Title: Numerical Methods for Information Tracking of Noisy and Non-smooth Data in Large-scale Statistics

Authors: Avinash B. S*[1]., Srisupattarawanit T[1]. and Ostermeyer H[1].

*[1] - *Institute for Dynamics and Vibration, Technical University of Braunschweig, Germany*

Authors' contributions:

This work was the collaborative effort of all authors. Author BS managed the literature search for the present review article and wrote the first draft of the manuscript. Authors ST and OH assisted in the arrangement of article and finalized the draft. All the authors read and approved the final manuscript.

ABSTRACT

In our universe, there is a presence of random bit of disorder in every field that has to be contemplated and understood clearly. This random bit of disorder in a physical system is known as noise. Noise in the field of statistics can be defined as an additional meaningless information that cannot be clearly interpreted which is present in the entire dataset. In large-scale statistics, noisy data has an adverse effect on the results and it can lead to skewness in any data analysis process, if not properly understood or handled. The adverse effect on the results is mainly due to uncorrelated (zero autocorrelation) property of noise. This makes it completely unpredictable at any given point in time, hence thorough investigation and removal of noise plays a vital role in data analysis process. In the field of engineering, measurement of experimental data obtained by using scientific instruments consists of some values that are independent of the experimental setup. One of most widely technique is the optimization methods viz, gradient descent, conjugate gradient, Newton's method etc. Most of these methods require the determination of derivative of a function specified by the dataset (using finite-difference approximation). If the noisy data is approximated using a specific finite difference method this results in the amplification of noise present in the data. In order to overcome the aforementioned problem of amplification of noise in the derivative of a function, various regularization methods are employed. The parameter that plays a vital role in these methods are termed as regularization parameter. One of the most important technique used in the field of regularization is known as total variation regularization. This review aimed at gathering the disperse literature on the current state of various noises and their regularization methods.

Key Words: Large-scale statistics; Noisy data; Regularization; Data driven methods; Amplification

---

* Corresponding author - bapuavinash6@gmail.com; +49 176 72100929

1. INTRODUCTION

In the modern field of engineering, we deal with a lot of experimental data that may consists of errors. These errors possess the properties of randomness and non-correlation meaning that they are completely unpredictable in nature. Hence the knowledge behind these errors, proper handling and removal techniques are prioritized during the early phase of data analysis [1]. Various numerical method for approximating the derivative of functions like finite-difference methods have taken center stage in many engineering interdisciplinary for optimization purposes. Application of these finite-difference methods to the noise contaminated dataset leads to intensification of already present noise. These amplification in the derivatives can be suppressed by applying total variation (TV) regularization technique. TV deals directly with the process of differentiation. This process of regularization assures that the calculated derivative of the function adheres to a certain degree of regularity [2]. The successful implementation of this methods hinges on one aspect, i.e., clearly understandingand determination of regularization parameter.

There are various methods that facilitates the determination of optimal regularization parameter One of the most important and widely used is the L-curve method. This method provides information on the regularization parameter based on the residual norm (L2) and the solution norm (L1) [3]. The graphical representation between the two for different regularization parameter provides an intersection point that stabilizes the effect of both the residual and the solution. This point is chosen as the optimal regularization parameter by using curvature plot [4].

A method that completely focuses on extensive analysis of residual vector is the normalized cumulative periodogram [5]. The selection of optimal regularization parameter is based on Kolmogorov-Smirnov test i.e., the cumulative periodogram must strictly lie within the confidence interval of 95% [6]. In these circumstances, the user is generally in a tough spot. Hence the generalized cross validation method is employed to overcome complexities of unknown exact data or the variance of noise [7].

These optimal parameters can then be used in the data-driven (sparse regression) method in order to determine the PDE of the governing equation. This method provides good ap- proximation of the system as this uses brute-force search and the sparse regression technique for sparse nonlinear time series matrix in order to achieve its goal [8]. With this background, an attempt has been made in this study to investigate the implications of noisy data in large scale statistics and regularization of noisy data in order to retrieve vital information.

2. GENERAL CONSIDERATION

Raw data collection, different types of noise present in a general system, processing and regularization are the important steps of this study. There are many regularization methods, few of the commonly used in the field of signal processing are: Ridge regression; Least Absolute Shrinkage and Selection Operator (LASSO) and Total Variation Regularization or Rudin–Osher–Fatemi model. The collected data must then be organized for future analysis. This process of organization of collected data is known as data processing. Example of data processing is the

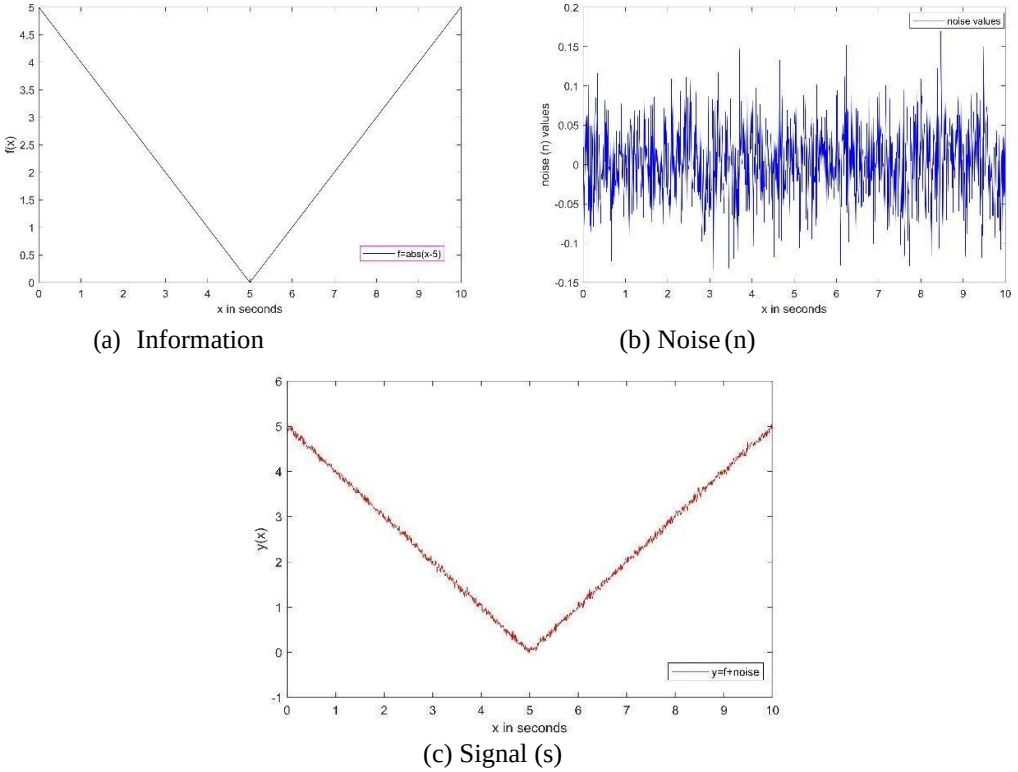
placement of data into columns and rows with respective variable names in a statistical software (Microsoft® Excel or Minitab™).

2.1 DIFFERENT TYPES OF NOISE PRESENT IN A GENERAL SYSTEM

In order to maximize the potential of the aforementioned regularization methods, we shall start with the brief understanding of different types of noise present in a general system described in equation 2.1 with Fig. 1:

$$s = i + n$$

where, s = Signal; i = Information and n = Noise



(a) Information (b) Noise (n)

(c) Signal (s)

Fig. 1: Graphs depicting the general system in equation 2.1

2.2 DATA ANALYSIS AND STEPS INVOLVED IN THE PROCESS

The process of obtaining raw data and its conversion into information which is useful for decision-making by the user, this is known as data analysis. The various steps involved in data analysis are shown in Fig. 2.

