Tuning Regularization Parameter for Solving ill-posed Inverse Problems using Subspace Methods

Amartansh Dubey
ECE Department, HKUST
adubey@connect.ust.hk

Samruddhi Deshmukh
ECE Department, HKUST
ssdeshmukh@connect.ust.hk

Nilesh Kumar Jha
ECE Department, HKUST
nkjha@connect.ust.hk

Abstract

One of the major challenges in solving ill-posed inverse problems is to achieve uniqueness and stability in the solution. Such problems usually arises from an under-determined (known measurements $\ll$ unknowns variables) and unstable system of equations where small perturbation in measurements can lead to large changes in the solution. Various regularization techniques have been proposed to solve these problems which enforces certain priors on the optimization variable and helps to achieve stable solution. Tuning the regularization parameter (λ) for solving ill-posed problems is difficult as small changes in λ can cause significant perturbation in the solution. The most common approach to tune regularization parameter (λ) is cross validation. This is a supervised learning approach as we require ground truth data in validation set for comparing performance of the regularized solution for various values of λ. This way of selecting λ fails if the test data is out of the training and validation data distribution. This is usually the case in the field of inverse scattering and radio imaging where scatterer (object to be imaged) can take any possible shape, size, location and physical properties (like refractive index) and it is difficult to create a comprehensive training and validation dataset. In this report, we propose a simple but efficient PCA and Sparse-PCA based unsupervised approach to tune λ. We demonstrate with the physics based simulation of electromagnetic scattering that the proposed method can select optimum λ for a given ill-posed problem and do not require additional data for cross validation based tuning. We also compare it to L-curve method which is another commonly used unsupervised approach for tuning λ

1 Introduction

Solving ill-posed inverse problems is a crucial step in wide range of research and commercial fields including medical imaging (Tearney et al., 1995), inverse scattering imaging (Lu et al., 2010; Depatla et al., 2015; Chen, 2018), image denoising (Yang et al., 2010), security scanning (Liu et al., 2007), signal processing (Wilson and Patwari, 2010) and many more. The key challenge is to solve an under-determined system of equations to converge to an unique solution that is stable (small perturbation in measurements do not significantly affects the parameter estimation). These under-determined systems are usually solved by imposing certain prior assumptions on the unknowns (like sparsity) which effectively reduces the number of unknowns compared to the known measurements. Such prior based approaches can be categorized into two classes, 1) maximum a posteriori probability (MAP) estimation and 2) Bayesian estimation. In this report we focus on MAP estimation problem. The widely used regularization methods including LASSO, Ridge, Total Variation, ELASTIC NET, all are the MAP estimates.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

A general inverse problem can defined as,

$$\underset{x}{\text{minimize}} \quad f(y - Ax) \tag{1}$$

where, $y \in \mathbf{C}^{m\times 1}$ is measurement vector (m known observations) and $x \in \mathbf{C}^{p\times 1}$ is unknown variable (p unknowns). $A \in \mathbf{R}^{m\times p}$ is a model matrix which transforms the variable vector into the measurement space. Matrix A depends on the domain of inverse problem. In physics and signal processing, matrix A is also considered as channel matrix which encapsulates the physical behaviour of the system under test. When $m = p$ and A is full rank, x can be estimated as $x = A^{-1}y$. For overdetermined system ($m >> p$), least square estimate (LSE) can give unique and stable solution as $x = (A^H A)^{-1}A^H y$. The challenging part is when problem is under-determined ($m << p$). This is solved using MAP estimation by including additional prior term (also called regularization),

$$\underset{x}{\text{minimize}} \quad ||(y - Ax)||_2^2 + \lambda f(x) \tag{2}$$

where, f is operator on the variable vector x and is used to impose prior on the x. The selection of f provides wide range of regularization techniques, for example,

$$\begin{aligned} f(x) &= ||x||_2^2 \text{ (Ridge Regression)} \\ f(x) &= ||x||_1 \text{ (LASSO Regression)} \\ f(x) &= \sum_{\forall i} |x_{i+1} - x_i| \text{ (Total Variation regression)} \end{aligned} \tag{3}$$

1.1 Brief Introduction to Inverse Scattering Problem

For this report, we use our research problem to generate the model and data. The problem statement is to solve inverse scattering problem (ISP) for indoor imaging. ISP models the interaction of electromagnetic waves (like radio waves from WiFi) with various objects in an given domain of interest (DOI) in terms of reflection, refraction, diffraction and scattering (Chen, 2018). The objects scatters the waves in various directions and these scattered signals are captured by the wireless sensors to generate the measurement data. A pictorial illustration is shown in Fig. 1. This measurement data can be used to solve a inverse problem to get the image of the objects. Fig. 1 shows the problem setup where transceiver nodes are placed at boundary of a imaging domain of interest (this can be a indoor room). There is a arbitrary shaped object (or scatterer) placed inside DOI which scatters the signals transmitted from the nodes. The forward problem is to find the received signal given image and properties of the scatterer. It can be simulated using volume source integral known as Lippmann-Schwinger equation which can be solved using method of moment approach (Bourlier et al., 2013). We have used forward simulations to generate the data.

