e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

DETERMINATION OF OPTIMAL REGULARIZATION PARAMETER AND ITS IMPACT ON INFORMATION TRACKING

Avinash Bapu Sreenivas[*1], **T. Srisupattarawanit**[*2], **H. Ostermeyer**[*3]

[*1]M.Sc. Graduate for Dynamics and Vibration, Technical University of Braunschweig, Germany.

[*2,3]Professor Institute for Dynamics and Vibration, Technical University of Braunschweig, Germany.

ABSTRACT

The derivative of a function in the field of engineering has numerous applications, one of the major application is in the optimization process. These functions are generally specified by a large-scale dataset in almost all engineering disciplines. In many instances, the function may be obtained from a large-scale dataset that consists of some type of noise. The analysis of noisy function has ramification on the entire process of data analysis specifically during numerical differentiation. The reason behind this is that the derivative of noisy function tends to amplify the already present noise. Hence hampering the process and it is reflected in the solution of the same. Over-fit condition is a common occurrence while fitting a noisy function. This can be overcome by an additional penalty term. This is known as regularization. The derivatives are slightly treated in a different manner, i.e., the numerical differentiation process itself is regularized to minimize the amplification of noise. This method is known as total variation regularization and it forms the basis of this research article. The successful implementation of this method mainly revolves around a factor known as regularization parameter. Hence, we employ various methods to determine the optimal regularization parameter for different test functions and noise levels. Some of the methods require the user to provide a range of regularization parameter and based on that and various philosophies behind each method, they provide the optimal value. Therefore, optimization of these methods is vital. This forms the aim of the research article.

KEYWORDS: Amplification of Noise, Total Variation Regularization, Optimal Regularization Parameter.

I. INTRODUCTION

Experimental data is an essential part of modern-day engineering interdisciplinary. These experimental data may consist of errors. Randomness and non-correlation are the properties of the aforementioned error that renders it completely unpredictable in nature. These random and non-correlated errors are termed as noise. Hence the knowledge behind these errors/noise, its impact on data analysis process, proper handling and removal techniques are prioritized during the early phase of data analysis[1, 2]. Numerical methods in order to approximate the derivatives of functions such as finite- difference methods have stolen the limelight in many engineering interdisciplinary for optimization purposes. Differentiation of noisy dataset leads to the amplification of already present noise. These amplification in the derivatives can be overcome by applying Total Variation (TV) regularization technique [1]. 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[1]. The successful implementation of this methods hinges on one aspect, i.e., clearly understanding and determination of regularization parameter. There are various methods that facilitate 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, 4]. 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 selected as the optimal regularization parameter by using curvature plot [5]. Another method known as Normalized Cumulative Periodogram (NCP) bases its philosophy on extensive analysis of residual vector. Optimal regularization parameter is selected based on Kolmogorov-Smirnov test i.e., the cumulative periodogram must strictly lie within the confidence interval of 95% [4]. These optimal parameters can then be used to extract vital information from a noisy dataset which facilitates a more accurate data analysis process. With this background, an attempt has been made in this research article to investigate the implications of noisy data and also regularization of noisy data in order to retrieve vital information.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

II. METHODOLOGY

Total Variation Regularization is a common technique used in the field of engineering and scientific computing. The basic principle of this method [1] is to determine the derivative of function "f", which can be obtained by minimizing the following equation,

$$F(u) = \alpha R(u) + DF(Au - f) \quad (1)$$

where,

R(u)	=	Regularization or penalty term
A(u)	=	Anti-differentiation term which is given by
Au(x)	=	$\int_0^x u$
α	=	Regularization parameter
DF(Au-f)	=	Data Fidelity term

Note:

- The role of regularization term is to penalize the irregularities in the "u" (solution)
- The role of the data fidelity term is to penalize the discrepancy between Au and f
- The role of the regularization parameter is to maintain a balance between regularization and data fidelity terms

Now, equation 1 is the general form of a regularization technique. In total variation regularization the derivative of the function is determined by minimizing the functional [1] of the length [0,L] as shown below,

$$F(u) = \frac{1}{2}\int_0^L \|Au - f\|_2^2 + \alpha \int_0^L |u'|_1 \quad (2)$$

where,

F	=	Functional defined on bounded variation $[0, L]$
f	=	Given function & $f \epsilon L^2$
u	=	Solution
α	=	Regularization parameter

2.1 Implementation

In section 2, our aim is to determine the derivative of "f" by minimizing the equation 2. This can be achieved by using the gradient descent method [6]. Gradient descent is a method employed to determine the local minimum of function. The aim of gradient descent is to initialize the step size that is proportional to the negative of the approximated gradient of the given function at each current iteration point. Therefore, for each iteration the gradient slowly tends towards the local minimum of the given function and once the convergence criteria is satisfied, the point at which this occurs is labeled as final value or minimum value [6] of the given function as shown in figure 1.

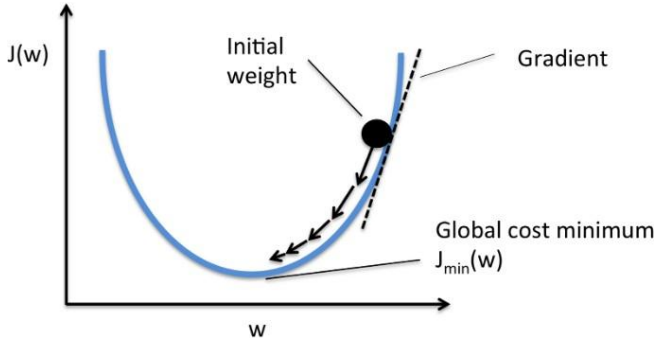

Fig-1: A graph depicting the convergence of a function using gradient descent method [7]

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

Gradient descent method uses the principle of Euler-Lagrange differential equation. A general form of Euler-Lagrange equation is represented in the following equation,

$$J = \int f(t, y, \dot{y}) \tag{3}$$

where,

J = Functional

f = Given function depending on $(t, y, \dot{y})$

$\dot{y}$ = $\frac{dy}{dt}$

If the following Euler-Lagrange differential equation is satisfied, then "*J*" has a stationary value,

$$\frac{\partial f}{\partial y} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{y}}\right) = 0 \tag{4}$$

Note:

The time derivative in equation 4 can be replaced by space derivative as shown below,

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial \dot{y}}\right) = 0 \tag{5}$$

The Euler-Lagrange differential equation using space derivative [8], as shown in equation 5 is applied to the total variation regularization functional in equation 2. This leads to the following system of equations in order to understand the gradient descent method.

$$\partial F(u) = A^T(Au - f) - \alpha \frac{d}{dx}\left(\frac{u'}{|u'|}\right) \tag{6}$$

The above equation can be solved by using steepest gradient descent method. In order to achieve the aim of method, the minimum can be determined by applying the following condition,

$$\frac{\partial u}{\partial t} = -\partial F(u) \tag{7}$$

Therefore, we arrive at the final equation,

$$\frac{du}{dt} = \alpha \frac{d}{dx}\left(\frac{u'}{|u'|}\right) - A^T(Au - f) \tag{8}$$

Note:

- In equation 8, in order to overcome the problem of getting undetermined solution (division by zero) the denominator $|u'|$ is substituted with $\sqrt{|u'|^2 + \epsilon}$ The characteristics of are as follows,
  - $\epsilon > 0$
  - very small constant
- Numerical approximation of the solution in 8 is determined by employing the explicit method.
- Discretization of u_t in equation 8 is performed using forward difference method for a fixed time step Δt as show below,

$$\frac{u_{n+1} - u_n}{\Delta t} \tag{9}$$

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

One of the disadvantages of using the gradient descent method for the determination of the minimum of the functional in equation 2 is the slow rate of convergence. Hence, to overcome the aforementioned disadvantage the nonlinear differential operator $u \rightarrow \left(\frac{d}{dx}\right)\left(\frac{u'}{|u'|}\right)$ is substituted with a linear operator $u \rightarrow \left(\frac{d}{dx}\right)\left(\frac{u'}{|u'_n|}\right)$ for each iteration in equation 8.