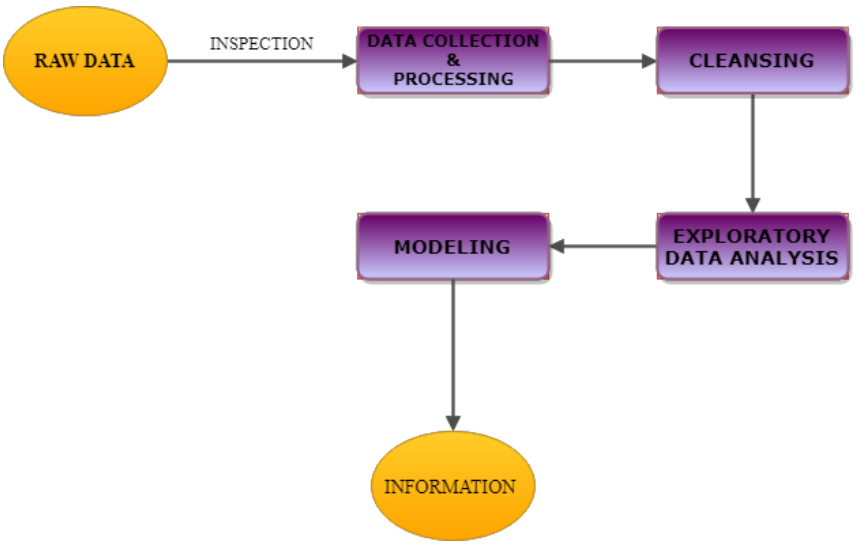


Fig. 2: A picture showing the steps involved in data analysis

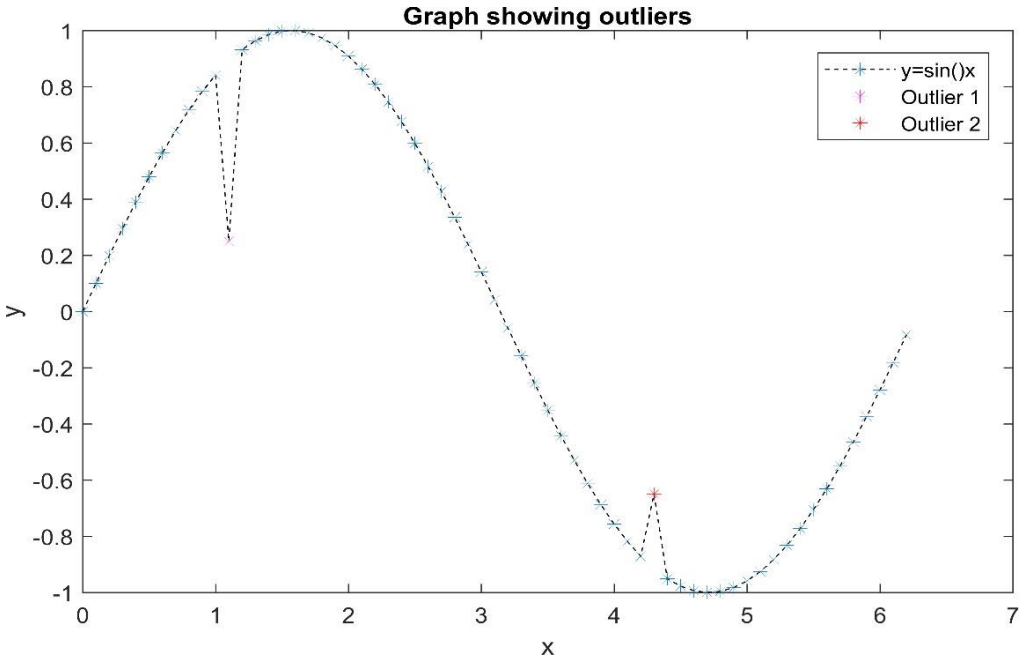
2.2.1 DATA COLLECTION AND PROCESSING

Data in general can be collected from various number of sources. Digital sources of data collection are some of the most convenient and trusted forms. In today's world where technological advancement is at its peak, sensors form a large part of data collection [9]. They are reliable, accurate and can transmit data round-the-clock to computers which can then be analyzed by the engineers. Temperature sensors in nuclear power plants, on aircraft to monitor engine temperature, seismic sensors in high earthquake prone regions in world are few examples that can provide engineers and scientists' accurate data that can save lives during critical situations.

The collected data must then be organized for future analysis. This process of organization of collected data is known as data processing. Example of data processing is the placement of data into columns and rows with respective variable names in a statistical software (Microsoft® Excel or Minitab™).

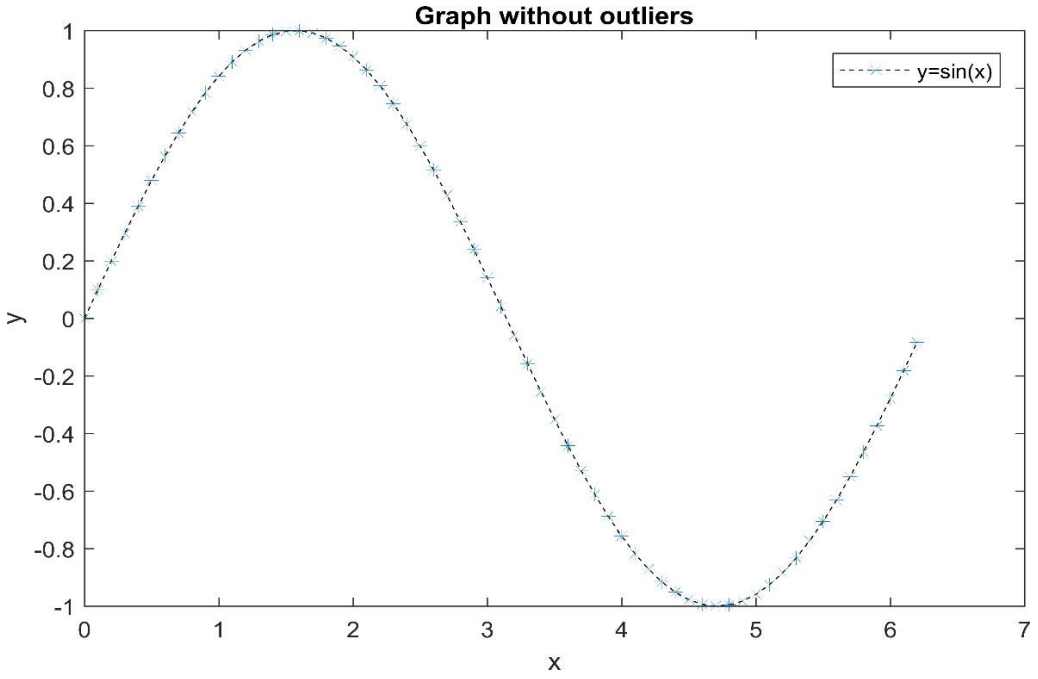
2.2.2 CLEANING OF PROCESSED DATA

Data cleaning (cleansing) is the process of understanding, collection and then removal of errors that may be present in the processed data [10]. This process is very critical during the final step of data analysis as it improves the accuracy of results. When dealing with quantitative processed data using various outlier removal methods forms the part of data cleaning. Outliers are values or observation in processed data that lie far part from the main pattern of the entire dataset. Fig. 2 shows a process with (Fig. 2a and without outliers 2b).



Time in seconds (s)

(a) Graph outlier



Time in seconds (s)

(a) Graph without outlier

Fig. 3: Representation of Outliers in a process

There are various methods to detect outliers in a process, one of the most commonly used technique is the scatterplot. This is very easy and quick process to detect the number of points lying outside the standard pattern of the whole process (Fig.3).