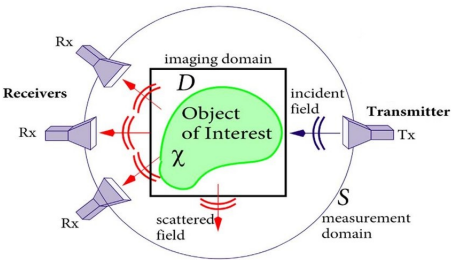


Figure 1: Illustration of Inverse Scattering Imaging Setup.

The inverse problem is to find the image of the scatterer given the received signals. The inverse problem is considered highly non-linear and non-convex (Chen, 2018; Bates et al., 1991) which is beyond the scope of this report. For this report, we used simplified linear approximation of the inverse problem which works for given physical constraints (size of scatterer has to be smaller than wavelength and permittivity of scatterer has to be small). This linear model is briefly described below in terms of measurements, model matrix and variable (more details can be found in (Depatla et al.,

2015)):

$$\begin{aligned}
&\text{Measurement Vector:} \quad y = P_{Tx}^{tot}(\mathbf{r}_{Rx})[\text{dB}] - P_{Tx}^{inc}(\mathbf{r}_{Rx}), \\
&\text{Model Matrix:} \quad A = 10\log_{10}e^{2k_0^2} \cdot \text{Re}\left[\frac{\sum\limits_{\forall n|\mathbf{r}_n \in \mathcal{D}} g(k_0 r_{n,Rx}) E_{Tx}^i(\mathbf{r}_n)\Delta a}{E_{Tx}^i(\mathbf{r}_{Rx})}\right], \\
&\text{Variable Vector (2D image unrolled as 1D):} \quad x = \chi(\mathbf{r}_n) = (\epsilon_r(\mathbf{r}_n) - 1)
\end{aligned} \tag{4}$$

Where, the measurement vector $y \in \mathbf{R}^{m\times 1}$ is difference in received WiFi signal power at receiver Rx (in presence and in absence of the scatterer), Model matrix $A \in \mathbf{R}^{m\times p}$ is derived in (Depatla et al., 2015), where $g(k_0 r_{n,Rx})$ is 2D greens function (Hankel function), $E_{Tx}^i(\mathbf{r}_n)$ is incident field from transmitter to object and $E_{Tx}^i(\mathbf{r}_{Rx})$ is the direct incident field from transmitter to receiver. The point r_n is any point inside object. The image vector which needs to be estimated is $x \in \mathbf{R}^{p\times 1}$ which is contrast in dielectric constant ϵ_r of object with respect to free space. This forms a linear inverse problem where $y \approx Ax$ is given system of equations with x being unknown. In this report, we used our forward research models to accurately generate y and A and then solved the corresponding inverse problem to find x when problem is severely under-determined ($m << p$). Given y and A, x can be estimated using regularized optimization in (2 which requires careful tuning of regularization parameter.

1.2 Common Methods to Tune Regularization Parameter (λ)

The value of λ trade-offs between extent of prior being enforced on x and actual data fitting using least squares. Since, λ is a hyper-parameter, there is no unique solution to (2) and accuracy will depend on the choice of λ. Furthermore, in applications such as 2D signal reconstructions (like MRI imaging, indoor imaging, etc), the sensitivity of solution of (2) increases w.r.t to λ as there can high spatial correlations in both horizontal and vertical directions in x. The most common approach for selecting optimum λ^* is by cross validation, which simply use validation dataset to check which regularization parameter performs well and then use that value for the test data. This is effective when enough data is available for training, validation and testing purposes. However, if the enough data is not available or there is high probability that test data can be out of the validation data distribution, such the cross validation will not give good results. This is also the case with our problem of interest in inverse scattering imaging using radio waves where the scatterer (object of interest) can take any possible shape, size, location, physical properties like refractive index. For such cases, it is better to solve (2) as an stand-alone optimization problem by only relying on the given measurement vector and model matrix (and not rely on creating training, validation datasets using collection of labeled pair of measurement and variable vector). For this, we need unsupervised learning approach for tuning λ instead of supervised cross validation method.

One such unsupervised approach is L-curve method which do not need additional data (Hansen and O'Leary, 1993; Calvetti et al., 2000). Instead, it uses trade-off curve between LSE residual term ($||(y - Ax)||_2^2$) and regularized term $f(x)$ for different regularization parameters and then select the value where the trade-off curve has a knee point (more details of L-curve can be found here (Hansen and O'Leary, 1993)). This provides huge advantage because L-curve approach uses just a measurement vector and model matrix to find λ instead of relying on additional validation data. Therefore, L-curve method can find λ in a unsupervised manner whereas cross validation approach requires additional labeled data for tuning λ. This helps in generalizing the solution for out-of-training datasets and hence L-curve method is preferred over cross validation approach in ill-posed inverse problems in physics.

