A Loss Landscape Perspective and Simulations for Imaging Inverse Problems based on AI and Neuron Network Training Method

Mingyong Zhou

School of Electronic Engineering
GXUST, China
Email: zed6641@hotmail.com

***Abstract* - The purpose of Imaging Inverse problems are to recover original true signals/images from observations by plethora physical methods. The recovered signals/images are measured by appropriate quantity and quality measures. Given a specific physical process, imaging is classically a mathematical algorithm design to estimate the original signals/images. In this paper starting from inverse imaging problem of Electrical Impedance Tomography (EIT), we will address several aspects of loss landscape and its regularization properties of a neuron network solution for imaging inverse problems. Loss landscape and regularization strategy for imaging inverse problems is first introduced in I), and is followed by II) Sampling from unbounded landscape to bounded discrete. In III), an asymptotic approximation property when neuron network width goes to infinity is analyzed. Complexity of training and approximation error is discussed in IV). In V), case study for indicator function approximation and application in imaging EIT by neuron network is highlighted. Finally in VI), discussions will be devoted to the limitations of regularization capability of neuron network as a new tool, especially in medical imaging applications.**

Keywords* - *ill-posed, regularization, loss function, convex optimization, indicator function, imaging inverse problem, AI, neuron networks.

I. INTRODUCTION

Well-posed problem is defined originally by J. Hadamard as a problem in which there exists stable, unique solution and solution is continuously depending upon input. Imaging inverse problems are often ill-posed problems, and thus are well investigated as mathematical problems using appropriate regularization models for optimization purpose. For example, in biomedical engineering, imaging problem is a type of inverse problem and widely applied in medical and target tracking. It involves the estimation of inverse functions by regularization methods and optimizations. On the other hand, neuron networks by nature is good at regularization in that it generalize and generate rules from large discrete samples. Even if the inverse problem is an ill-posed problem (unbounded at some values, undefined at some values, jumping or abruptly changing at some values) in context of continuous meaning[1], we will show that neuron network can regularize the illness in inverse problems with large sample data. First with discrete data samples, neuron network can learn and regularize possible illness by interpolating the discrete samples by constraints of inverse complexity number. This in fact is a first step towards regularization for discrete samples that could be generated from ill inverse function. Secondly with correct and appropriate regularization method, the loss function for a neuron network could be a convex optimization problem. There have been some investigations on how to convert a non-convex regularization optimization into a dual convex optimization problem for a neuron network [1][8]. In large parameters optimization, convex optimization is important to achieve the global minimum for a large neuron network model.

In biomedical imaging problems, the problem itself is complicated in that the values of original images are in a large space including real values and jumping up and down values for example. By reducing the large space into an assumed smaller pace, space complexity of inverse problems can be reduced. However the practical accuracy cannot guaranteed [9]. In this paper we will focus our discussions on the abrupt changing values in original images, as well as continuous images. The motivation is clear: to trace simultaneously continuous images as well as the abrupt changing "dot" images of small inclusions usually relating to illness in a patient. These dot images relates to indicator functions in mathematics and were investigated in [7] as well.

In fact the loss function definition itself needs to be well investigated in that the loss function itself must be well defined first in a reasonable way to reflect the problem itself, and must be feasible to be calculated secondly including its convexity property. We will show that in fact a loss function including a convex regularization items can be constructed based on descriptions in [2]. We in addition propose and show that such a regularization includes both small weights as well as large weights in order to trace jumping and abruptly changing, as follows:

$$L1 = \lambda_1 \| R - G^{(w,b)}(V) \|_p + \lambda_2 C\left(G^{(w,b)}\right) + \lambda_3 \{\| w \|_p + \frac{1}{\| w \|_p}\} \quad (1)$$

[1] Electrical Impedance Tomography (EIT) is a good example in which EIT forward problem is formulated as Maxwell partial differential equations in its continuous formula while its inverse problem is ill-posed.

$$L2 = \lambda_1 \| R - G^{(w,b)}(V) \|_p + \lambda_2 \frac{1}{C(G^{(w,b)})} + \lambda_3 \{ \| w \|_p + \frac{1}{\| w \|_p} \} \quad (1a)$$

$$L3 = \lambda_1 \| R - G^{(w,b)}(V) \|_p + \lambda_2 \{ C(G^{(w,b)}) + \frac{1}{C(G^{(w,b)})} \} + \lambda_3 \{ \| w \|_p + \frac{1}{\| w \|_p} \} \quad (1b)$$

where $\lambda_1 + \lambda_2 + \lambda_3 = 1.0, (0 \le \lambda_1, \lambda_2, \lambda_3 \le 1)$, $0 < C < 1.0$ denotes complexity regularization[14], and could be a complexity measure such as Rademacher or Hessian complexity measure e.g., R denotes an Impedance vector of $n \times 1$ and V is a voltages measurement vector $m \times 1$ in EIT case. $G^{(w,b)}$ denotes the neuron network with appropriate weights and bias (W, b) with n inputs and m outputs. $L1, L2, L3$'s correspond to regularization of minimal complexity, highest complexity and most appropriate complexity respectively.