2.2 A brief insight about Lagged diffusivity fixed point method

We are now going to discuss the method of lagged diffusivity for smaller problems [1] in detail. The three principles of this method are,

i) "u" is constructed on a uniform grid i.e.,
$\{x_i\}_0^L = \{0, \Delta x, \Delta 2x, \Delta 3x \dots \dots \dots L\}$
ii) Derivative of "u" is determined halfway between the grid using the forward difference method i.e.,
$Du(x_i + \Delta x/2) = u(x_{i+1}) - u(x_i)$
iii) Similarly, the integral of "u" is determined halfway between the grid using trapezoidal rule

The table 1 summarizes the formula employed during various phases in the lagged diffusivity fixed point algorithms and the pseudo code is presented in 2.1.

Variable Name	Formula
E_n	$\sqrt{((u_n(x_i) - u_n(x_{i-1}))^2 + \epsilon)}$
L_n	$\Delta x D^T E_n D$
H_n	$K^T K + \alpha L_n$
g_n	$K^T(Ku_n - f) + \alpha L_n u_n$

Table 1: A Table Summarizing The Important Formula Required In The Lagged Diffusivity Method

By utilizing the formula in table 1, the updated value shown in equation 10 forms the solution to equation 11required over each iteration point,

$$s_n = u_{n+1} - u_n \tag{10}$$

where,

$$s_n = -H_n^{-1} g_n \tag{11}$$

Algorithm 2.1: Lagged diffusivity fixed point pseudo code

```
1. Niter ← Initialize        %Specify number of iteration
2. α ← Initialize            %Specify regularization parameter
3. u ←[0;diff (Data);0]      % Naive derivative
4. for n ← 1:Niter
        g_n ← K^T(K u_n − f) + α L_n u_n      % Gradient
        H_n ← K^T K + α L_n                   % Cost function (Hessian approximation)
        s_n ← −H^{−1} g_n   % Determined using preconditioned conjugate gradient
        u_n ← u_n − s_n     % Update
   end
```

Note:

- The second and third principles of lagged diffusivity method leads to the formation of a differentiation matrix" D" and" A" of the size $L \times (L + 1)$ respectively.
- The advantage of this method is that it prevents the need to deal with the boundary conditions required during differentiation process.
- Computationally, this method provides better results.

2.3 Importance of Total Variation Regularization

This section allows us to understand the importance employing TV regularization. TV regularization can be applied to a process irrespective of the type of noise it contains. Hence, this method can be employed in various fields. Generally the regularization or penalty term shown in equation 1 is given by,

$$R(u) = \int_0^L |u'|^2 \tag{12}$$

where,

R(u) = Regularization or penalty term

Equation 12 constraints the minimizer to be continuous. Hence it leads to inaccurate differentiation of the given function. In order to overcome and avoid the aforementioned difficulties, the total variation regularization method represented in equation 2 is employed [1]. The advantages of this methods are described below,

- This method helps to keep a check on the noise present in the data as it has a large total variation.
- Unlike in equation 12, total variation regularization considers the (jump) discontinuities.
- Total variation regularization facilitates the computation of discontinuous derivatives and characterizes the noisy data clearly.

Note:

- The regularization parameter plays a vital role in balancing the regularization term and data fidelity term as explained in equation 1.
- The choice of the regularization parameter is generally based on" eye-balling" technique for very small data set.
- Real world application usually deals with large data set, hence the" eye-balling" technique is computationally time consuming and reduces the efficiency of the process.

Due to the disadvantages of" eye-balling" technique, automated techniques are designed and understood in order to improve the efficiency and reduce the computational time for large data set.

2.4 Determination of optimal regularization parameter

There are 2 major techniques which are employed to overcome the aforementioned problem and they are,

- L-curve method
- Normalized Cumulative Periodogram (NCP) method

A brief explanation of the above 2 techniques are discussed below,

1. **L-curve method:**
   The" L-curve" method is implemented by using the principles of the total variation regularization as explained in section 2. The solution is determined for each of the user defined regularization parameter,

$$\alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4 \dots \dots \dots \alpha_n \tag{13}$$

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

This solution is denoted as "u_α ". Then, for each of the respective"u_α " residual between the solution and given function is estimated. This is represented by " $Au_\alpha - f$ ".
Finally, the L1-norm of the solution i.e., "$|u_\alpha|_1$" and the L2-norm of the residual, i.e., " $\|Au_\alpha - f\|_2^2$" are determined. A graph of $(\|Au_\alpha - f\|_2^2, |u_\alpha|_1)$ is plotted. This results in graph characterized by the" L-curve" as shown in figure 2.

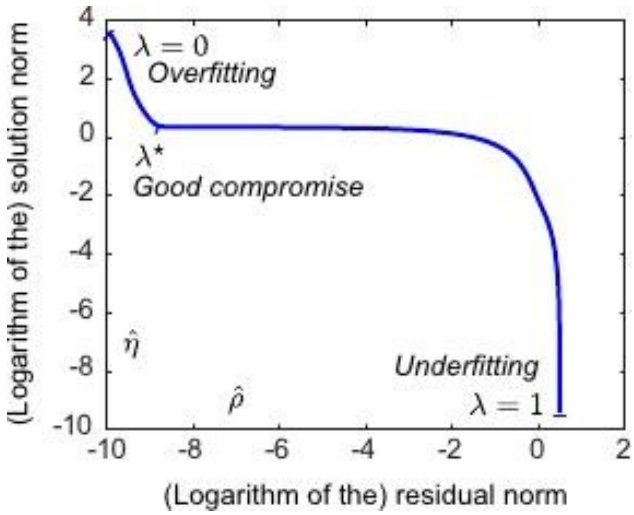

Fig-2: A Graph Showing The General Form Of The "L-Curve" [9]

As seen the figure 2, there is clear representation of" corner". The role of the" corner" is to maintain balance between over-fit and under-fit condition. The corner also is an indication of regularization and perturbation errors [10]. Finally, the aim of this method is to extract the optimal regularization parameter. This aim is now achieved by the determination of the curvature using the following equation 14 for the figure 2.

$$\hat{C} = 2\frac{\xi\rho}{\xi'}\left[\frac{\alpha^2\xi'\rho + 2\alpha\xi\rho + \alpha^4\xi\xi'}{(\alpha^2\xi^2 + \rho^2)^{\frac{3}{2}}}\right] \tag{14}$$

where,

$\hat{C}$ = Curvature
ξ = $|u_\alpha|_1$
ρ = $\|Au_\alpha - f\|_2^2$
ξ' = $\partial\xi/\partial\alpha$

Note:

Equation 14 deals with the determination of "ξ" which signifies the derivative with respect to "α", hence making it difficult to calculate the curvature values. In order to overcome the aforementioned complexity, we employ an alternative method (based on coordinates) which is described below,

$$a = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$
$$b = \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2}$$
$$c = \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}$$
$$A = 0.5 \cdot |(x_1 - x_2) \cdot (y_3 - y_2) - (y_1 - y_2) \cdot (x_3 - x_2)|$$

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

$$\hat{C} = \left(\frac{4A}{a \cdot b \cdot c}\right) \tag{15}$$

The entire process of" L-curve" method is summarized in the following flowchart 3.