However it has been shown that for severely underdetermined large dimensional problems with large measurement noise, L-curve may not guarantee a good selection of λ as shown in (Xu et al., 2016). Furthermore, L-curve method may also give sub-optimal value of λ if variable vector x has intricate correlations (temporal or spatial) and the model matrix derived from domain information is approximation to actual underlying non-linear model (as we shall show this later in results). Furthermore, selecting the only on value of λ (at the knee point of L-curve) means throwing away some other potentially useful information which might be achieved by other values of λ. This leads to loss of structure in imaging.

In this report, we propose a simple yet effective subspace decomposition approach which removes the need of regularization parameter tuning and combines the desirable structural similarity among multiple solutions achieved by different values of λ.

2 Proposed Method

The optimum estimate of variable $x^*(\lambda) \in \mathbf{R}^{P\times 1}$ in terms of regularization parameter λ can written as follows,

$$x^*(\lambda) = \underset{x}{\text{minimize}} \quad ||(y - Ax)||_2^2 + \lambda f(x) \tag{5}$$

The regularization parameter can lie in range $\lambda \in (0, \lambda_{max})$ where for several regularization kernels f, the analytical expression for λ_{max} is available (Friedman et al., 2001). For example, in LASSO regression ($f(x) = ||x||_1$), $\lambda_{max} = |A^T y|_\infty$. The optimum value for regularization parameter λ^* can lie anywhere in this range of 0 to λ_{max}. The selection of λ^* depends on various factors including dimensionality constant of the problem ($c = m/p$), noise in y and accuracy of model matrix A.

For both methods, 1) selection via cross validation 2) L-curve method, we have to sample sufficient values of λ from $\lambda \in (0, \lambda_{max})$ and find corresponding estimates of $x^*(\lambda)$. The best among these estimates can be selected using cross validation or L-curve method as explained in section 1.2. If there is no analytical expression of λ_{max} available for particular choice of regularization function f, then λ_{max} can be chosen to be large enough so that all the coefficients of $x^*(\lambda_{max})$ shrinks towards zero (this is for case of sparsity priors (Friedman et al., 2001)). For other types of priors such as TV, we can select λ_{max} such that successive difference in coefficients of $x^*(\lambda_{max})$ becomes zero (Strong and Chan, 2003). In this way, we can select λ_{max} based on the properties of the prior applied in the regularization.

We sample n values of λ from $(0, \lambda_{max})$, i.e. $\lambda_i = (\lambda_1, \lambda_2, ..., \lambda_n)$ where $\lambda_1 = 0$, and $\lambda_n = \lambda_{max}$. The estimates of vector $x^* \in \mathbf{R}^{p\times 1}$ for each of these λ_i can be written in form of a $p \times n$ matrix denoted as X as,

$$X(\in \mathbf{R}^{p\times n}) = [x^*(\lambda_1), x^*(\lambda_2), ..., x^*(\lambda_n)]^T \tag{6}$$

Each column of matrix X in (6) represents an estimate of optimization variable x using certain value of regularization parameter λ_i. There are n such columns as we have n values of λ_i. Aim is to find which column of X gives the estimate of x which corresponds to the optimal value of λ. Alternatively, aim can be to find what linear combination of columns of X can give best reconstruction of x^* so that we can exploit best common structure among solutions given by multiple values of λ.

Our idea is to use subspace matrix decomposition tools such as PCA on the covariance matrix of X to find the common structure in the columns of X and find a direction (vector) along which most of the correlated structure is preserved.

2.1 Interpretation

While solving our original under-determined inverse problem ($m << p$) in (5), if we select value of λ which is too small, the estimate x^* will give solution which is close to the least square solution (LSE). It is well known that for an under-determined problem, the LSE solution gives highly unstable and noisy solution because the pseudo inverse matrix in $(A'A)^{-1}A'y$ is poorly conditioned and eigenvalues of $(A'A)$ are close to zero. In other words, for very small values of λ, we get the *high-variance noisy solution*.

On the other hand, if we select extremely large values of λ in solving (5), the coefficients in x^* are over-shrinked and provides high bias solution and extremely small values of coefficients of x^*. For example in case of LASSO with large λ, all coefficients of x^* will shrunk towards zero and hence do not provide meaningful results. In other words, the solutions are highly shrinked towards the prior. This gives *high-bias noisy solution*.

Therefore, there must be a optimal value of λ in between which can provide better solution and overcome the trade-off between high variance(LSE solution) and high bias (over-regularized solution) solutions. In eigen-space, we can interpret it as pulling up the eigenvalues which were very close to zero whereas pulling down the over-estimated eigenvalues.

To summarize, first few columns of X contains noisy and high variance estimates of x whereas last few columns contains highly biased and over-shrunk solution. There will be some columns in X which actually provide best estimates of x. However, despite noise in the estimates, all the columns in X should contain some common structure which captures some part of the optimal solution.