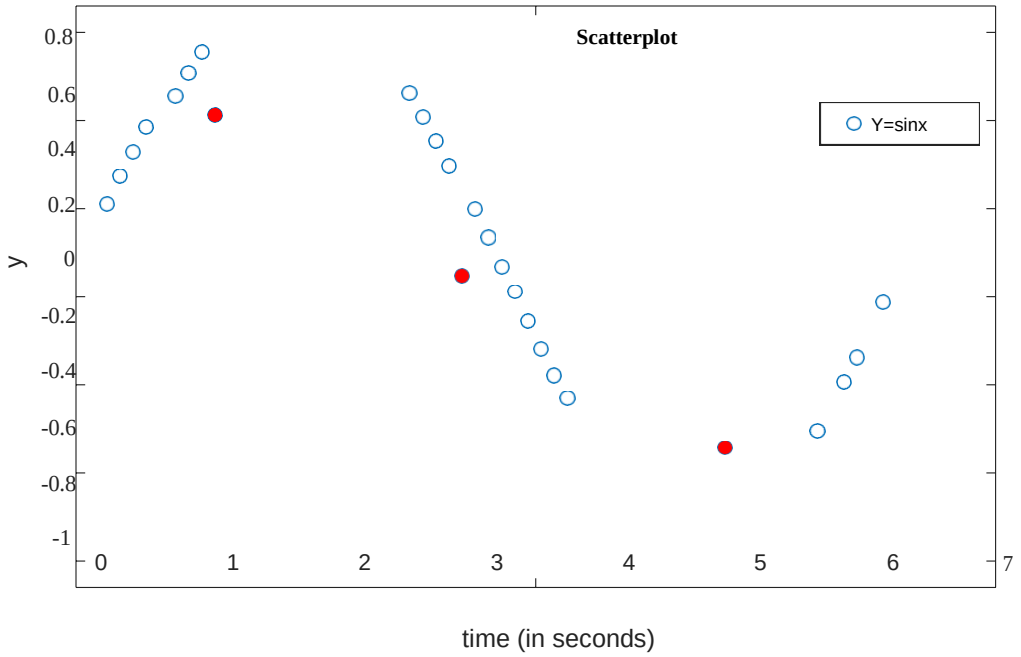


Fig. 4: A scatterplot showing the process trend and the detected outliers

There are many other techniques like the box plot that are used in the detection of outliers in a process. The advantage of using box plot is that it provides clear information on mild and extreme outliers. Box plot also has the option of detecting outliers by using median, 1st and 3rd quartile principle. A typical boxplot is shown in Fig. 5.

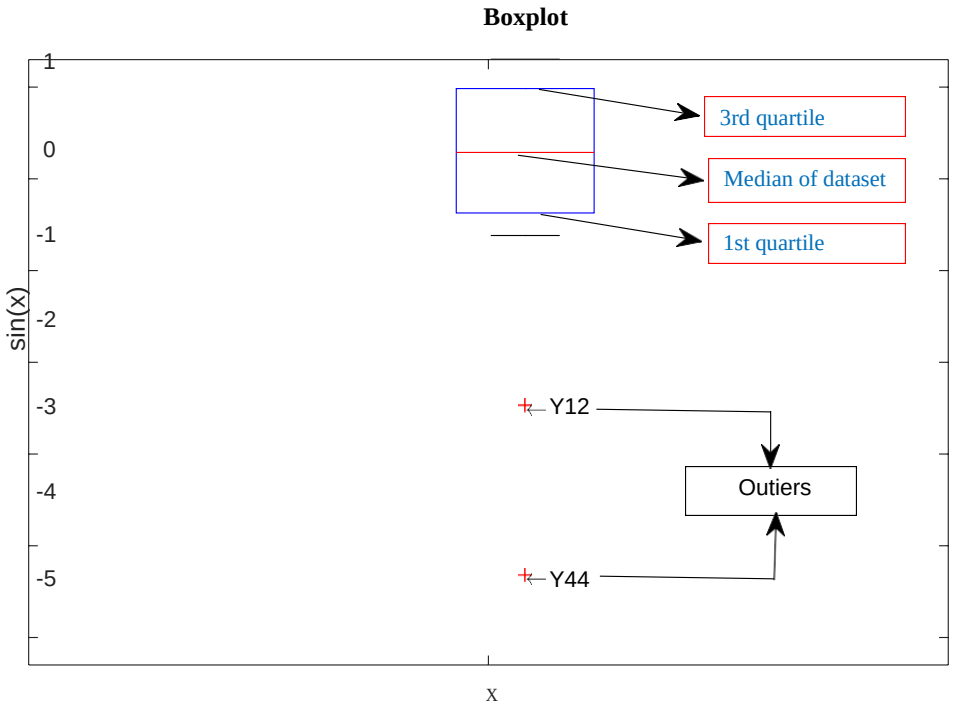


Fig. 4: A boxplot showing the detected outliers

After the detection of outliers, one cannot simply employ univariate and multivariate methods to remove the detected outliers as it can have adverse effect on the entire process. So using robust

techniques like "Minkowski error" method helps to reduce the impact of outliers on the dataset (or model). The major advantage of "Minkowski error" over RSS is that it reduces the effect of outliers by taking the power of error terms lesser than 2 [11].

In certain scenarios, processed data and/or processed data after treating outliers may be skewed. This type of skewed data needs to be transformed using certain transformation techniques before analyzing exploratory. The most common method employed for skewed data is the Box-Cox (or power) transformation.

$$x(\lambda) = \frac{(x^{\lambda} - 1)}{\lambda} \qquad \lambda = 0 \quad (2.2)$$

$$x(\lambda) = \ln(x) \qquad \lambda = 0 \quad (2.3)$$

where, $x(\lambda)$ = Transformed data; x = Skewed data; λ = Box-Cox parameter

But the best way [12] to select "λ" is by using LLF (logarithm of likelihood function).
This marks the conclusion of cleansing of processed data.

2.2.3 EXPLORATORY DATA ANALYSIS

The process of deciphering the cleaned data extensively by using visualization techniques, calculation of vital descriptive statistics (like mean, median, mode etc.) is known as exploratory data analysis. This helps the user to comprehend the meaning behind the obtained dataset. Hence it translates to exploring the cleaning data from all possible angles. It consists of many sub-tasks like, re-cleansing (if necessary), procurement of additional data, calculation of descriptive statistics and visualization

2.2.4 DATA MODELING

The final step in process of data analysis is data modeling. The knowledge obtained from exploratory data analysis steps plays a vital role in the identification of certain relationship between variables. Various relationships such as regression analysis, correlation can be obtained by compiling specific algorithms and/or applying specific mathematical formulae. Finally, the user can construct descriptive models for analysis [13]. The results obtained can be termed as information, this can help the user to understand the datasets and certain changes can be made in order to improve the efficiency of the process for future studies.

