

# Analytics of Unconstrained Convex Problems in Sparse and Redundant Representations: A Case Study on Image Processing Applications

Pooja Patil  
PESIT-Bangalore, South Campus  
Dept. of Electronics and Communication  
Engineering  
Bengaluru, Karnataka, India

Subhash S. Kulkarni  
PESIT-Bangalore, South Campus  
Dept. of Electronics and Communication  
Engineering  
Bengaluru, Karnataka, India

## ABSTRACT

The key objective of sparse and redundant representations is all about the introduction of a highly elegant data model with well-defined mathematical foundations. The structure of this model has been emerged from the conventional transforms represented in redundant unitary form. This draws out a new flavor of treatment to data. The appeal of this model is attributed to compact representation it facilitates. A lot of freedom persists in model adaptation to fit the data depending on the application.

Many models are available aiming to serve the data of interest. Models can be given as mathematical descriptions or the conditions that the underlying signal of interest are believed to obey. Models that are available in image processing are Discrete Cosine Transform (DCT), Principal Component Analysis (PCA), wiener filtering, anisotropic diffusion, etc. Posing a new model requires a delicate attention to attain the simplest possible and reliable model while justifying the actual content of the data. To this Sparseland model emerges out as a new universal data model serving many applications in image processing.

A class of applications in image processing demand the recovery of clean image from the naturally available perturbed images, which are treated as inverse problems. Some examples of such problems are image deblurring, image inpainting, etc. This paper discusses sparse approach to an inverse problem with image deblurring as a case study. The comparison of various shrinkage algorithms supporting sparse approach, used to serve this application is discussed along with the key constraints involved in these algorithms.

## Keywords

Unconstrained convex problems, Sparse solutions, Inverse problems, Optimization, Regularization, Shrinkage algorithms

## 1. INTRODUCTION

As linear system of equations are often encountered in many engineering problems, a vast literature exists on the subject of linear algebra for solving these equations [5], [16], [15] and the knowledge of which can be practically deployed in real time applications. However, in practice, a system of linear equations with matching unknowns are very rare and often ends up being sparse. The sparse solution of the equations also find numerous applications, especially in image processing [25], [9], [4], [17]. The field of sparse representation elicits its structure from the conventional transforms which ensures simplicity of processing, efficiency of representation, speed etc. The main idea of the Sparseland

model is all about dimensionality reduction aiming to provide an alternative and more concise description of the incoming data in an attempt to claim that the data worked upon is simpler than it may appear. Another dimension to this model is redundancy, i.e., using a representation longer than the data. In this context, from image processing perspective, the sparse model is based on the assumption that a given image has a sparse representation w.r.t. a specific redundant dictionary A which is described subsequently [10].

## 2. SPARSE APPROACH TO INVERSE PROBLEMS IN IMAGE PROCESSING

There exists a class of problems in image processing that demand the recovery of clean image from a corrupted version of it. In most applications, the problems are underdetermined. Consider an undetermined system  $Ax=b$  which has more unknowns than equations where A is an  $m \times n$  ( $m < n$ ) matrix called dictionary, b is the image and x is its sparse representation. Under the assumption that A is full rank, this system has infinitely many solutions and hence is impractical to model. This gives rise to the process of regularization in which the basic idea is to assign a penalty  $P(x)$  to every possible solution and select one that gives the lower value. While minimizing w.r.t. x, search is made for the sparsest solution among those that satisfy the equation  $Ax=b$ , i.e.,

$$\min_x P(x) \text{ s.t. } Ax = b \quad (1)$$

Wiener filtering is an approach derived based on Minimum Mean Squared Error (MMSE) criterion. This approach for image restoration needs power spectrum of the undegraded image which is generally not available. Alternately solution is sought using wiener filter by approximating the ratio of power spectrum of degraded to undegraded image [13]. This approach is not suitable. On the contrary, sparse approach to image recovery overcomes this limitation by a combination of  $l_1$  ( $\|\cdot\|_1$ ) and  $l_2$  ( $\|\cdot\|_2$ ) - norm constraints against  $l_2$ -norm constraint used in wiener filtering. This results in a convex optimization problem and hence yields a global minimum solution. Thus, this optimization problem, with reference to equation (1), can be formulated as

$$f(x) = \frac{1}{2} \|b - HAx\|_2^2 + \lambda 1^T \rho(x) \quad (2)$$

where  $\rho(x)$  is a sparsity promoting function and changing the constraint  $Ax=b$  in equation (1) to penalty (equation (2)) will introduce a parameter  $\lambda$ , the choice of the value of which is always controlled by the application.