Remarks on loss landscape denoted by equation (1)-(1b) about their convexity and other property aspects:

-1 the first item in *L1, L2, L3* of (1)-(1b) are based on empirical observed data samples. Assumed that $p = 2$, the second items *L1* of (1) penalizes the higher complexity solutions while *L2* penalizes lower complexity. Both *L1, L2* 's are of non-convexity while *L3* selects an optimal complexity and is a convex optimization problem. In fact the geometry property of the first item in *L1, L2, L3*'s were well discussed by Y. Cooper [15] on the loss landscape's geometric property. The global minimum is shown to be a sub-manifold instead of a function from R^n to R for over-parameterized neuron model.

-2 When *P=2*, $\lambda_1 \| R - G^{(w,b)}(V) \|_p + \lambda_3 (\| w \|_p + \frac{1}{\| w \|_p})$ could be a strong convex as well depending upon the quantity of all items. The last regularization items with respect to W were motivated from and explained in [2].

-3 $\lambda_2 C(G^{(w,b)})$, $\lambda_2 \frac{1}{C(G^{(w,b)})}$'s are associated with complexity regularization which was also discussed in [14]. A discussion on upper and lower bounds for Rademacher complexity of Banach space is addressed in [10]. In [18] it is shown that deep neural network converges into low complexity solution without regularization. In contrast to [18], *L2* of (2) in fact penalizes low complexity solutions. One needs to investigate the landscapes represented by *L1, L2, L3* in terms of it dual convexity and convergence property.

In theory if one can prove that if landscapes (1)-(1b) are strong convex landscapes or is equivalent to a dual convex optimization[2], solution is unique for $\min_{w,b,C} L1, \min_{w,b,C} L2, \min_{w,b,C} L3$ in (1)-(1b), otherwise there exists non-unique solutions to $\min_{w,b,C} L1, \min_{w,b,C} L2, \min_{w,b,C} L3$ which leads to uncertainty in imaging observed in some applications of CT imaging[1]. In fact the definitions of loss landscapes in (1)-(1b) play important roles in approximation errors as argued in [14] as well. As will be shown in next section, the regularization will lead to a smoothed landscape/function to approximate the ill-posed inverse landscape, as a first step to incur an approximation error. An upper bound of statistics of errors is discussed in [14]. The neuron network denoted by $G^{(w,b)}$ in the next step is applied to approximate the regularized landscape.

II. Sampling from Unbounded Landscape to Bounded Discrete - Neuron Network is a Natural Regularizer

With the digital computers, discrete samples are obtained for a CT imaging for example usually for a better quality and higher resolution. In imaging problems such as EIT, forward problem formulated by Maxwell differential equations with initial boundary values can be solved with finite element method (FET). The inverse problem however is often ill-posed illustrated by Figure 1 in which discontinuous, unbounded and jumping intervals exist. However in EIT case one can measure as large as possible $\vec{y}$ as impedance with given voltages observations as $\vec{x}$ denoted by "+" in Figure 1. We demonstrated in [11] that with proper regularization method, an good approximation using one hidden layer by the smoothed (regularized) curve is possible for one dimension case.

Now we assume that a lower bound for a complexity denoted by $C(G^{(w,b)})$ exists as defined in loss landscape (1), and with appropriate training data the inverse landscape/function can be learned by a neuron network with minimum complexity. With the complexity to be constrained as minimal, the neuron network with a loss landscape of (1) is a natural smoothed curve /regularizer in figure1. In fact there are mathematician who successfully derived inverse formula such as for Radon forward transform [16]. But due to the inverse computation burden, such inverse formula for Radon transform is for characterization only instead of practical applications. Neuron network on the other hand provides a natural smoothed/regularization tool as long as the complexity of inverse landscape/function can be well understood. Figure 1 demonstrates an approximation to infinity and jumping values for an ill-posed function $y = g(x1, x2)$,usually relating to inverse function

[2] In [1] a dual convexity theory is proposed for loss definition $L = \| R - G^{(w,b)}(V) \|_2 + \lambda_1 \| w \|_2$ where the loss definition is slightly different from landscapes defined in equations (1)-(1b) in our paper .

in imaging problems ,where g is ill-posed and_hard to be described by a simple mathematical equation.