3. A BRIEF DISCUSSION ABOUT NOISE

This section focuses on the different types of noise and its characteristics encountered in various statistical and signal processing fields. As shown in equation 2.1, noise "n" can be classified as shown below,

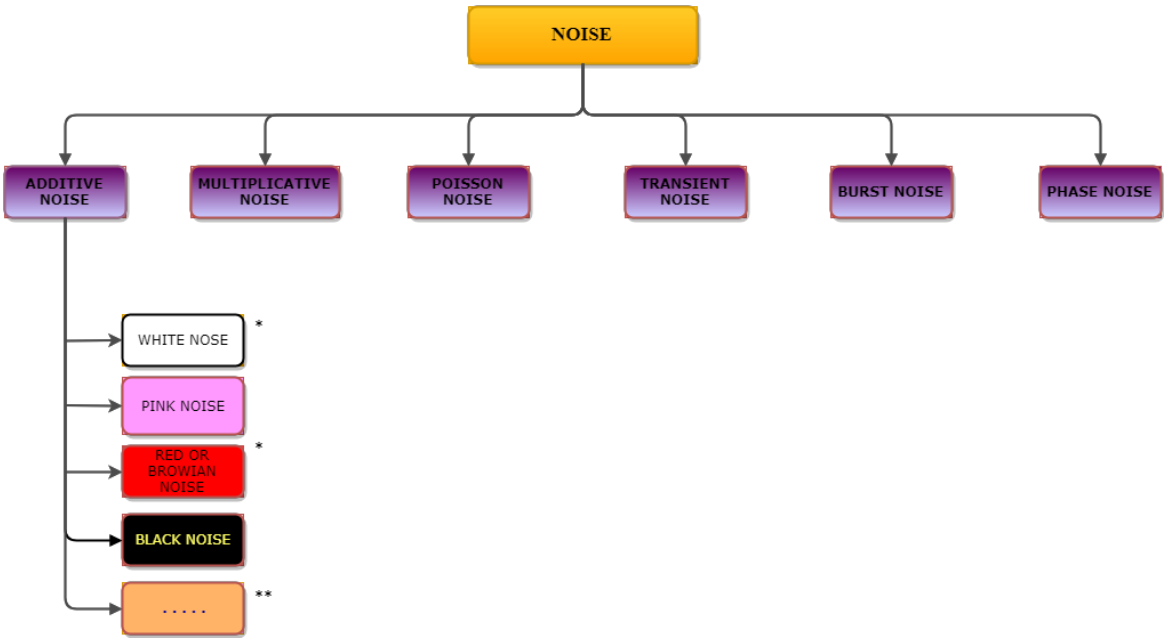


Fig. 5: Classification of noise

Note: As seen from Fig. 5,

$(*)$ $\Longrightarrow$ main focal point. Hence it is explicitly described in 3.3.

$(**)$ $\Longrightarrow$ additive noise includes many other slightly less significant subdivision.

3.1 DIFFERENT TYPES OF NOISE (Fig 5) are explained below[14]:

3.1.1 Multiplicative noise:

In a given system, if the random term depends on the state of that system, this type of noise is termed as multiplicative noise. In terms of dataset, we can say that the noisy data is the resultant of noise multiplied to the data vector. This can be clearly interpreted with the help of a following system(model).

$$s = i \cdot n \qquad (2.4)$$

where s= Signal; i=Information (true signal) and n=Noise

Denoising of multiplicative noise requires a transformation of the model in equation 2.4 into additive noise. Logarithmic transformation is very helpful tool in denoising multiplicative noise as this provides an additive form.

$$\log(s) = \log(i \cdot n) \qquad (2.5)$$

$$\log(s) = \log(i) + \log(n) \qquad (2.6)$$

where s=Signal; i=Information(true signal) and n=Noise

Now, equation 2.6 clearly represents an additive system and various denoising techniques can be applied. Finally, inverse logarithm ($\log^{-1}$) of the denoised signal provides the solution to the original system.

3.1.2 Poisson Noise:

Poisson noise is also termed as shot noise (Fig. 6). Shot noise is mainly observed in electronic devices. This type of noise is generated when a charge carrier such as electrons or ions travel through a gap results in random fluctuation in electric current. This random fluctuation is known as shot noise [15].

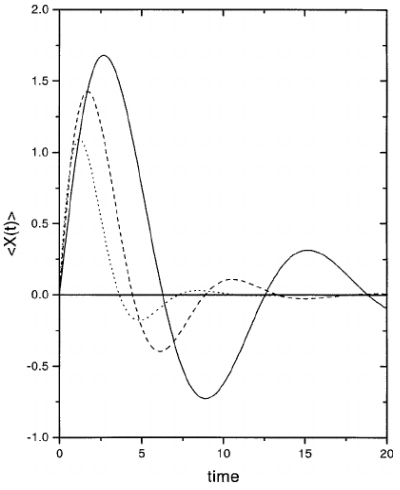


Fig.6: Poisson Noise

3.1.3 Transient Noise:

This type of noise is very common in the field of communication systems like mobile phones and hearing aids. The background noise that hinders communication in the field of communication systems is termed as transient noise (Fig.7)

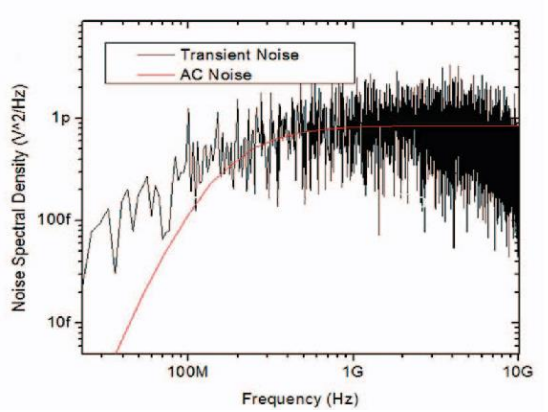


Fig.7: Transient noise

3.1.4 Burst Noise:

Burst noise is also termed as Random Telegraph Signal (RTS) and "popcorn" noise. It is very similar to the shot noise and generated at low frequencies. When a single charger carrier is captured by a single trapping center, this leads to the generation of burst noise as shown in Fig. 8.

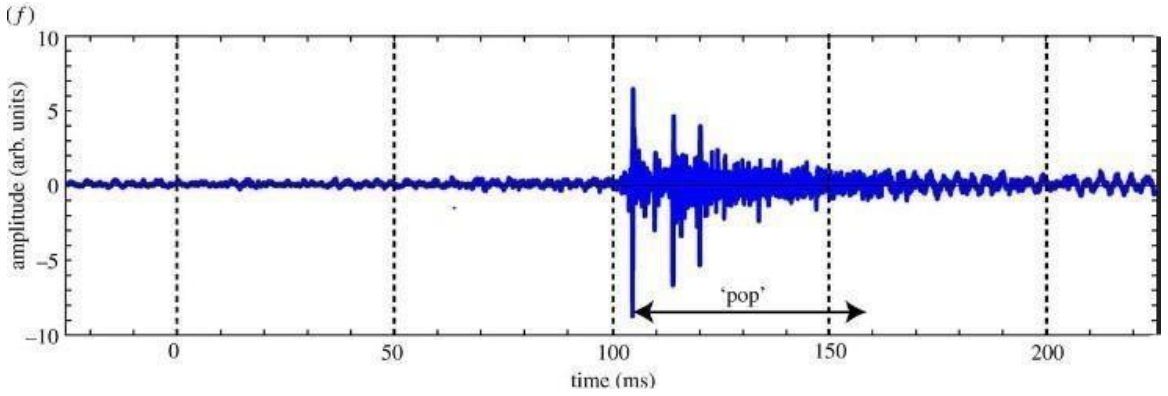


Fig.8: A graph showing the generation of pop (burst) noise

3.1.5 Phase noise:

In order to understand the meaning and definition of phase noise, let us define the term "phase". Phase in a waveform cycle is defined as the position of a point in time. Three types of phases in a wave is shown in Fig.9. Square, triangle, sinusoidal complex are a few examples of different types of waveforms shown in Fig.9. The random and rapid variation of phase in a signal (waveform) caused by time domain instability is known as phase noise.