We aim to use techniques such as PCA, Sparse PCA (SPCA), Robust PCA (RPCA) to find the direction in p dimensional feature space of X which captures this common structure. For finding the direction of maximum variance (common structure/correlation among columns of X), the simplest approach can be to perform eigen-decompisition on the covariance matrix of X and the direction of first eigenvector can be the direction of maximum structural similarity. This eigenvector can be used as the solution of (5) instead of using any one of the columns of X as a solution.

As explained above, the matrix $X \in \mathbf{R}^{p \times n}$ is our data matrix of interest which can be considered as p-dimensional features space on which n samples are scattered. The $p \times p$ sample covariance matrix of X can be defined as,

$$\Sigma_{SCM} = \frac{1}{N}(X - \mu)(X - \mu)^T \tag{7}$$

Where, μ is $p \times 1$ mean of all p rows of X. Assuming features are zero mean, we can simply define sample covariance matrix as, $\Sigma_{SCM} = \frac{1}{N} X X^T$.

To find the direction of highest structural correlation among columns of X, we want a vector in p dimensional feature space which captures the maximum variance (or in other word, which captures the maximum common structure among the columns of X). Let this vector be $\tilde{x} \in \mathbf{R}^{P \times 1}$. Mathematically, we want to maximize following objective function (find a unit vector on which projection of X is maximized),

$$\begin{aligned} \underset{\tilde{x}}{\text{maximize}} \quad & var(\tilde{x}^T X) \\ \text{subject to} \quad & ||\tilde{x}||_2 = 1 \end{aligned} \tag{8}$$

where, the objective function $var(\tilde{x}^T X)$ can be expanded as,

$$\begin{aligned} var(\tilde{x}^T X) &= E[(\tilde{x}^T X - E(\tilde{x}^T X))(\tilde{x}^T X - E(\tilde{x}^T X))^T] \\ &= \tilde{x}^T (E[(X X^T)])\tilde{x} \qquad\qquad if\ E[X] = 0 \\ &\approx \tilde{x}^T (\Sigma_{SCM})\tilde{x} \end{aligned} \tag{9}$$

Using (9), the objective function in (8) can alternatively be written as,

$$\begin{aligned} \underset{\tilde{x}}{\text{maximize}} \quad & \tilde{x}^T (\Sigma_{SCM})\tilde{x} \\ \text{subject to} \quad & ||\tilde{x}||_2 = 1 \end{aligned} \tag{10}$$

The problem in (10) is well known maximum eigenvalue problem, where the maximum value of objective function is the maximum eigenvalue of Σ_{SCM} which is achieved for $\tilde{x}$ being the eigenvector of Σ_{SCM} corresponding to the maximum eigenvalue. In context of our problem in (5), the $\tilde{x}$ achieved from (10) is the solution which captures the structural similarity between the columns of original data matrix X. Therefore, using any one of the column of X as best estimate of x^* for solving (5), we can use $\tilde{x}$ (maximum eigenvector of Σ_{SCM}) as the solution which can capture the useful structures in multiple columns of X. Hence instead of selecting one column of X using L-curve or cross validation, we combine multiple columns of X as a potential improved solution.

The objective function in (10) can be further improved if the sparsity prior is favorable in the underlying inverse problem. This can be achieved by imposing additional cardinality constraint in (10) to get,

$$\begin{aligned} \underset{\tilde{x}}{\text{maximize}} \quad & \tilde{x}^T (\Sigma_{SCM})\tilde{x} \\ \text{subject to} \quad & ||\tilde{x}||_2 = 1 \\ & ||\tilde{x}||_0 = k \ \ (1 < k < p) \end{aligned} \tag{11}$$

The solving problem in (11) gives direction of maximum variance but now the components of the vector $\tilde{x}$ will have $p - k$ coefficients zero, hence we get a sparse solution. The problem in (11) is a non-convex and the relaxed convex formulation can be done using tools from Semidefinite programming (Yao, 2021, 2004). This approach is also known as sparse PCA (SPCA). There are

many other improved variants of PCA and SPCA Udell et al. (2014); Yao (2021), however, in this report we explore only these two. Future work can use more advance matrix factorization tools.