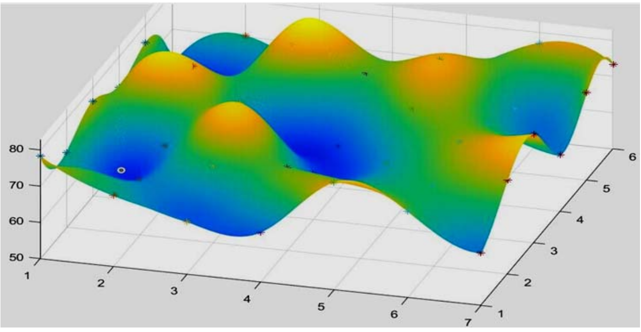

Figure1. Approximation to $y = g(x1, \ x2)$ (where g is ill-posed with infinity and singularity (+ denotes training data samples as input to neuron network to approximate the ill-posed function/landscape)

In fact a conjecture can be formulated on the statistical approximation error for loss landscape definition (1):

Conjecture 1:Define $Q = \min_{w,b,C} L1$,where $L1$ is defined in (1). If e is defined as an statistical approximation error between regularized landscape/curve and original landscape as illustrated in Figure1, following the arguments in [14], e is upper bounded by $e < O(\sqrt{Q})$.

In [11] simulations are carried out using a shallow neuron network (with only one hidden layer) to interpolate the values between the measured data samples in Figure 1, and enough training data samples are obtained for the error estimations. Rectified Linear Unit (ReLU) performs best as candidates of activation functions. In [12] simulations for EIT using deep neuron network (with 2 hidden layers, $L1$ is used for loss landscape with $\lambda_1 = 1.0, \ \lambda_2 = \lambda_3 = 0$), enough training data samples are measured in circuit for training purpose, and a preliminary empirical result using only eight voltage measurements is satisfactory for eight impedance tomography result. Such result is considered as a low complexity solution with convergence as in [18]. However to reach a higher complexity solution with reduced error estimates, it is shown in [18] that training samples must be enhanced to offset the higher complexity in original data function space.

III. Asymptotic Approximation Property When Neuron Numbers Goes To Infinity

In this section, we provide theorems to show in quantity that the accuracy of function approximation for a bounded and continuous landscape (illustrated as regularized smoothed curve in Figure 1) is in order of $O(N^{q-1}), (q < 1)$, where N is the number of neurons.

Theorem 1: At any given regularized bounded and smoothed landscape point, the absolute value of approximation for regularized landscape is in the order of $O(N^{q-1}), (q < 1)$, where N is the number of neurons, and $q < 1$.

We provide a proof for one dimension case. Higher dimension cases can be proved similarly. The proof is derivation by two steps. First the absolute error for area by integration for each division is shown to be approximating to zero if one notices that for each i'th division:

($i = 0,1,2 \ldots N$) [2, **pp.143**]

$$\frac{2}{N} f_{min}(x_i) \leq S_i \leq \frac{2}{N} f_{max}(x_i). \qquad (2)$$

Therefore with the assumption that the function is bounded,

$$\lim_{N \to \infty} S_i = 0. \qquad (3)$$

When x_i is approaching x at the i'th division, since the area is approaching zero then

$$\lim_{N \to \infty} \frac{2}{N} \left(\lim_{x_i \to x} (f(x_i) - f(x)) \right) = 0. \qquad (4)$$

$\lim_{x_i \to x} (f(x_i) - f(x)$) must convergence to 0 in order of $O(N^{q-1}), (q < 1)$ because otherwise (4) is not valid.

Therefore when q = ½ approximation error for large N is $O(1/\sqrt{N})$. Obviously q is relevant to complexity of landscape associated with inverse problem.