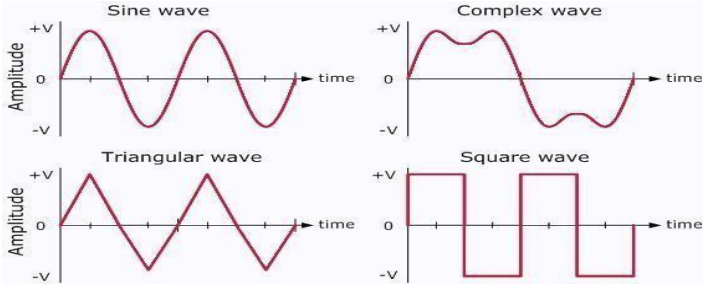
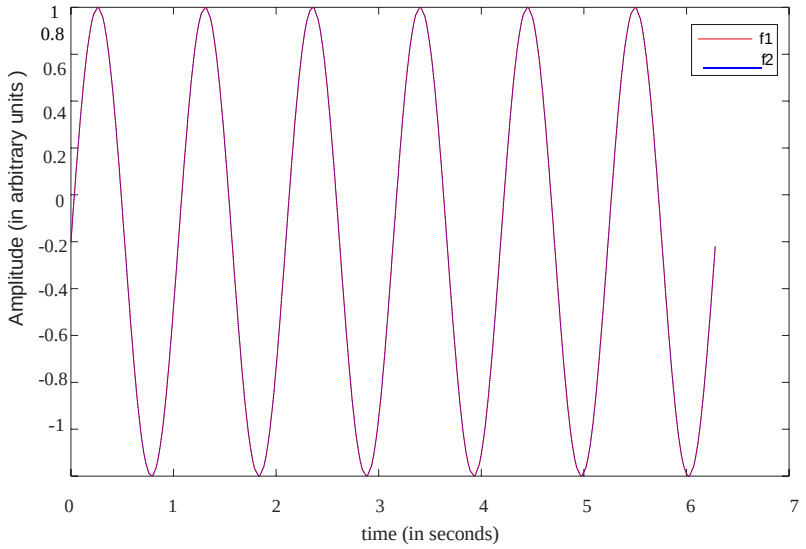
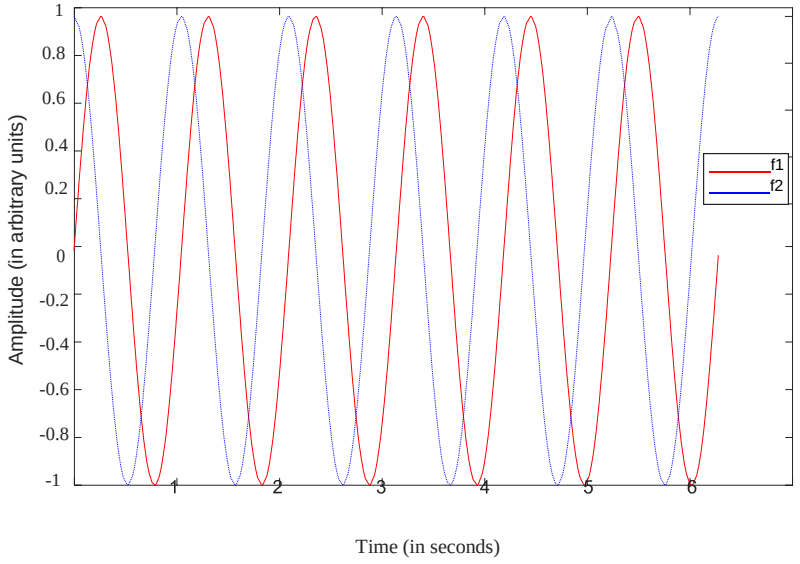


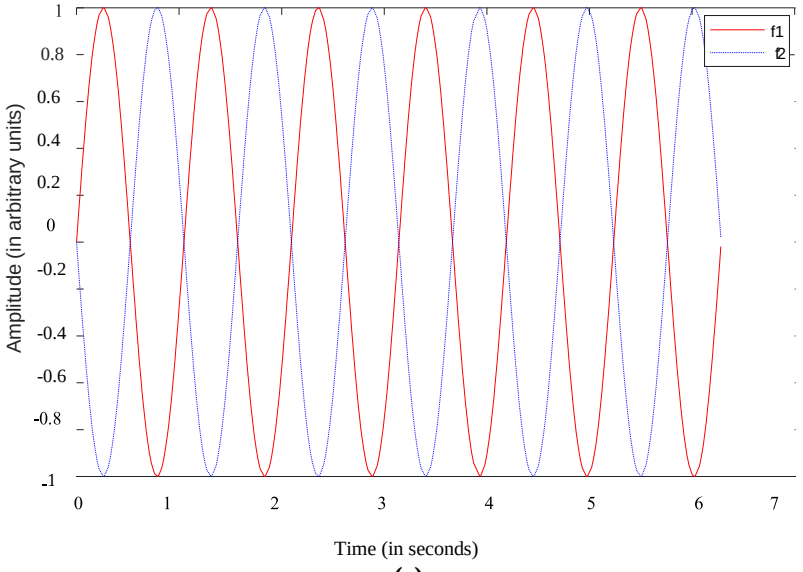
Fig.9 A picture showing common types of waveforms



(a)



(b)



(c)

Fig.10: Graphs showing two waves in phase (a), out of phase (b) & completely out of phase (c)

3.2 Additive White Gaussian Noise (AWGN)

Before jumping into the deep end regarding the explanation of AWGN, let us first break down and understand the terminology "Additive White Gaussian Noise".

3.2.1 Additive =⇒This type of noise are additive in nature. This means that the received signal is the resultant of information added with some noise as shown in equation 2.1.

3.2.2. White =⇒ It is mixture of all types or colors of noise. White light is mixture of all the frequencies or wavelength of visible spectrum (shown in Fig. 11). This definition of white light is literally translated into white noise [16].