Here is the step by step summary of the proposed algorithm,

1. Solve regularization problem in (5) for n values of $\lambda \in (0, \lambda_{max})$. This gives n number of $p \times 1$ dimensional solutions x, each of these estimates can be denoted as $x^*(\lambda_i), i = 1, 2, ..., n$.
2. Stack these n estimates of $p \times 1$ column vector into a $p \times n$ matrix denoted as X. The ith column of this matrix represents a solution to regularization problem in (5) using certain value of regularization parameter λ_i.
3. Apply subspace decomposition methods such as PCA, SPCA or RPCA to find the direction which captures maximum structural similarity among n columns of matrix X.
4. In case of PCA, the direction of maximum structural similarity will be simply first principal component (maximum eigenvector) of sample covariance matrix Σ_{SCM} of X. It is denoted as $\tilde{x}$.
5. The estimate of $\tilde{x}$ can be used as a better solution of (5) instead of selecting any one column $x^*(\lambda_i)$ of the X as a solution.
6. $\tilde{x}$ will contain structure reconstructed by multiple regularization parameters while solving (5) and hence it can be considered better than selecting any one of the regularization parameter value for getting the solution.
7. Tools from Random Matrix Theory (Bun et al., 2016; Ledoit and Wolf, 2017, 2004) can also be used to clean Σ_{SCM} to improve the results (This can be part of future work).

3 Numerical Analysis and Results

In this section we solve an ill-posed inverse scattering problem (explained in section 1.1) to demonstrate how our proposed method can provide a stable solution. Fig. 2 (similar to Fig. 1) shows a geometric setup where a concentric ring shaped scatterer is placed at the center of a DOI. DOI is of dimension 3 x 3 m^2. 20 wireless transceiver nodes are placed at the boundary of the DOI. When one node transmit 2.4 GHz radio waves towards the DOI, the object inside the DOI scatters the waves and these scattered waves are received by other nodes in the receiving mode. There are total $m = 20 \times (20-1)/2 = 190$ unique link (from transmitting to receiving nodes) and hence number of measurements are $m = 190$. Hence, the measurement vector y is a 190×1 vector (see section 1.1 for expression of y). The aim is to image part of the DOI where scatterer is located as shown in the magnified area of 0.5 x 0.5 m^2 on right in Fig. 2. This is done because solving for whole 3 x 3 m^2 DOI can be computationally expensive. For imaging, we discretized the 0.5 x 0.5 m^2 DOI into 50×50 small pixels (total p=2500 pixels). Each of these pixels has a associated refractive index value which needs to be estimated. Using 190 measurements to estimate refractive index of all 1600 grids is the inverse problem at hand. Therefore, knowns are $m = 190$ and unknowns are $p = 2500$ which makes it a severely under-determined problem. The model matrix A is a 190×2500 matrix derived using the model given in (4).

To solve this problem, we solve Norm-2 regularization (Ridge regression). Any other kind of regularization can also be used, the process described in the previous section will remain same. We chose Ridge as it is faster to solve (using analytical expression) and also it gives a lot of noisy coefficients in the reconstruction of DOI profile. Testing against high noise data gives better idea about the performance of the proposed method. Using ridge regression, original problem in (5) can be written as,

$$\begin{aligned} x^*(\lambda_i) &= \underset{x}{\text{minimize}}\ ||(y - Ax)||_2^2 + \lambda_i ||x||_2^2 \\ &= (A^T A + \lambda_i I_{p\times p})^{-1} A^T y \end{aligned} \tag{12}$$

The estimated $x^*(\lambda_i) \in \mathbf{R}^{p\times 1}$ vector can be reshaped into $\sqrt{p} \times \sqrt{p}$ 2D image of DOI (where p = 2500, hence image will be 50x50 pixels). Solving (12) for n values of $\lambda_i = \lambda_1, \lambda_2, ..., \lambda_n$, we get a $p \times n$ matrix X (same as (6)). The value of λ_{max} cannot be easily estimated for Ridge regression, unlike in Lasso where λ_{max} is any value above which $\sum_{\forall i} x^*(\lambda_{max}) = 0$. For ridge, easiest way to select λ_{max} is to make it large enough such that most coefficients tends to zero. For our inverse

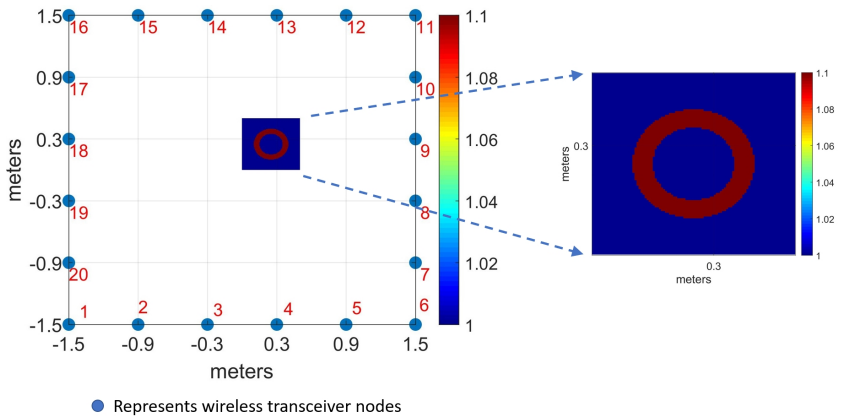


Figure 2: Illustration of Imaging Domain of interest (DOI). The DOI is 3 x 3 m^2 in area and 20 wireless sensors are placed around this DOI (see blue circles placed at the boundary of DOI, numbered from 1 to 20). On right, is the magnified region which consist of the ring shaped scatterer with relative permittivity = 1.1.

scattering problem, we select λ_{max} such that more than 90% of coefficients of $x^*(\lambda_{max})$ are below 10^{-5}. The value 10^{-5} is arbitrary selection which simply means small enough to be considered as zero.