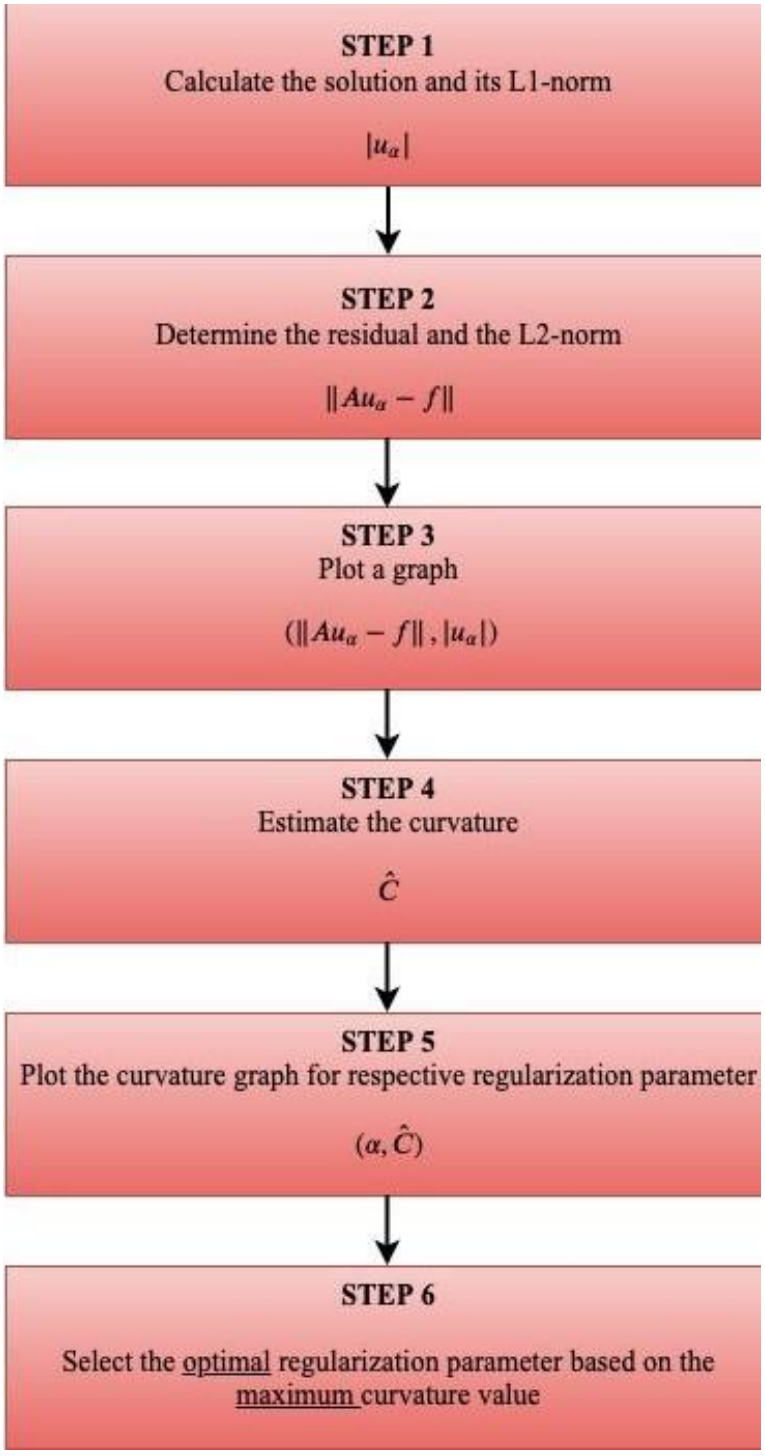

Fig-3: A Flowchart Describing The Process Involved In L-Curve Method

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

The table 2 below summarizes the method by brief indicating the pro and cons of employing the L-curve method.

Advantages	Disadvantages
Highly robust method	Special cases: Presence of 2 "corners" namely a global corner and" New corner". One must then check manually for better yielding solution
Clear indication of over-fit and under-fit condition	Increase in problem size ("*n*") leads to over-regularization or large parameter value.
"corner" also indicates a balance between the two conditions	

Table-2: Table Showing A Brief Advantages And Disadvantages Of L-Curve Method

2 Normalized Cumulative Periodogram (NCP):

NCP solely requires the determination of residual vector (r) as this methods aims to perform statistical analysis that focuses completely on the determined residual [4, 11]. NCP methods helps to isolate the noise from relevant information in the dataset. Regularization parameter which succeeds at achieving the aforementioned task is chosen as the optimal parameter [10]. The aforementioned residual vector can be simply represented as follows,

$$r = Au_\alpha - f \tag{16}$$

Based on the perception of the residual vector, we can draw the following conclusion,

i) All information in the given function" f" is not extracted when" α" is large.
ii) Only noise remains in the residual vector when" α" is small.

Hence, emphasis is laid completely upon the determination of the optimal regularization parameter, which in turn helps to extract all the relevant information from the noisy data or function (f). General form of the NCP [4] is as follows,

$$\hat{r} = fft(r) \tag{17}$$

where,

r = Residual vector
$fft(r)$ = discrete Fourier transforms of the residual vector

Then the periodogram of residual vector, "" is represented as,

$$p = \left(|\hat{r_1}|^2, |\hat{r}_2|^2, |\hat{r}_3|^2, \dots \dots |\hat{r}_q|^2\right)^T \tag{18}$$

where,

q = $[n/2\,] + 1$
n = Length of residual vector "r"

NCP for residual vector "r" is shown below,

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

$$c(r)_k = \frac{\|p(2:k+1)\|_1}{\|p(2:q)\|_1} \qquad k = 1,2,3 \ldots\ldots\ldots q-1 \tag{19}$$

where,

$c(r)$ = NCP of residual vector "r"
p = Periodogram or power spectrum of residual vector "r"

Upon the derivation of "$c(r)_k$" in equation 19 from the general form of NCP shown is equation 17, our focus now shifts towards achieving the aim of NCP. The main objective of NCP method is to determine whether the residual contains only" white noise". A graph of $(k\,, c(r)_k)$ is plotted. The regularization parameter which lies closest to the Gaussian white noise characteristics line (dotted black line in figure 8) is chosen as the optimal regularization parameter.

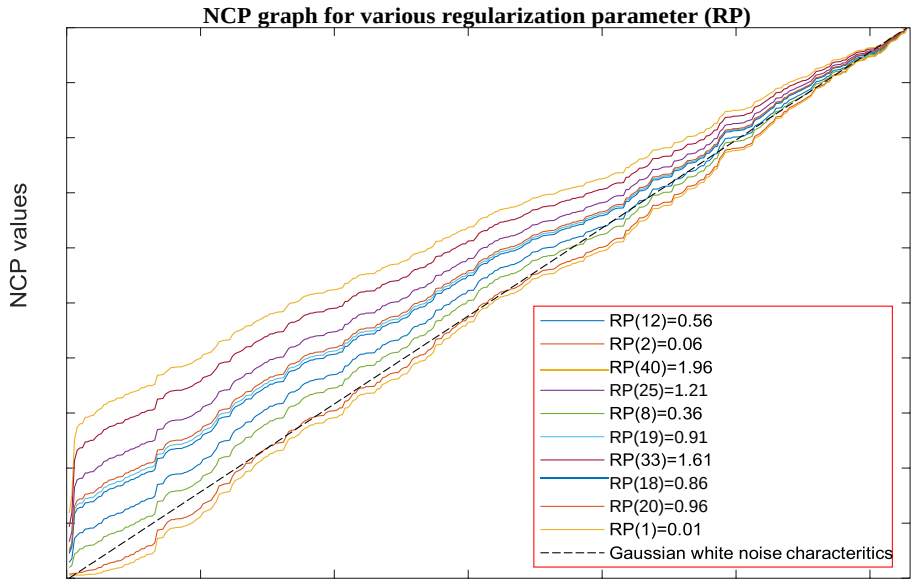

Fig-4: A Graph Showing The Behavior Of NCP Values For Ten Different Regularization Parameter

The table 3 below summarizes the method by brief indicating the pro and cons of employing the NCP method.