3.2.3 Gaussian =⇒ This type of noise follows normal probability distribution Function (pdf), classified as shown below:

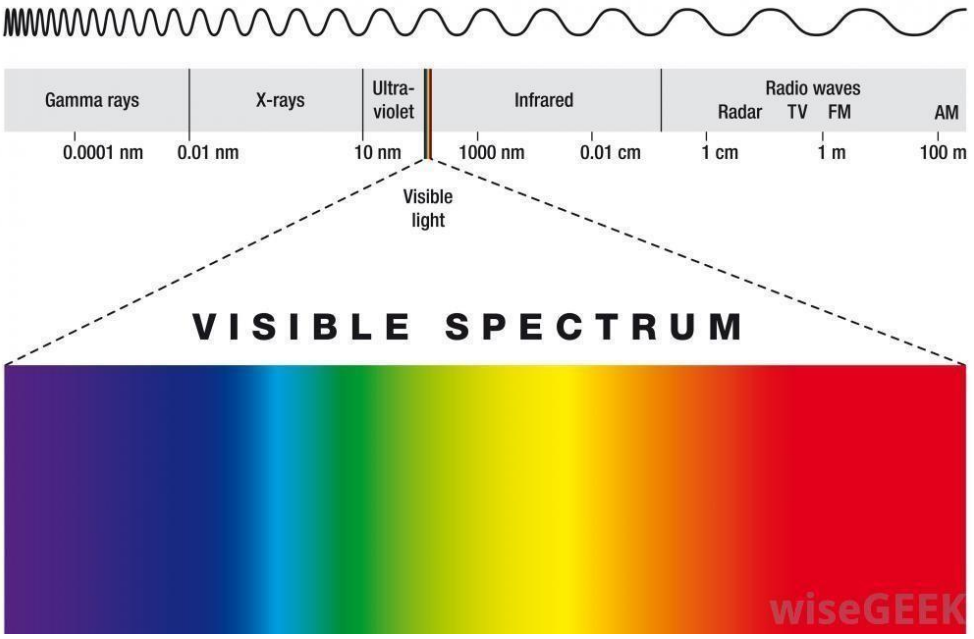
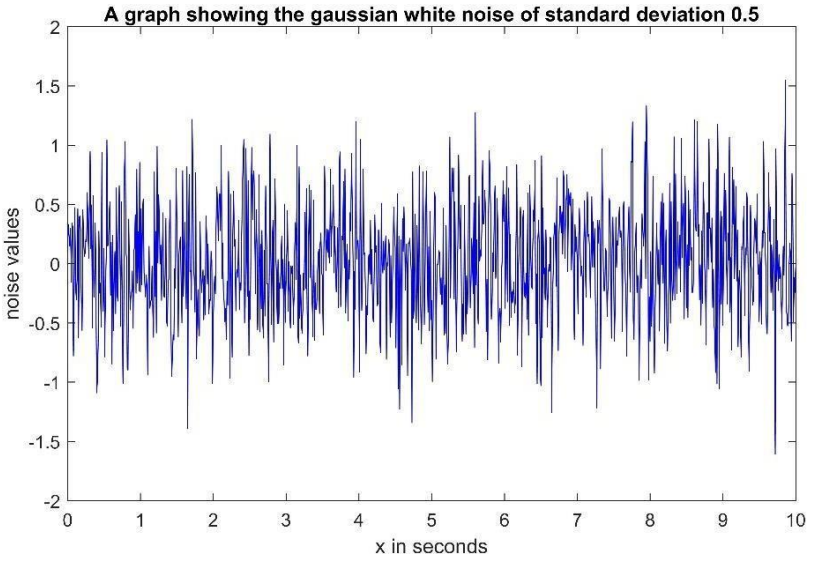
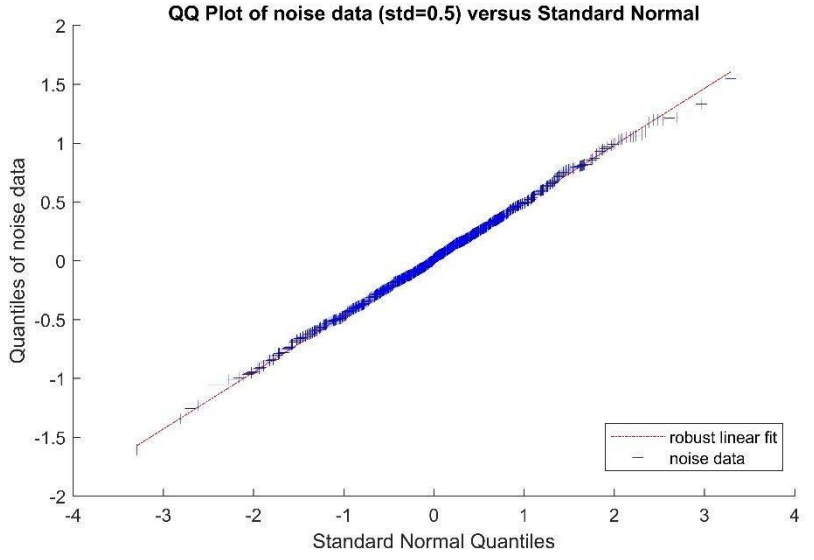


Fig. 11: A picture showing the visible spectrum

White noise with respect to a signal and its source is a statistical model having constant power spectral density (PSD), which means that it is a random noise having equal intensity for different frequencies. An example of the Gaussian white noise is shown below:



(a)



s

(b)

Fig. 12: Representation of Gaussian white noise and its quantile-quantile plot

3.3 Brownian Noise

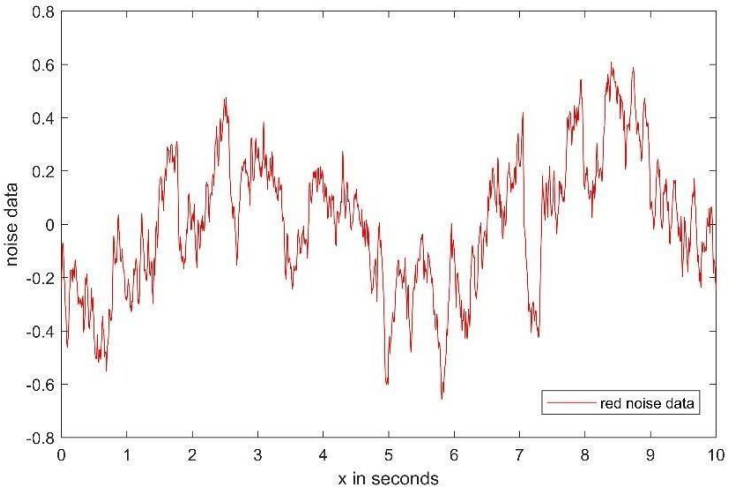
Brownian noise [17] is also known as,

3.3.1 Red Noise –Longer wavelength produces stronger noise similar to radio waves shown in Fig.13, hence the term "red" noise **Brown noise**–Robert Brown discovered Brownian motion. Hence it's also coined as "brown" noise.

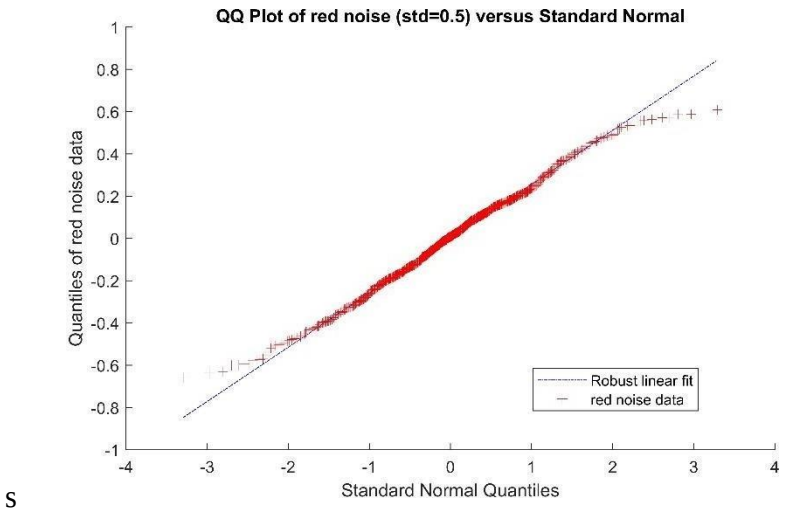
The characteristics of red noise are briefly discussed below,

Red noise has more energy at lower frequencies $\Rightarrow P(f) \propto 1/f^2$.
Power spectrum is denoted by P (f)
Frequency is denoted by f Integration
of white noise → Red noise

An example of the Brownian or red noise is shown below,



(a)



s

(b)

Fig. 13: Representation of Brownian/red noise and its quantile-quantile plot

With this brief understanding of different types of noise, let us now dive into the concepts surrounding important regularization methods.

3.4 A brief discussion regarding regularization methods

As mentioned earlier, the 3 widely used regularization techniques are

1. Ridge regression or Tikhonov regularization method
2. Least Absolute Shrinkage and Selection Operator (LASSO)
3. Total Variation Regularization or Rudin–Osher–Fatemimodel

Before we step into each of the aforementioned regularization techniques, let us define the term regularization. Regularization is defined as a method that helps to overcome the problem surrounding over-fitting of penalized regularization coefficients [18]. This aim of regularization is achieved by the introduction of additional information to solve ill-poised problems. Due to the fact that minimization of residual sum square are highly unstable in nature, regularization methods proves to be all the more important in many scientific fields.