IV. Complexity Of Training And Approximation Error

In imaging inverse problems, for space R^n it can be shown that to reach an accuracy of $\frac{1}{\varepsilon}$, the complexity during training is $O(n^{E})$ [6]. However we insist that with super-computing power and increased training time, training computing with $O(n^{E})$ is not impossible in practice. In particular, with the space R^n to be confined into a predefined space, we can handle the complexity easier as shown in [13]. This observation also leads to a discussion of complexity C in loss definition (1). This implies that to obtain a higher accuracy in space R^n, to reach an accuracy of $\frac{1}{\varepsilon}$, the complexity C is a larger number in (1) which in turn leads to a higher upper bound in approximation error in **Conjecture 1**.

Corollary 1: Any bounded functions with jumps can be approximated up to any accuracy in order of $O(N^{q-1}),(q<1)$, where q value depends upon the jump sharpness as well as complexity of landscape.

This corollary is achieved from theorem 1 if one applies construction of neuron network in [2].

In fact for a type of smooth enough function, an upper bound for the approximation error can be derived as in theorem 2.

Theorem 2: For a type of smooth enough function $f(t)$ an upper bound for approximation error can be derived as

$$|f_M(t)-f(t)|<C+\frac{1}{2}(|a_M|+|b_M|), \qquad (5)$$

where C is a positive constant, $f_M(t)$ is a trigonometric series expansion of $f(t)$ and a_M, b_M are the coefficients of trigonometric series expansions. $\lim_{M\to\infty}\{a_M, b_M\}=0$

The proof is a direct derivation based on Lebesgue's theorem on convergence series and is omitted here.

V. SIMULATIONS FOR EIT IMAGING PROBLEM BY FULLY CONNECTED FORWARD NEURON NETWORK

In [12][13], a combination optimization training method using neural network is used for detecting black and white discrete dot imaging, assumed that imaging are comprised of discrete black and white with small enough dots. For imaging confined to predefined space, such combination optimization approach is feasible, if only enough training samples are provided in theory. We apply a fully connected forward neuron network with three hidden layers each of which has 128, 64 and 6 neurons respectively. Three sets of 10 impedance and 10 voltages measurements are used for training and EIT tested results are obtained. 10 true impedance values and tested values are compared for each of 10 impedance values. 3sets samples of tested result are illustrated in the following figures.

Simulations Results for 3 set samples each standing for 10 voltage measurements and 10 Impedance is demonstrated as follows. Figure 2-Figure 4 illustrate the estimation results using loss function equation (1) without regular items while Figure 5-7 illustrate the estimation results using loss function equation (1) with regularization item $(\|w\|_2+\frac{1}{\|w\|_2})$. The performance improvement by including a regularization item $(\|w\|_2+\frac{1}{\|w\|_2})$ is hardly observed in figure 5-figure7. However more configurations of impedance and test data are needed for any conclusions.

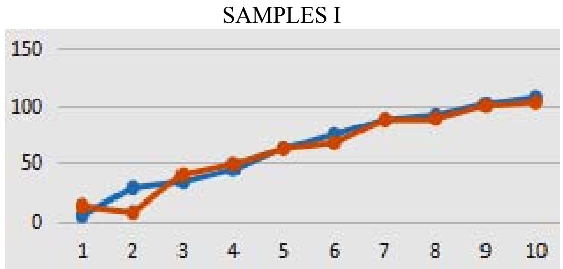

Figure 2. Testing after training results for samples I (Orange indicate tested impedance values vs blue trues)

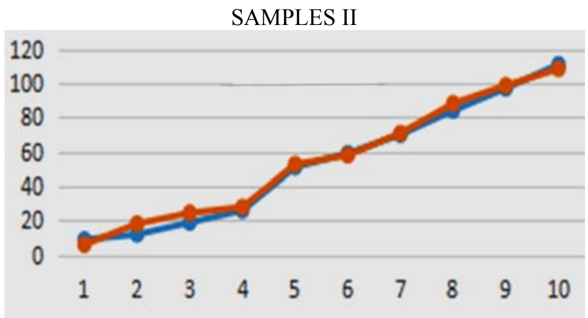

Figure 3. Testing after training results for samples II (Orange indicate tested impedance values vs blue trues)

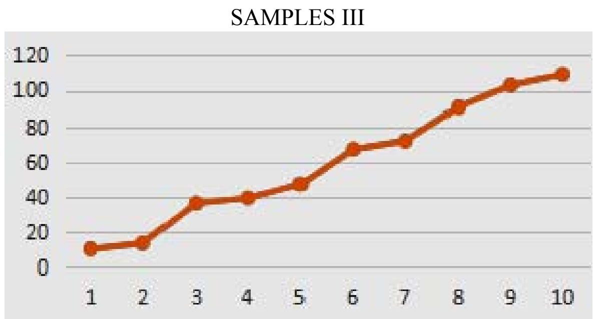

Figure 4. Testing after training results for samples III (Orange indicate tested impedance values vs blue trues)