Advantages	Disadvantages
Computationally inexpensive	Deals mainly with white noise • Constant spectral density • Random errors independent to each other
Deals with only residual vector • more stable for high regularization parameter	Dealing with different types of noise • Covariance matrix must be know

Table-3: Table Showing A Brief Advantages And Disadvantages Of NCP Method

III. RESULTS

The important terminology and the methods explained in section 2 provides a strong foundation to apply the acquired knowledge on various test functions. This section is segregated into various case studies, each dealing with different aspects and its results are summarized. In order to achieve satisfactory results using TV regularization, we must first determine the optimal regularization parameter explained in section 2.4. To begin with, we consider and perform analysis on 2 different test functions. They are described below and summarized

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

in tables 4 and 5. Gaussian white noise of a specific standard deviation is added to a function" f " which yields a noisy "y" function. We then implement various methods to determine the optimal regularization parameter which helps in extracting vital information from a noisy dataset (E.g. measurement data)

Note: All the graphs were generated using MATLAB®.

3.1 Test function 1

Data points	$x = 0 : 0.01 : 5$
Given function	$\|x - 2.5\|$
Standard deviation of AWGN	0.05
Noisy function	$y = f + \eta$

Note:
The variable" η" is AWGN noise (vector) which is generated using" randn" a MATLAB® syntax.

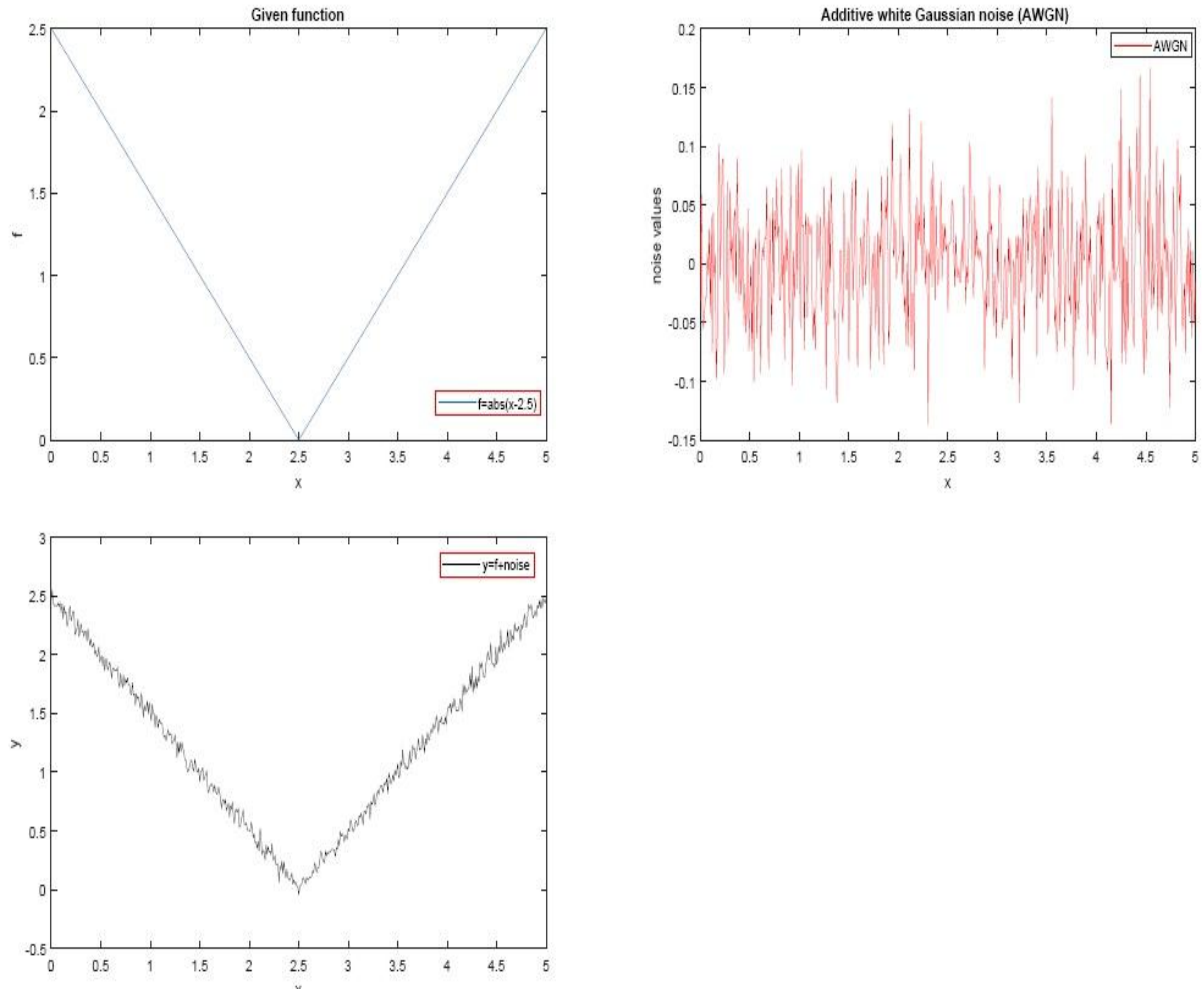

Fig-5: Graphs Depicting The Respective Information Provided In Table 4

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

3.1.1 Eye-balling method

It is also known as trial-and-error method. The user needs to start with an initial value and check if the solution is either over-fit or an under-fit condition. This suggest whether one must increase or decrease the choice of regularization parameter. This process continues until a certain value that provides a good fit is selected. This does not ensure that the selected parameter is optimal. This method proves to be tedious for a user when dealing with large scale data and also an unfeasible technique. The results of a few parameters are shown below,