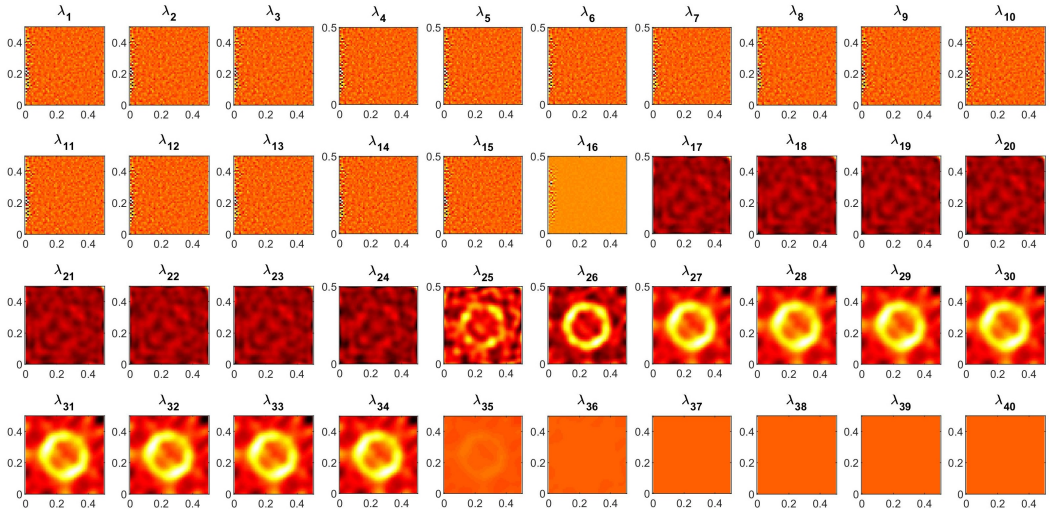
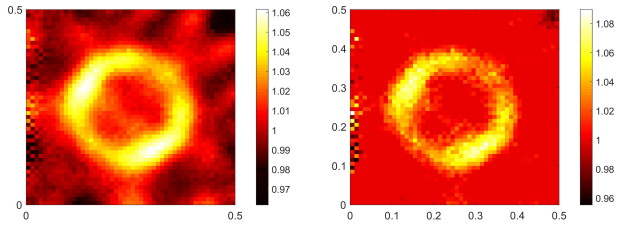


Figure 3: 40 reconstructions obtained by solving Ridge regression in (12) for 40 values of λ.

The Fig. 3 shows reconstruction of DOI profile using $n = 40$ values of $\lambda \in (0, \lambda_{max})$. The first subplot in Fig. 3 is for $\lambda_1 = 0$ whereas last subplot (bottom-right) is for $\lambda_{40} = \lambda_{max}$. As explained in section 2.1, it can be clearly seen from Fig. 3 that when regularization parameter λ is too small, the reconstruction just looks like random noise with no structural similarity with the original ring profile in Fig. 2. On the other hand, when λ is too large (example, λ_{40}), the reconstruction looks overly smooth with highly blurred biased image of the DOI profile.

On stacking these 40 reconstructed images (as column vector) into one matrix, we get 2500×40 data matrix X. We apply PCA and SPCA on the sample covariance matrix of X as explained in section 2.1 (equations (10) and (11)). Fig. 4(a) shows the image of first principal component (PC) of covariance matrix using traditional PCA as defined in (10). It can be clearly seen that first PC preserves the shape of the DOI profile and reduced the noisy reconstructed pixels which were arising due to under or over regularization. Fig. 4(b) shows image of first principal component of covariance matrix using sparse PCA as defined in (11). It can be clearly seen that there is improvement in terms of capturing the structure of the profile and most of the background pixels are zero, hence achieving a

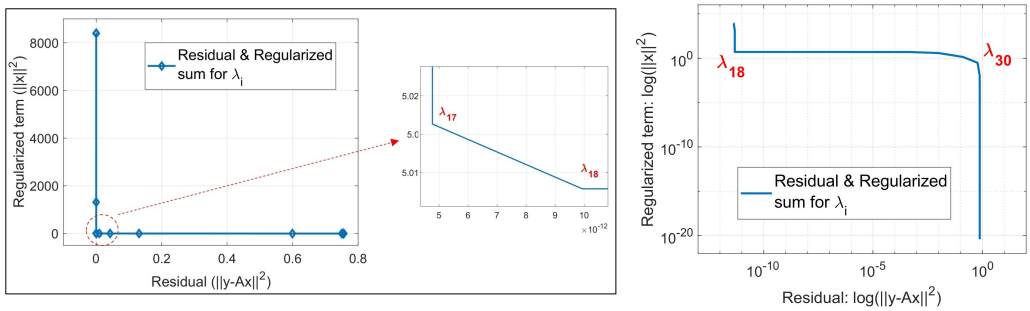
sparse reconstruction (NOTE: pixels having zero values are shown as pixel value of one in Fig. 4 because actuall relative permittivity is written as = 1 + x^* as explained in (4)).