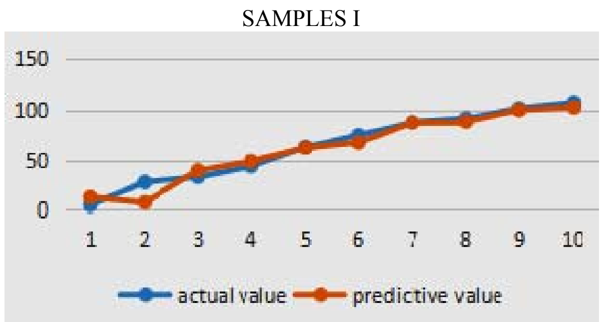

Figure 5. Testing after training results for samples I (Orange indicate tested impedance values vs blue trues)

SAMPLES II

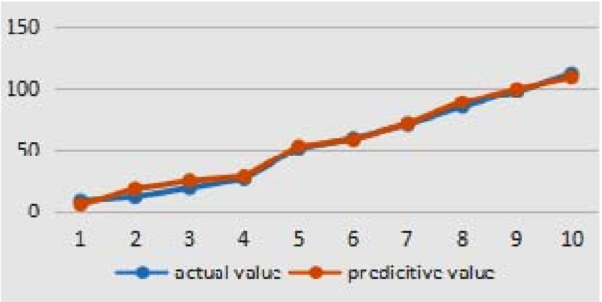

Figure 6. Testing after training results for samples I
(Orange indicate tested impedance values vs blue trues)

SAMPLES III

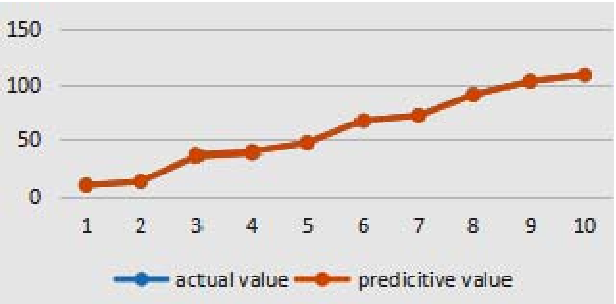

Figure 7. Testing after training results for samples I
(Orange indicate tested impedance values vs blue trues)

To compare with traditional regularization method, we apply the same configuration setting into an EIDORS platform with 10 impedance and 10 voltages measurements. Figure 8 illustrates the diagonal setting for 10 impedance next to each other. Figure 9 is the imaging result for 10 impedance. Initial subjective judgement is that 10 impedance cannot be easily identified from the EIDORS imaging results. Definitely more simulations are needed for future investigations on performance comparing study.