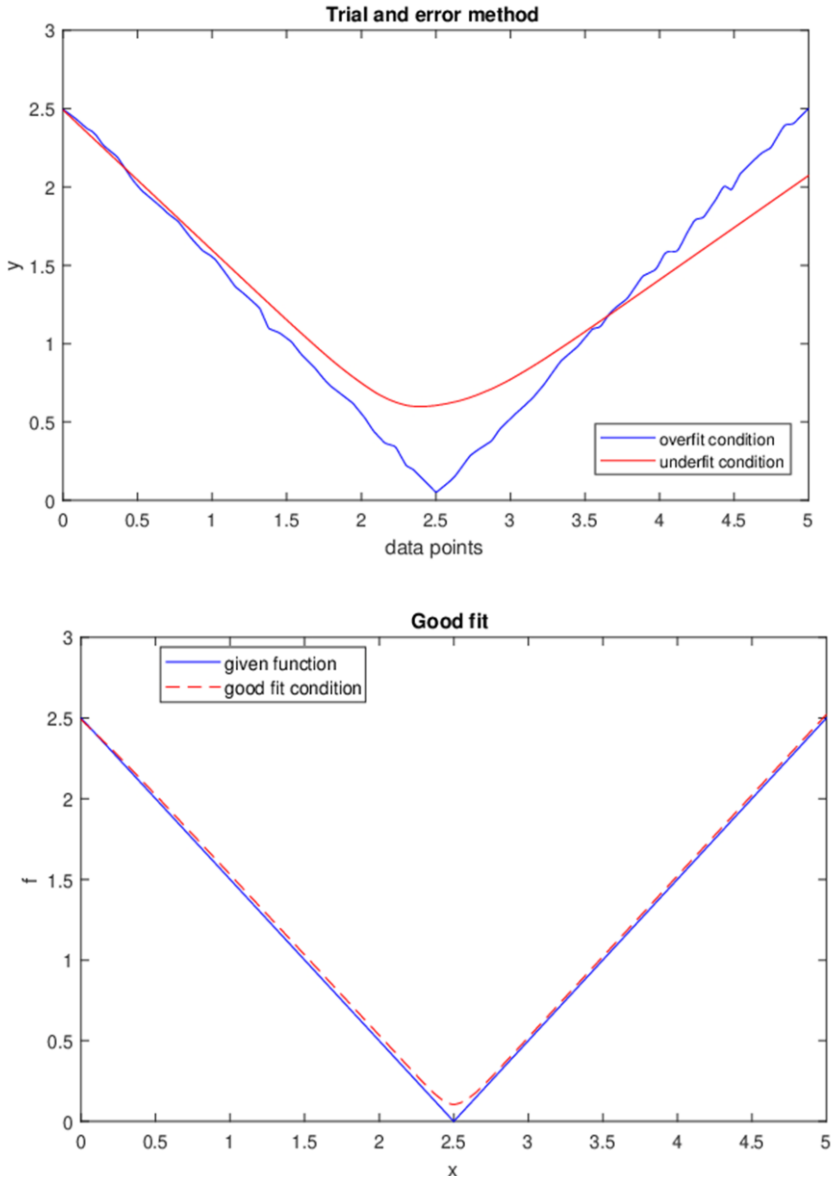

Fig-6: Graphs depicting the over-fit and under-fit condition (a) and good fit (b) for test function 1

3.1.2 L-Curve method

We now implement the L-curve method and its curvature plot as explained in section 2.4 to determine the optimal regularization parameter that helps to balance the under-fit and over-fit condition. The results are shown below,

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

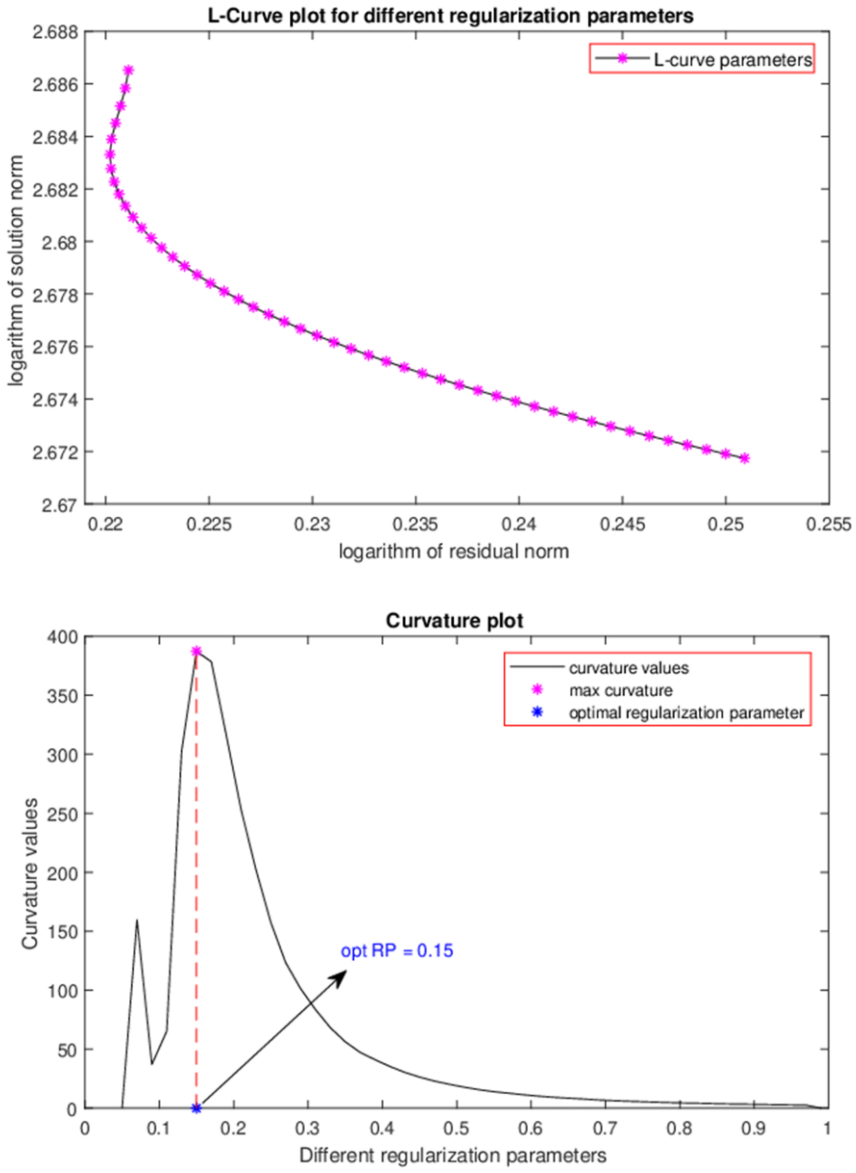

Fig-7: L-curve results for test function 1

3.1.3 NCP

Now, we implement NCP method as explained in section 2.4 to determine the optimal regularization parameter. The graph below depicts the NCP values for various regularization parameter for test function 1. As seen below, this graph alone renders it impossible to choose an optimal parameter.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

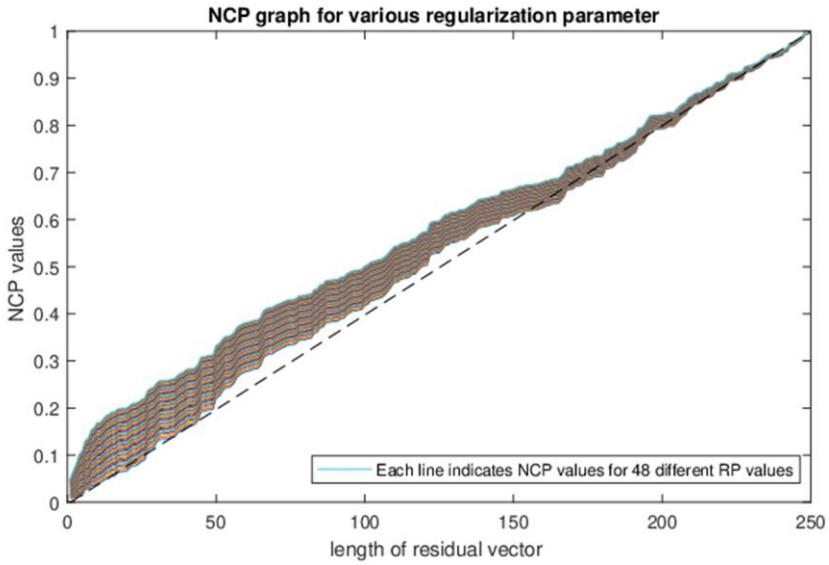

Fig-8: A graph showing the behavior of NCP values for various different regularization parameters

The figure 8 makes our inference regarding the NCP a bit tedious. Hence, we isolate the optimal regularization parameter from our entire set of regularization parameter and this is shown in figure 9.
The selection of the optimal value can be performed by determining the L2-norm between Gaussian white noise characteristic line (dotted black line in figure 9) and various NCP values for respective regularization parameter. One with the least L2-norm can be selected as optimal because this particular line (blue line in figure 9) lies closest to the desired.