(a) (b)

Figure 4: (a) First Eigenvector of Sample Covariance Matrix of X by solving (10). It captures common features in 40 images shown in Fig. 3. (b) shows principal component using Sparse PCA as formulated in (11)

Fig. 5(a) shows L-curve achieved after solving ridge regression (12) for 40 values λ. The knee point can be seen at λ_{17} and λ_{18}. It can be seen from Fig. 3 that solutions corresponding to λ_{17} and λ_{18} do not provide good reconstruction of the scatterer profile. Our proposed method using PCA and SPCA in Fig. 4 is much better. However, if we replot the L-curve in log scale as shown in Fig. 5(b), we can see another knee point around λ_{30}. The solutions corresponding to λ_{30} in Fig. 3 is much closer to actual scatterer profile. However, it is important to note that knee point around λ_{18} is much more steep and hence, without knowledge of ground truth, one can select λ_{18} as optimal value which will lead to wrong result. To overcome this problem of false knee point, more advance versions of L-curve can be explored in future work (Xu et al., 2016).



(a) (b)

Figure 5: (a) L-curve for Ridge regression (12) applied using 40 values of λ. The points around knee point is shown as magnified image on right. The knee point can be seen at λ_{17} and λ_{18} (b) Log scale plot of L-curve shown in (a)

Using previous numerical example, we showed that subspace based methods to tune regularization parameter is a promising approach. It do not require cross validation or Lcurve methods and can provide optimum accurate solution without requiring any validation dataset. The proposed method captures important features from various solutions obtained using multiple regularization parameter. Furthermore, the proposed method is utilizing information from solutions obtained from all n values of λ instead of selecting any one optimum λ.

Fig.6 shows similar results but for different profile. Here we used more intricate shape of the profile by taking one arbitrary sample of digit 3 from the MNIST data set and treat it like a scatterer kept inside the DOI. Fig.6 shows reconstruction of this profile using same process as Fig.3. Once again, it can be seen that first component of PCA and SPCA decomposition can successfully capture the original profile from 40 reconstructions shown in Fig.6.

4 Conclusion

In this report, we present a subspace based method to tune the regularization parameter for solving ill-posed inverse problems. The technique can be used for any regularization formulations which

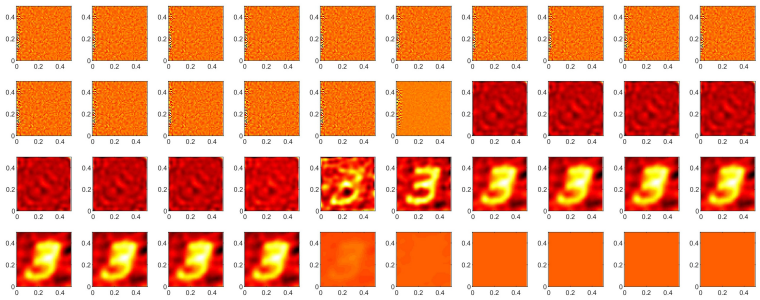
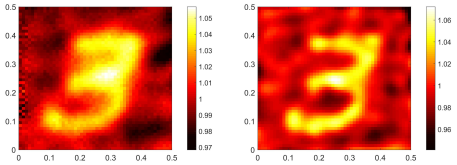


Figure 6: 40 reconstructions obtained by solving Ridge regression in (12) for 40 values of λ.



(a) (b)

Figure 7: (a) First Eigenvector of Sample Covariance Matrix of X in Fig. 6 by solving (10). It captures common features in 40 images shown in Fig. 6. (b) shows principal component using Sparse PCA as formulated in (11)

has a least square data term along with a regularization terms. The proposed method extracts the significant and common structural similarity from the multiple estimates given by multiple values of regularization parameter. This is done by finding direction of maximum variance across the solutions obtained using multiple values of regularization parameter. We also used SPCA instead of PCA to get sparsely constructed result. We also showed that L-curve method can give misleading values of optimum λ in inverse scattering problem considered. The proposed method is a unsupervised way of utilizing multiple solutions given by a regularized regression without need of cross validation with additional labeled data.

5 Future Work

Future work can include:

1. Explore Random Matrix Theory Tools to clean the Sample Covariance Matrix for improved results (Bun et al., 2016; Ledoit and Wolf, 2017, 2004).
2. Combine results from multiple regularization techniques (like sparsity from LASSO and piece-wise continuity from TV regularization) and find common structure in the solutions using proposed method. This can help in enforce multiple priors in the solution.
3. Explore more advance versions of PCA and SPCA for improving the results (Yao, 2021; Udell et al., 2014).