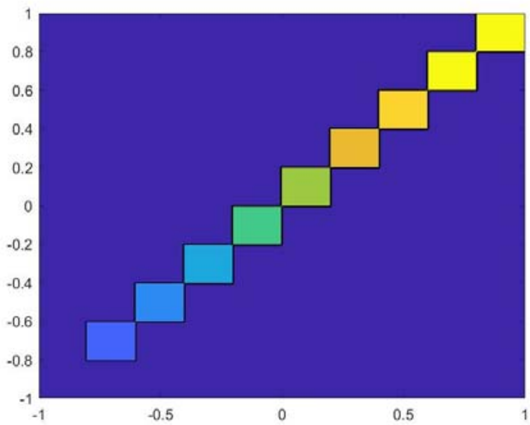

Figure 8. 10 Impedance Configurations Settings for Testing

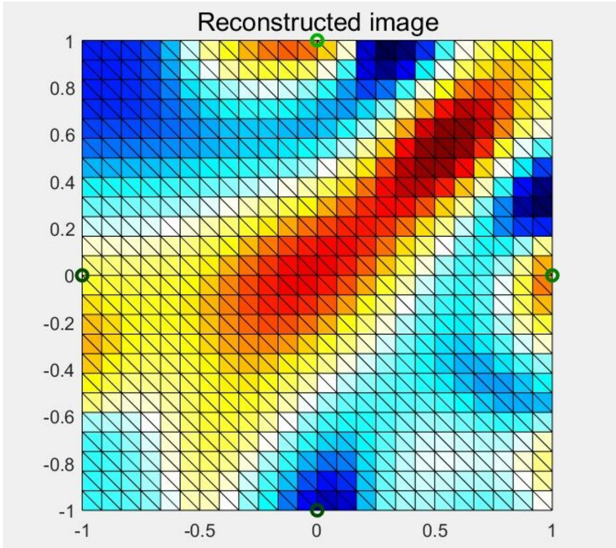

Figure 9. EIDORS imaging result based on diff GN-1 STEP

VI. DISCUSSIONS ON REGULARIZATION CAPABILITY OF NEURON NETWORKS

The loss landscape associated with neuron network architecture needs to be carefully investigated in the first place. The guarantee of well-posed solution to imaging inverse problem needs to be ensured while one investigates approximation accuracy and complexity regularization etc.

Regularization method based on neuron network is first observed by its natural property of neuron network approximation capability for any smoothed functions. It is totally data driven and based, therefore it is a natural smoother/regularizer in the first place. The theory behind is that it is based on large data sample data for over-parameterized neuron model and could be generalized well up to complexity constraints, partly due to recent benign over-fitting observations[17]. Such generalization capability is based on neuron network property and is thus a combination of neuron network with plethora regularization method that could eventually enhance the final performance such as in EIT imaging. Future work will be devoted to a comparing study on the pros and cons of AI based methods as compared to traditional imaging method such as quasi Newton -Raphson method in EIT.

VII. ACKNOWLEDGEMENT

The author acknowledges the preparation of figure1 in the paper by Li Yancheng of GXUST and Ji Zhengjia to prepare data from Figure 5 to Fgure7. Figure 8 and 9 were prepared by Li Yongqi for whom the author appreciate. Financial support from GXUST fund no. 19Z39 is acknowledged.

REFERENCES

[1] M.Pilanci, " The Hidden Convex Optimization Landscape in Deep Neuron Network", DeepMath 2020, update at DeepMath 2021 talk.
[2] Michael Nelson, "Neuron network and Deep Learning", 2020 August, Translated version, Telecommunication Press, Beijing.
[3] B.Kiani, "Joint embedding in Self-Supervised Learning in Kernel Regime", DeepMath 2022, San Deigo, also in arXdiv:2209.14884 [CS].
[4] Simon Haykins, " Neural Networks and Learning Machines", China Machine Press, 2017, July.
[5] Itay Hubara et al "Binarized Neuron Network", 30th international conference on NIPS, 2016, Barcelona, Spain.
[6] Chen Tianping, " A Note on ReLU Function ", Note on ReLU Function.
[7] A. Caragea, "Neural network approximation and estimation of classifiers with classification boundary in a Barron class", arXdiv [MATH.FA], 2011.09363.
[8] M.Pilanci, "Neural Networks are Convex Regularizers: Exact Polynomial-time Convex Optimization Formulations for Two-layer Networks, PMLR 119, 2020 .
[9] Weinan E, "Machine Learning and Computational Mathematics", arXiv:2009.14596v1[math.NA] 23 Sep., 2020.
[10] Weinan E, "On the Banach Space Associated with Multi-layer RELU Networks", arXiv:2007.15623v1[stat.ML] 30 Jul 2020.
[11] Xie Zongrui, "A Comparing study on activation functions", GXUST undergraduate thesis 2023, under supervision of Dr. Zhou Mingyong.
[12] Guangtao, "An EIT algorithm design and simulations using Neuron Network", GXUST undergraduate thesis 2023, under supervision of Dr.Zhou Mingyong.
[13] Yuxiang Huang, "Combination optimization for EIT problem using Neural Networks",GXUST undergraduate thesis 2021 under supervision of Dr.Zhou Mingyong.
[14] Andrew Barron,"Statistical Property of Artificial Neural Network", IEEE Proceedings of the 28th Conference on Decision and Control, Tampa, Florida, pp.280-285,1989.
[15] Y.Cooper, "The Loss Landscape of Over-parameterized Neural Network", https://doi.org/10.48550/arXiv.1804.10200.
[16] A.K. Louis, "Ghosts in tomography - the null space of radon transform", Mathematical Methods in the Applied Sciences,3(1):1-10, 1981.
[17] Peter Bartlett, "Benign over-fitting", Lecture1, CIRM, December 31, 2021.
[18] L.Wu and Weinan Er, "Towards Understanding Generalization of Deep Learning: Perspective of Loss Landscapes", aXiv: 1706.10239v2 [cs.LG], 28 Nov 2017.

.