3.4.1 Ridge regression (L2 regularization)

The aim of ridge regression is to minimize the ordinary least square with an added penalty term. This penalty term is the square of the magnitude of the coefficients. This explanation is summarized in equation 2.8.
The ridge regression solution "$\hat{x}_{ridge}$" solves the following minimization problem for a given system Ax = b,

$$\underset{x\in\mathbb{R}^m}{\operatorname{argmin}} \sum_{i}^{n} \left(\sum_{j}^{m} a_{ij} x_j - b_i^2 \right)^2 + \alpha \sum_{j}^{m} x_j^2 \tag{2.7}$$

The equation 2.7 can be represented in a simpler form as,

$$\underset{x\in\mathbb{R}^m}{\operatorname{argmin}} \underbrace{||Ax - b||_2^2}_{Residual} + \alpha \underbrace{\left||x^2\right||_2^2}_{Penalty} \tag{2.8}$$

Where,

$b \in R^n$ = Response vector;

$A \in R^{n\times m}$ = Predictor matrix

α = Regularization parameter

In matrix notation equation 2.8 becomes,

$$C_{ridge} = (A\,x - b)^T (A\,x - b) + \alpha\, x^T x \tag{2.9}$$

Expanding and simultaneous simplification of equation 2.9 results in the following [7]

$$C_{ridge} = x^T A^T A\,x - b^T A\,x - x^T A^T b + b^T b + \alpha\, x^T x \tag{2.10}$$

$$= x^T A^T A\,x - x^T A^T b - x^T A^T b + b^T b + x^T \alpha I \tag{2.11}$$

x

$$= b^T b - 2\,x^T A^T b + x^T A^T A\,x + x^T \alpha I\,x \tag{2.12}$$

$$C_{ridge} = b^T b - 2\,x^T A^T b + x^T (A^T A + \alpha I)x \tag{2.13}$$

The objective function in 2.7 can be minimized by taking the partial derivative of 2.13 with respect to "x".

Minimization condition =⇒ the gradient of the objective function must be equal to zero.

$$\frac{\partial C_{ridge}}{\partial x} = 0 \tag{2.14}$$

$$\Rightarrow \frac{\partial C_{ridge}}{\partial x} = -2\ A^T b + \underbrace{2(A^T A + \alpha I)x}_{*} \tag{2.15}$$

$$\Rightarrow -2\ A^T b + 2(A^T A + \alpha I)x = 0 \tag{2.16}$$

$*$ indicates that the specific part of the equation was achieved by successfully applying matrix (symmetric) differentiation rule

Simplification of the equation 2.16 leads to the ridge regression solution i.e., "$\hat{x}_{ridge}$"

$$\hat{x}_{ridge} = (A^T A + \alpha\ I)^{-1} A^T b \tag{2.17}$$

where, I = Identity matrix (n × m); α I = Ridge term

Advantages ridge term,

i Facilitates invertibility of resultant matrix and it gets added to the principle diagonal

ii consistently achieves a unique solution

The equation 2.8 can be interpreted geometrically as shown below:

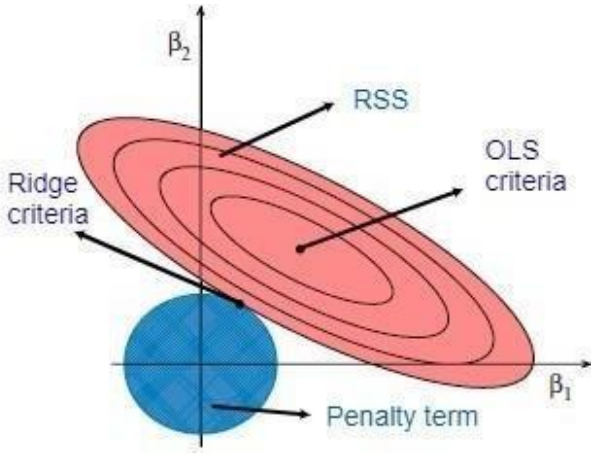


Fig. 14: Geometric representation of ridge regression [19]

The Fig.14 clearly depicts the aim of ridge (L2-regularization) regression i.e., minimization occurs simultaneously between the RSS (ellipse) and the penalty term (circle) mentioned in equation 2.8. The simultaneous minimization occurs at "$\hat{x}_{ridge}$" shown in equation 2.17.

3.4.2 LASSO

LASSO aims to minimize the ordinary least square with an added penalty term [20]. In case of L1- regularization, the penalty term is the sum of the absolute value of the regression coefficients. Hence LASSO is also known as the L1-regularization [21] .

$$\underset{x\in\mathbb{R}^m}{\operatorname{argmin}} \sum_{i}^{n} \left(\sum_{j}^{m} a_{ij} x_j - b_i^2 \right)^2 + \alpha \sum_{j}^{m} |x_j \tag{2.18}$$

The equation 2.7 can be represented in a simpler form as,

$$\underset{x\in\mathbb{R}^m}{\operatorname{argmin}} \underbrace{||Ax-b||_2^2}_{Residual} + \alpha \underbrace{|x|_1}_{Penalty} \tag{2.19}$$

where, $b \in R^n$ = Response vector; $A \in R^{n\times m}$ = Predictor matrix; α = Regularization parameter

The first part of the derivation is similar to L2-regularization, in matrix notation equation 2.8 becomes,

$$C_{lasso} = (A\,x - b)^T\,(A\,x - b) + \alpha\,|x|_1 \tag{2.20}$$

Expanding and simultaneous simplification of equation 2.20 results in the following,

$$C_{lasso} = x^T A^T A\, x - b^T A\, x - x^T A^T b + b^T b + \alpha\ |x|_1 \tag{2.21}$$

$$= x^T A^T A\, x - x^T A^T b - x^T A^T b + b^T b + \alpha\, |x|_1$$

$$C_{lasso} = b^T b - 2\, x^T A^T b + x^T A^T A\, x + \alpha\, |x|_1$$

The equation 2.7 can be represented in a simpler form as,

$$\underset{x \in \mathbb{R}^m}{\operatorname{argmin}} \underbrace{||Ax - b||_2^2}_{Residual} + \alpha \underbrace{|x|_1}_{Penalt} \tag{2.22}$$

where, $b \in R^n$ = Response vector; $A \in R^{n \times m}$ = Predictor matrix; α = Regularization parameter

The first part of the derivation is similar to L2-regularization, in matrix notation equation 2.8 becomes,

$$C_{lasso} = (A\, x - b)^T\, (A\, x - b) + \alpha\, |x|_1 \tag{2.23}$$

Expanding and simultaneous simplification of equation 2.23 results in the following,

$$C_{lasso} = x^T A^T A\, x - b^T A\, x - x^T A^T b + b^T b + \alpha\ |x|_1 \tag{2.24}$$

$$= x^T A^T A\, x - x^T A^T b - x^T A^T b + b^T b + \alpha\, |x|_1 \tag{2.25}$$

$$C_{lasso} = b^T b - 2\, x^T A^T b + x^T A^T A\, x + \alpha\, |x|_1 \tag{2.26}$$

Next, taking the derivative of equation 2.26, we get,

$$\nabla C_{lasso} = -2A^T b + 2A^T Ax + \nabla(\alpha\, |x|_1) \tag{2.27}$$

Due the face that equation 2.24 consists of the term "$\nabla(\alpha\, |x|_1)$", sub-differential helps us to arrive at the final solution. But before we step into sub-differential, let us assume that the $A^T A$ is equal to I and multiply "2" to the penalty term.

Equation 2.27 becomes,

$$\nabla C_{lasso} = -2A^T b + 2x + 2\nabla(\alpha\, |x|_1) \tag{2.28}$$

Now, the sub-differential becomes,

$$\nabla(C_{lasso}) = \begin{cases} 2x - 2A^T b + 2\alpha, & x > 0 \\ [-2\alpha, 2\alpha] - 2A^T b, & x = 0 \\ 2x - 2A^T b - 2\alpha, & x < 0 \end{cases} \tag{2.29}$$

Breaking down each of the 3 conditions mentioned in equation 2.29,

Case 1: when $x > 0$ $\quad 2x - 2A^T b + 2\alpha = 0$ (2.30)

Equation 2.30 must be satisfied.