References

GJ Tearney, ME Brezinski, JF Southern, BE Bouma, MR Hee, and JG Fujimoto. Determination of the refractive index of highly scattering human tissue by optical coherence tomography. *Optics Letters*, 20(21):2258–2260, 1995.

T Lu, Krishna Agarwal, Yu Zhong, and Xudong Chen. Through-wall imaging: Application of subspace-based optimization method. *Progress In Electromagnetics Research*, 102:351–366, 2010.

Saandeep Depatla, Lucas Buckland, and Yasamin Mostofi. X-ray vision with only WiFi power measurements using rytov wave models. *IEEE Transactions on Vehicular Technology*, 64(4): 1376–1387, 2015.

Xudong Chen. *Computational methods for electromagnetic inverse scattering*. Wiley Online Library, 2018.

Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 2010.

H. Liu, H. Zhong, N. Karpowicz, Y. Chen, and X. Zhang. Terahertz spectroscopy and imaging for defense and security applications. *Proceedings of the IEEE*, 95(8):1514–1527, 2007. doi: 10.1109/JPROC.2007.898903.

J. Wilson and N. Patwari. Radio tomographic imaging with wireless networks. *IEEE Transactions on Mobile Computing*, 9(5):621–632, 2010.

Christophe Bourlier, Nicolas Pinel, and Gildas Kubické. *Method of moments for 2D scattering problems*. Wiley Online Library, 2013.

RHT Bates, VA Smith, and Ross David Murch. Manageable multidimensional inverse scattering theory. *Physics Reports*, 201(4):185–277, 1991.

Per Christian Hansen and Dianne Prost O'Leary. The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6):1487–1503, 1993.

D Calvetti, S Morigi, L Reichel, and F Sgallari. Tikhonov regularization and the l-curve for large discrete ill-posed problems. *Journal of Computational and Applied Mathematics*, 123(1-2):423–446, 2000.

Yanbin Xu, Yang Pei, and Feng Dong. An extended l-curve method for choosing a regularization parameter in electrical resistance tomography. *Measurement Science and Technology*, 27(11): 114002, 2016.

Jerome Friedman, Trevor Hastie, Robert Tibshirani, et al. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

David Strong and Tony Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse problems*, 19(6):S165, 2003.

Yuan Yao. *CSIC 5011: Topological and Geometric Data Reduction and Visualization Spring 2021*. 2021. URL `https://yao-lab.github.io/2021.csic5011/`.

Yuan Yao. A mathematical introduction to data science. 2004.

Madeleine Udell, Corinne Horn, Reza Zadeh, and Stephen Boyd. Generalized low rank models. *arXiv preprint arXiv:1410.0342*, 2014.

Joël Bun, Romain Allez, Jean-Philippe Bouchaud, Marc Potters, et al. Rotational invariant estimator for general noisy matrices. *IEEE Trans. Information Theory*, 62(12):7475–7490, 2016.

Olivier Ledoit and Michael Wolf. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Studies*, 30(12):4349–4388, 2017.

Olivier Ledoit and Michael Wolf. Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119, 2004.

CONTRIBUTION		
AMARTANSH DUBEY (Student ID: 20413942)	SAMRUDDHI DESKHMUKH (Student ID: 20783488)	NILESH KUMAR JHA (Student ID: 20617354)
Brainstormed and co-developed the idea of using Subspace methods for regularization parameter tuning	Brainstormed and co-developed the idea of using Subspace methods for regularization parameter tuning	Brainstormed and co-developed the idea of using Subspace methods for regularization parameter tuning
Simulated suitable data for the report using Method of Moments code for electromagnetic scattering	Developed framework to solve the inverse scattering problem for a wide range of regularization parameters	Developed framework to implement Sparse PCA and apply it on the inverse scattering data
Wrote Abstract and Section 2 (Proposed method) for the project report	Wrote Introduction and Numerical section of the project report	Wrote Introduction and Numerical section of the project report
Co-developed subspace method codes for analyzing the data, specifically, performing eigendecomposition on the multiple estimates of inverse solution to provide preliminary verification of idea	Converted Matlab codes related to physics based electromagnetic forward and inverse simulations to Python for utilizing wider range of libraries and all in one platform for analysis	Converted Matlab codes related to physics based electromagnetic forward and inverse simulations to Python for utilizing wider range of libraries and all in one platform for analysis
Literature survey on challenges in tuning regularization parameter, specifically in inverse scattering problems and subspace methods. Currently working on developing RMT based techniques to clean covariance matrix for better results	Literature survey on challenges in solving high dimensional ill-posed problem and faster subspace decomposition. Currently exploring Learning based approaches (like NN) for subspace decomposition and dimensionality reduction	Literature survey on various prior enforcing optimization techniques. Currently exploring and working on Bayesian frameworks (variational inference) which can further help in tuning the regularization parameter.