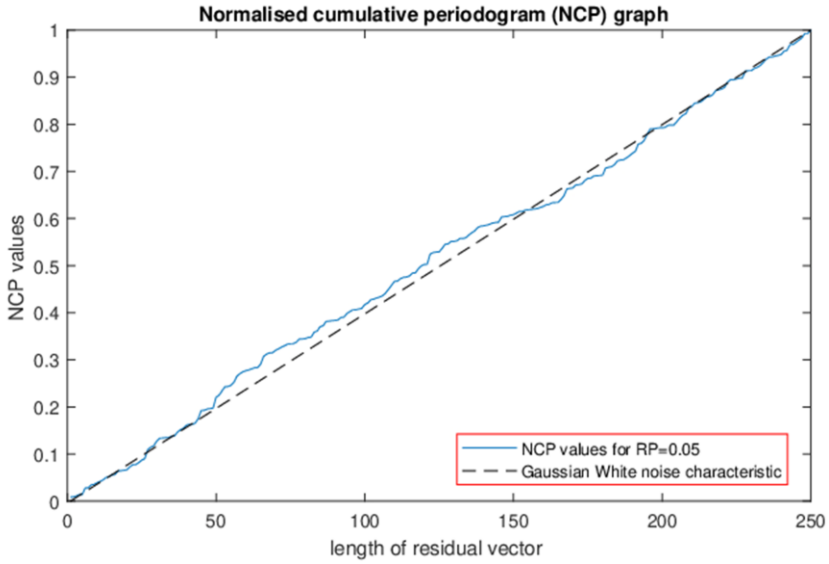

Fig-9: A Graph Showing The Behavior Of NCP Values For The Optimal Regularization Parameter

Finally, we employ the optimal value from L-curve method in TV regularization to obtain numerically determined function. As the L-curve method provides numerical solution of a much higher accuracy in comparison to NCP method. This is shown in the figure below,

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

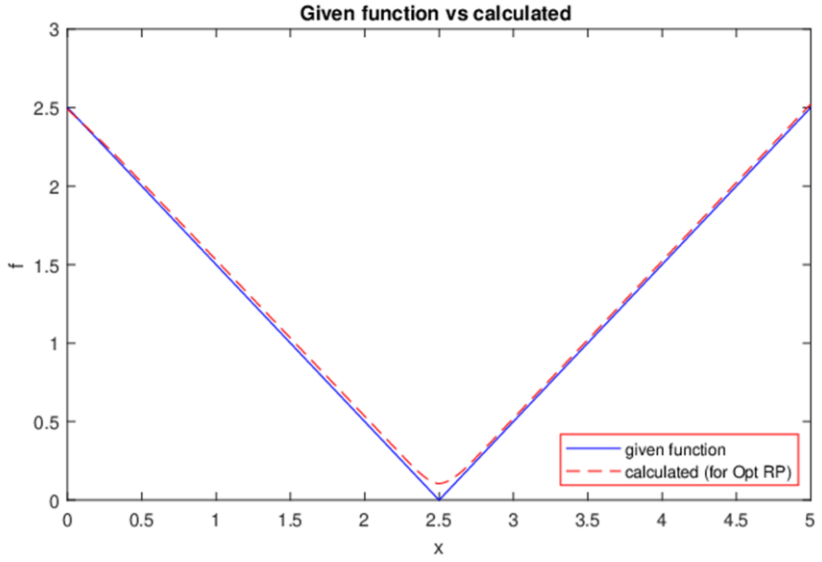

Fig-10: A graph showing the given function and numerically determined function

3.2 Test function 2

Data points	$x = 0 : 0.01 : 5$
Given function	$\exp(x)$
Standard deviation of AWGN	0.5
Noisy function	$y = f + \eta$

Table-5: Given information for test function 2

Note:

1. The variable" η" is AWGN noise (vector) which is generated using "randn" a MATLAB® syntax.
2. Since we are dealing with exponential function, the standard deviation should be higher for AWGN in order to affect the test function significantly. Hence, we choose a much higher standard deviation for test function 2 in comparison to test function 1 & 2.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

3.2.1 Eye-balling Method

Similar to the explanation in 3.1.1, this method is inefficient in nature. Few results are shown below,

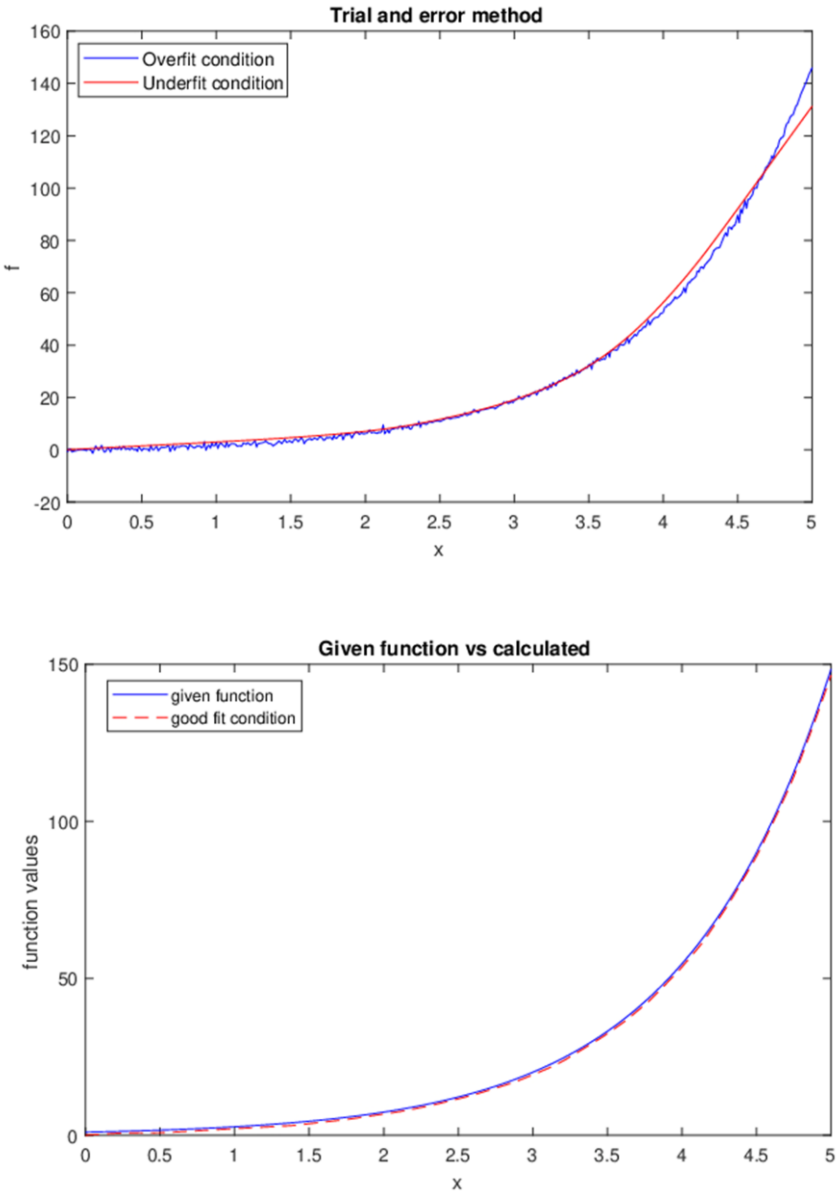

Fig-11: Graphs depicting the over-fit and under-fit condition (a) and good fit (b) for test function 2

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

3.2.2 L-curve method

The results for L-curve and its curvature plot are shown in figure 12 and 13 respectively.