Therefore, we get, $\quad x = 2A^T b - \alpha$ (2.31)

Case 2: when $x = 0$

$$0 \in [-2\alpha, 2\alpha] - 2A^T b \tag{2.32}$$

Therefore, we now have 2 sub-cases, i.e,

$$-2\alpha - 2A^T b < 0 \Rightarrow \alpha > -A^T b \tag{2.33}$$

$$2\alpha - 2A^T b > 0 \Rightarrow \alpha > A^T b \tag{2.34}$$

The sub-cases mentioned in equation 2.34 becomes,

$$\alpha > A^T b \tag{2.35}$$

when $x = 0$

Case 3: when $x < 0$

$$2x - A^T b - 2\alpha = 0 \tag{2.36}$$

Equation 2.36 must be satisfied.

Therefore, we get,

$$x = A^T b + \alpha \tag{2.37}$$

The aforementioned cases help us to arrive at the solution for LASSO and it summarized in the equation below,

$$\frac{\partial C_{ridge}}{\partial x}\hat{x}_{lasso} = \begin{cases} 0, & x_j > \left|A^T b\right| \\ A^T b - sign(A^T b) \cdot \alpha, & x_j \leq \left|A^T b\right| \end{cases} = 0 \tag{2.38}$$

The equation 2.22 can be interpreted geometrically as shown below:

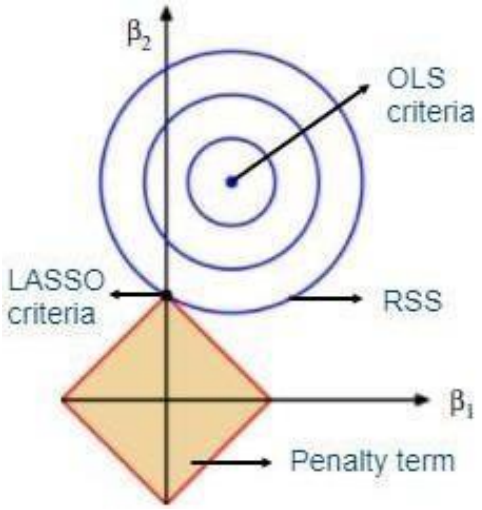


Fig. 15: Geometric representation of LASSO regression [22]

The Fig.15 clearly depicts the aim of LASSO (L1-regularization) regression i.e., minimization occurs simultaneously between the RSS (circle) and the penalty term (square) mentioned in equation 2.19. The simultaneous minimization occurs at "x^$_{lasso}$" shown in equation 2.38.

Table 1: Comparison between L1 and L2 regularization

Properties	**L1 regularization**	**L2 regularization**
Robustness	Penalty term Absolute value of coefficients $\boxed{|x|_1 \Rightarrow}$ Outliers are affected linearly This method is more robust	Penalty Term square of the coefficients $\boxed{\|x_2^2\|_2^2 \Rightarrow}$ Outliers are affected exponentially This method is less robust
Computational effort	Penalty term $\boxed{|x|_1 \Rightarrow}$ Non-differentiable term This method requires more computational effort	Penalty Term $\boxed{\|x_2^2\|_2^2 \Rightarrow}$ Closed form of solution Solution are obtained by using matrix form This method requires less
Sparsity	This method has the ability to shrink coefficients to zero Sparse solution.	This method spreads the error hindering sparsity

4.0 CONCLUSION

The fundamentals of large-scale statistics was focused with retrieving the information from noisy data in the present review article. The method of total variation regularization helps to study thoroughly and understand the concept behind regularization parameter on various test functions each at different amplitude of noise. The study behind the optimal parametervalue shines light on the fact that a stronger noise level in a large-scale dataset requires considerably strong optimal parameter. As we know that, in the real-life problems it is very difficult to define noise from the actual measurement data which needs the iterative process to automatically obtain regularization parameter. The information being tracked was implemented in the process of finding differential equations by using data-driven (or Sparse Regression) method.

5.0 COMPETING INTERESTS

Authors have declared that no competing interests exist.

6.0 REFERENCES

1. Rani NS, Rao PS, Anurag. Study an analysis of noise effect on big data analytics. Int. J. Management, Technology and Engineering. 2018; 8(XII):5841-5850
2. Chartrand R. Numerical differentiation of noisy, no smooth data. ISRN Applied Mathematics. 2011; 2011: 1–11(doi:10.5402/2011/164564).
3. Hansen PC. Analysis of discrete ill-posed problems by means of the l-curve. SIAM Review. 1992; 34(4): 561–580.
4. Belge M, Kilmer ME, Miller EL. Efficient determination of multiple regularization parameters in a generalized l-curve framework. Inverse problems. 2002; 18(4): 1161– 1183.
5. Hansen PC. Kilmer ME. A parameter-choice method that exploits residual information. PAMM. 2007; 7(1):1021705–1021706.
6. Rust BW, Dianne PO. Residual periodograms for choosing regularization parameters for ill-posed problems. Inverse Problems. 2008; 24(3):034005.
7. Jansen M, Malfait M, Bultheel A. Generalized cross validation for wavelet thresholding. Signal Processing. 1997; 56(1):33–44.
8. Rudy SH, Brunton SL, Proctor JL, Kutz, JN, Data-driven discovery of partial differential equations. Science Advances. 2017; 3(4):e1602614.
9. Hariri RH, Fredericks EM, Bowers KM. Uncertainity in big data analysis: Survey, opportunities and challenges. J.Big Data. 2019; 6:44 (https://doi.org/10.1186/s40537-019-0206-3).
10. Kumar RK, Chadrasekaran RM. Attribute correction-data cleaning using association rule and clustering methods. Int. J. Data Mining & Knowledge Management Process. 2011; 1(2): 22-32.
11. Ogu AI, Inyama SC, Achugamonu PC. Methods of Detecting Outliers in A Regression Analysis Model. West African J. Industrial and Academic Research. 2013; 7(1): 105-113
12. Martinez WL, Martinez AR, Solka J. Exploratory Data Analysis with MATLAB, Second Edition. Chapman & Hall/CRC. ISBN 9781439812204; 2010.

13. Mahmoudi A. Adaptive Algorithm for Estimation of Two-Dimensional Autoregressive Fields from Noisy Observations. Int. J.Stochastic Analysis. vol. 2014; Article ID 502406, 7 pages.
14. Alexander Ch. Sadiku M, Fundamentals of electric circuits. Fifth Edition. McGraw-Hill; 2013.
15. Bubba TA, Porta F, Zanghirati G, Bonettini S. A nonsmooth regularisation approach based on shearlets for Poisson noise removal of ROI tomography. Applied Mathematics and Computation. 2018; 318:131-152
16. Sliney DH. What is light? The visible spectrum and beyond. Eye (Lond). 2016; 30(2): 222–229.
17. Song S, Chandhuri K, Sarwate AD. Learning from data with heterogeneous noise using SGD. JMLR Workshop Conf.Proc., 2015; pp 894-902
18. Srivastava N, Hinton G, Krizhevsky A, Sutskeverilya E, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Over fitting. J. Machine Learning Research. 2014; 15:1929-1958.
19. Wieringen WNV, Lecture notes on ridge regression – version 0.20. 2018; https://arxiv.org/pdf/1509.09169.
20. Chang H, Zhang D. Machine learning subsurface flow equations from data. Computational Geosciences. 2019 (http://doing.org/10.1007/s10596-019-09847-2).
21. Hastie T, Tibshirani R, Wainwright M. Statistical Learning with Sparsity: The Lasso and Generalizations, CRC Press; 2015.
22. Jiang Y, Yunxiao H, Zhang H. Variable selection with prior information for generalized linear models is the prior lasso method. J. American Statistical Association. 2016; 111 (513): 355–376.