n = 500, m = 200, \|\mathbf{x}\|_0 = 18, \sigma = 0.05 \times \sum_{i=1}^m |\Phi^i \mathbf{x}| / m$. Choose $\Lambda = \{5 \times 10^{-4}, 10^{-4}, 5 \times 10^{-3}, 10^{-3}, 0.01, 0.05, 0.1, 0.5, 1, 2, 5\}$. Draw the non-zero elements of \mathbf{x} at randomly chosen location, and let their values be drawn randomly from $\text{Uniform}(0, 1000)$. The sensing matrix Φ should be drawn from $\pm\text{Bernoulli}$ with probability of $+1$ being 0.5. Now do as follows. Use the L1-LS solver from https://web.stanford.edu/~boyd/l1_ls/ for implementing the LASSO.

1. Plot a graph of VE versus the logarithm of the values in Λ . Also plot a graph of the RMSE versus the logarithm of the values in Λ , where RMSE is given by $\|\mathbf{x}_g - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$. Comment on the plots. Do the optimal values of λ from the two plots agree?

Answer: The plot of VE and $RMSE$ versus the logarithm of the values in Λ are as follows:

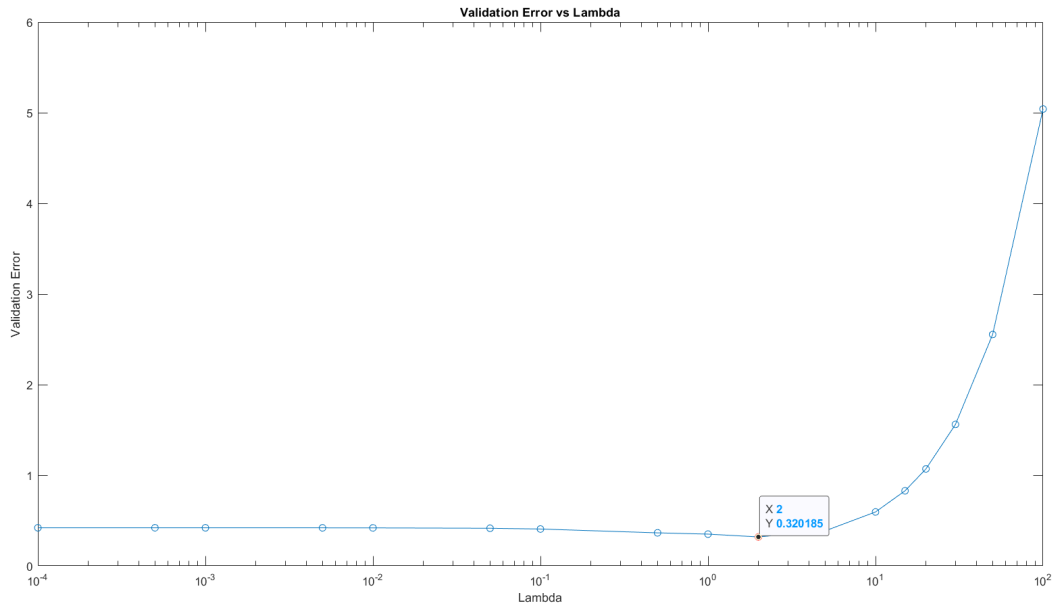


Figure 1: Plot of VE versus the logarithm of the values in Λ

From this plot we can easily see that the validation error is high for higher values of lambda. However, for the values of lambda lower than 5, the validation error is very low with very little variation. Also, the minimum validation error is obtained for $\lambda = 2$ as highlighted in the figure. Hence, the optimal value of λ is 2.