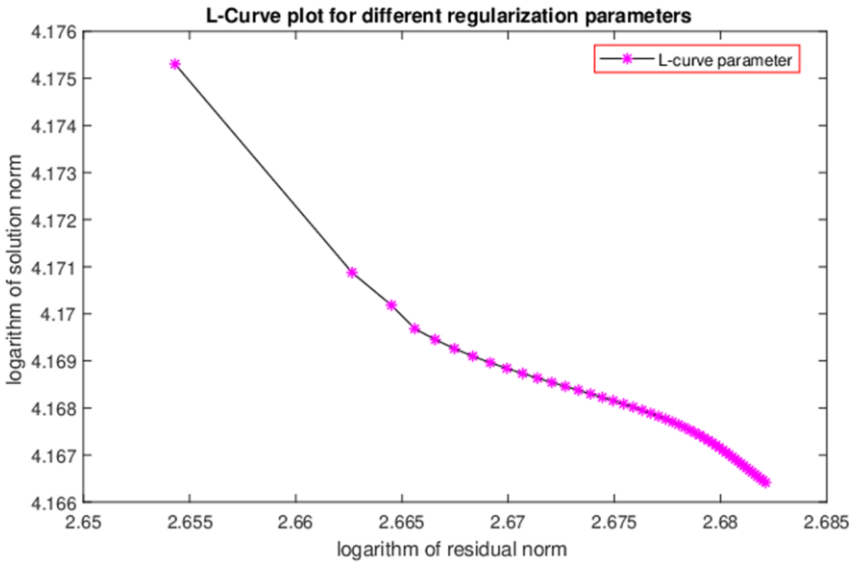

Fig-12: A graph representing L-curve

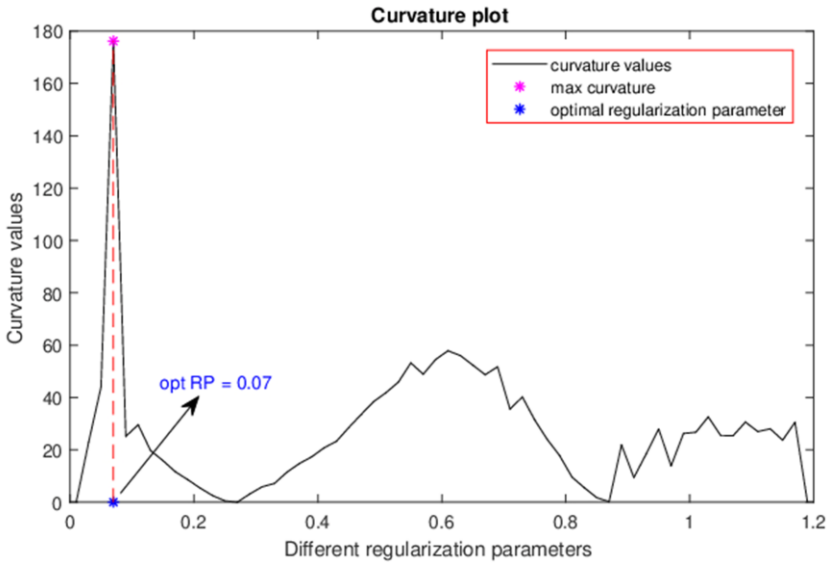

Fig-13: A graph showing curvature plot

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

3.2.3 NCP method

The below clearly depicts the nature of NCP values for various regularization parameter.

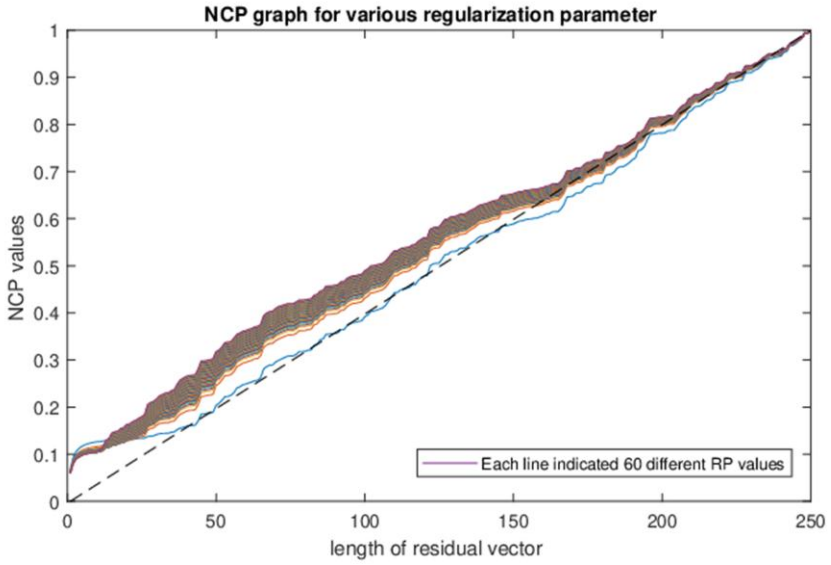

Fig-14: A graph showing the behavior of NCP values for various differen regularization parameters

Figure 14 indicates that the NCP values for the first regularization parameter (blue line) lies closer to the Gaussian white noise characteristic line (dotted black line). Hence, we isolate those particular values and it is shown in figure 15.

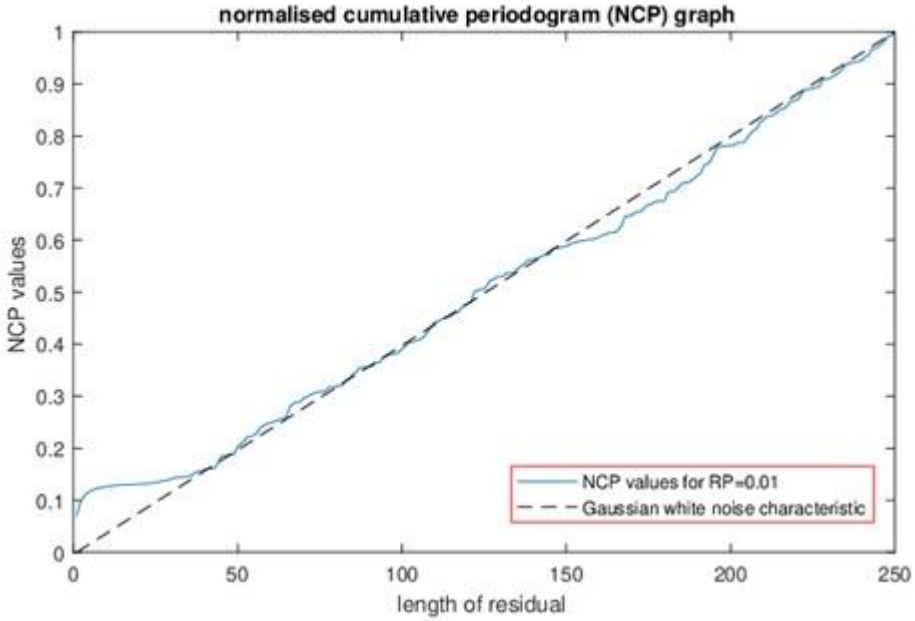

Fig-15: A graph NCP values for regularization parameter=0.01

In the case of test function 2, the optimal values from either L-curve or NCP provides satisfactory numerical obtained solution. As seen in figure 16 the chosen optimal values prove to be satisfactory as the given function is identical to the numerical one.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

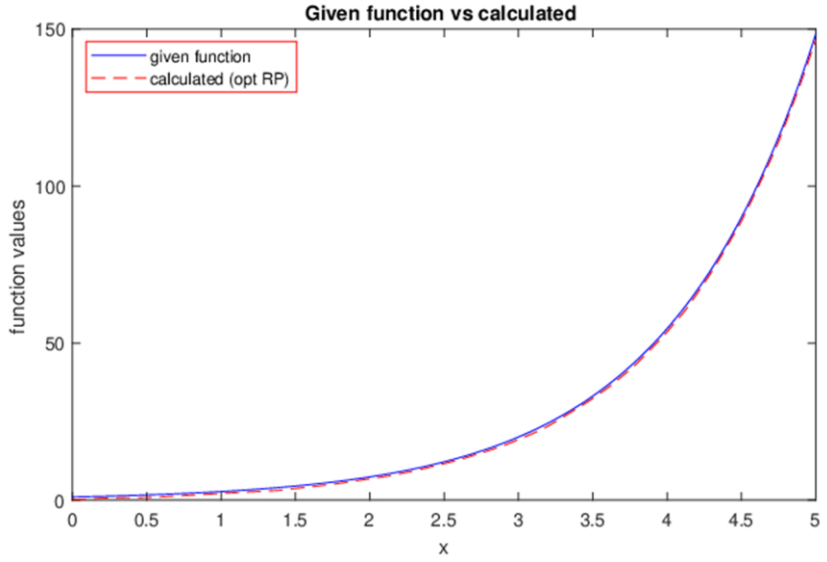

Fig-16: A graph showing given function and numerically obtained function for test function 2

3.3 Effect of different noise levels (standard deviation) on regularization parameter

In section 3.1 and 3.2, we provide the results using different test functions to determine the optimal values but for only one noise level (standard deviation). This section shines light on the effect of standard deviation of noise (AWGN) on regularization parameter. We consider the test function 2 and change the standard deviation as shown in the table below,

Standard deviation 1	0.02
Standard deviation 2	0.05
Standard deviation 3	0.1
Standard deviation 4	0.2

Table-6: Different standard deviation

As explained in 3.1.2, we perform the analysis for various standard deviation in table 6 and the appropriate results were collected. The graph below shown the relationship between standard deviation and regularization parameter.

Note:

The marked values in below figure 17 are the respective optimal values for the standard deviations mentioned in table 6. Hence, with the help of figure 17 we can draw the inference that higher standard deviation of noise (AWGN) requires stronger regularization parameter as the process or dataset has higher variations in noise levels.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume: 02/Issue: 04/April-2020 www.irjmets.com

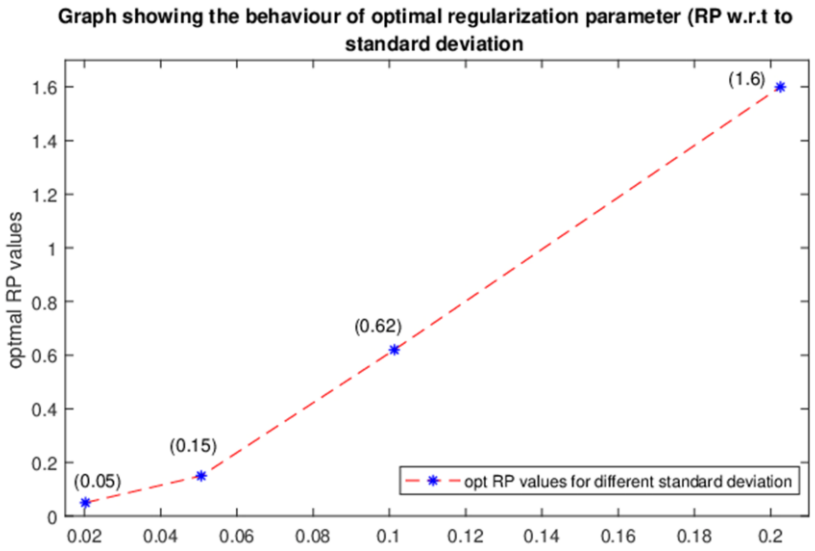

Fig-17: A graph showing the behavior of optimal regularization value w.r.t different standard deviation for test function 2

Note:
The marked values in figure 17 are the respective optimal values for the standard deviations mentioned in table 6.

Hence, with the help of figure 17 we can draw the inference that higher standard deviation of noise (AWGN) requires stronger regularization parameter as the process or dataset has higher variations in noise levels.

IV. DISCUSSION

Several methods can be adopted to reduce the variations that also help to facilitate better information retrieval from noisy data. Among the methods investigated in this article, it was found that eye-balling (trial-and-error) method was tedious, inefficient and computational infeasible in nature. In order to obtain the optimal values either L-curve or NCP (Normalized Cumulative Periodogram) provide satisfactory numerical solution. Also, higher standard deviation of noise (AWGN) requires stronger regularization parameter as the process or dataset has higher variations in noise levels. Murat *et.al* [5] have supported the present study with respect to L-curve method. Similar trends of using NCP method was followed by Bert and Dianne [11].

V. REFERENCES

[1] Rick Chartrand (2011) Numerical Differentiation of Noisy, Nonsmooth Data. *ISRN Applied Mathematics*, (2011),pp: 1–11. ISSN: 2090-5564. DOI: 10.5402/2011/164564.

[2] Avinash Bapu Sreenivas T. Srisupattarawanit H. Ostermeyer (2019) Numerical Methods for Information Tracking of Noisy and Non-smooth Data in Large-scale Statistics. *Journal of Engineering Research and Reports*, 6(4):2–11. ISSN: 2090-5564.

[3] Per Christian Hansen (1992) Analysis of Discrete Ill-Posed Problems by Means of the L-Curve. *SIAM Review* 34(4): 561–580. ISSN: 0036-1445. DOI: 10.1137/1034115.

[4] Per Christian Hansen and Misha E. Kilmer (2007) A parameter-choice method that exploits residual information. *PAMM*, 7(1): 1021705–1021706. ISSN: 16177061. DOI: 10.1002/pamm.200700264.

[5] Murat Belge, Misha E. Kilmer, and Eric L. Miller (2002) Efficient determination of multiple regularization parameters in a generalized L-curve framework. *Inverse Problems*, 18(4):1161–1183. ISSN: 02665611. DOI: 10.1088/0266-5611/18/4/314.

[6] Sarwate AD Song S Chaudhuri K (2015) Earning from Data with Heterogeneous Noise using SGD. *JMLR Work- shop and Conference Proceedings*, pp: 894–902.

[7] Sebastian Raschka (2018) MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software* 3(24):638 DOI: 10.21105/joss. 00638. URL: http://joss.theoj.org/papers/10.21105/joss.00638.

[8] Weisstein and Eric W. *Euler-Lagrange Differential Equation*. URL: http://mathworld.wolfram.com/Euler-Lagrange Differential Equation.html.

[9] The L-Tangent Norm. URL: http://www.brnt.eu/phd/node17.html.

[10] Introduction to Inverse Problems. Technical University of Denmark. URL: http://www2.imm.dtu.dk/~pcha/ DIP/chap5.pdf.

[11] Bert W Rust and Dianne P O'Leary (2008) Residual periodograms for choosing regularization parameters for ill-posed problems. *Inverse Problems*, 24(3): 034005. DOI: 10.1088/0266- 5611/24/3/034005. URL https://doi.org/10.10882F0266-56112F242F32F034005.