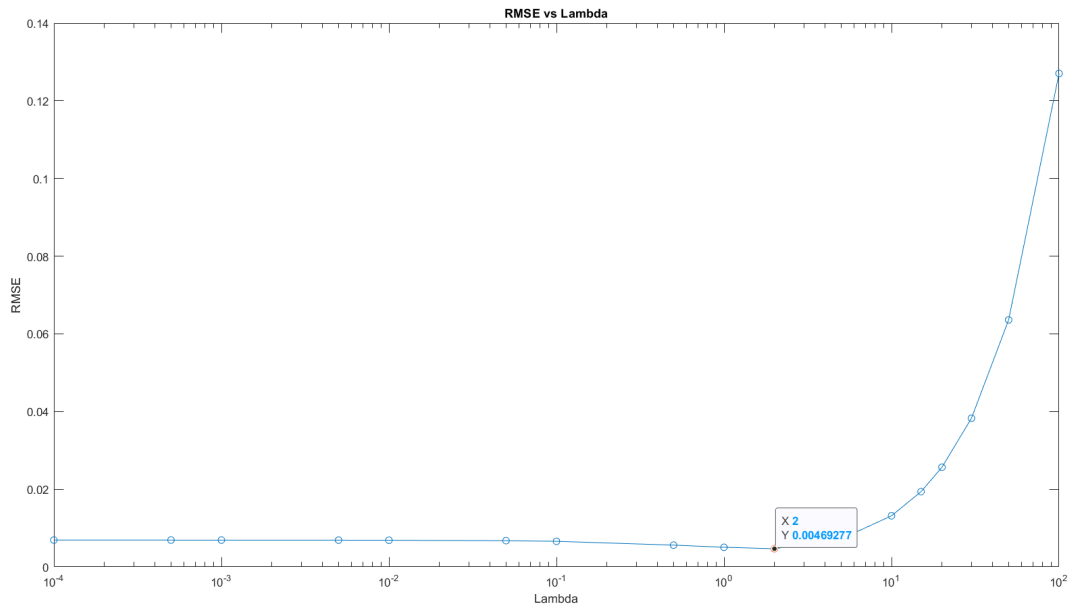


Figure 2: Plot of $RMSE$ versus the logarithm of the values in Λ

Clearly, the trend of the RMSE vs lambda plot is similar to that of the VE vs lambda plot. Similarly, the RMSE is very low for the values of λ greater than 5 and the optimal value of λ is 2.

Hence, the value of lambda that we obtain from both the methods is the same.

2. What would happen if V and R were not disjoint but coincident sets?

Answer: If V and R are coincident sets, then we are actually attempting to find the value of λ that minimizes the error on the reconstruction set itself. Clearly, that value of λ should be zero as by adding a positive term to $\|\mathbf{y} - \Phi\mathbf{x}\|^2$, we are effectively increasing the value of the objective function $J(\mathbf{x})$. Hence, the optimal value of λ that is obtained will be close to zero when V and R are coincident sets. This is the reason we need a disjoint validation set (for determining the optimal λ) which verifies the ability of the algorithm to reconstruct using unseen data.

3. The validation error is actually a proxy for actual mean squared error. Note that you can never determine the mean squared error since the ground truth \mathbf{x} is unknown in an actual application. Which theorem/lemma from the paper <https://ieeexplore.ieee.org/document/6854225> (On the theoretical analysis of cross-validation in compressed sensing) refers to this proxying ability? Explain how.

Answer: In the given paper, the **Theorem 1** and **Theorem 2** prove the proxying ability of the validation error.

Theorem 1 (Recovery error estimation) provides bounds on the recovery error ϵ_x (i.e. the actual MSE) in terms of the Cross validation (CV) residual ϵ_{CV} (i.e. the validation error). Moreover, it has been shown that this bound becomes tighter as the number of CV measurements (m_{CV}) increases. Hence if we have sufficiently large number of CV measurements, and we obtain the hyperparameters that reduce the validation error, then those hyperparameters also reduce the actual mean squared error.

Theorem 2 (Recovery error comparison) states that if \hat{x}^p and \hat{x}^q are two recovered signals such that the actual MSE $\epsilon_x^p \geq \epsilon_x^q$, then the validation errors also follow $\epsilon_{CV}^p \geq \epsilon_{CV}^q$ with probability that increases as the number of CV measurements increases. Hence, if we have sufficiently large number of CV measurements, and we obtain the hyperparameters that reduce the validation error, then those hyperparameters also reduce the actual mean squared error.

Hence, we can indeed use the validation error as a proxy for the actual mean squared error given that we have sufficiently large number of CV measurements.

4. In your previous assignment, there was a theorem from the book by Tibshirani and others which gave you a certain value of λ . What is the advantage of this cross-validation method compared to the choice of λ using that theorem? Explain.

Answer: The theorem from the book by Tibshirani and others gives us a bound on the value of λ as

$$\lambda_m > 2 \frac{\phi^T \eta}{m}$$

where ϕ is a $m \times n$ matrix. Furthermore, for Gaussian noise the authors showed that the appropriate choice with high probability is

$$\lambda_m = 2\sigma \sqrt{\frac{\tau \log(n)}{m}}$$

for some $\tau > 2$ and σ is the standard deviation of the noise.

However, this choice is only valid if sensing matrix ϕ satisfies the restricted eigenvalue condition with parameter τ for some cone C . The cross validation method does not require such assumption on

sensing matrix ϕ .

Moreover, checking the eigenvalue condition is computationally expensive compared to performing cross validation over a set of possible λ values. Hence, the cross validation method is more efficient.