**Table 1 : Description of various inverse problems**

Si. No.	Applications	Signal model	Problem formulation
1.	Image deblurring	$\hat{y} = Hy + v$	$\hat{x} = \operatorname{argmin}_x \frac{1}{2} \ \hat{y} - HAx\ _2^2 + \lambda \cdot \rho(x)$
2.	Image denoising	$y = y_0 + v$	$\hat{x} = \operatorname{argmin}_x \lambda \ Wx\ _1 + \frac{1}{2} \ \hat{y} - HAx\ _2^2$
3.	Image separation	$y = y_c + y_t + v$	$\hat{x}_c, \hat{x}_t = \operatorname{argmin}_x \lambda \ x_c\ _1 + \lambda \ x_t\ _1 + \frac{1}{2} \ \hat{y} - A_c x_c - A_t x_t\ _2^2$
4.	Image inpainting	$y = My_0$	$\hat{x}_c, \hat{x}_t = \operatorname{argmin}_x \lambda \ x_c\ _1 + \lambda \ x_t\ _1 + \frac{1}{2} \ \hat{y} - MA_c x_c - MA_t x_t\ _2^2$

Various techniques have been proposed to tackle such problems, forming a rich and exciting field of research that stands at the intersection of signal and image processing, estimation theory, optimization, and applied mathematics. With all this well equipped and guaranteed literature, let us now see how some of the image processing problems, tabulated in table (1), which are modeled such that they seek an approximate solution to the proposed model justifying the sparsity criterion.

In table (1), the squared  $l_2$ -norm of the difference between the original image and it's estimate  $HAx$  is the loss or error, the Lagrange multiplier  $\lambda$  is a regularization parameter which controls the amount of penalty, using the sparsity promoting function  $\rho(x)$ , that can be added to the loss to sparsify the solution. Here,  $x$  is the sparse representation of the image,  $H$  is the blur kernel. This will be an identity operator for image denoising and a mask  $M$  containing the missing pixels in case of image inpainting. All these applications are described considering the noise  $v$  because of the perturbations in the naturally available images. More details on these parameters can be found in section 3

Section 3 gives a description of the parameters involved, section 4 addresses the iterative shrinkage algorithms and their acceleration methods, section 5 discusses the constraints on the parameters for practical point of view and section 6 summarizes the observations and conclusions drawn.

### 3. PARAMETERS FOR UNIQUE SOLUTIONS IN SPARSE FRAMEWORK

With all the problem formulations, there is a lot of flexibility in choosing each of the parameters and their constraints. These decisions are critical to the success or failure of the developed arena. This section gives a brief description of the choice of parameters and their constraints as follows:

#### 3.1 Redundant Dictionary Matrix, A

In the concept of Sparseland model the dictionary  $A$  is assumed to be a low rank, unitary and redundant  $m \times n$  ( $m < n$ ) matrix [23], [7], [22]. The optimal approximation error in sparse approximation deteriorates as the number of atoms in the dictionary increases. Thus, the amount of redundancy of the dictionary attributes to the stronger uniqueness. Some examples of the unitary transforms that can be used to create the dictionaries are Hadamard bases, DCT, wavelets such as Haar and its descendants, viz: curvelets [19], [18], contourlets [11], [8], [12] etc.

It is built with transform co-efficients in order to better understand and explore the query data. An ideal transform would lead to set of independent co-efficients while concentrating most of the energy in as few co-efficients as possible. Due to the multitude of choices available, having the effective and right choice has always been controlled by the application, one is working with.

A classical method for assessing the quality of a specific transform  $A$ , as the sparsifying operation, is based on the  $m$ -term approximation. This approximation error is expected to

decay as  $m$  grows. The speed of this decay defines the goodness of the transform for handling the query signals. It can be observed that this process is based on thresholding algorithm as they pursuit in order to evaluate the representation. Clearly this is not the best way to estimate the signal representation. However, it suits best to analyze theoretically [3].

In order to get a mathematical expression for the rate of decay, the true data must be replaced by a mathematical description of it. In the context of images, the model was chosen as piecewise smooth regions separated by piecewise smooth edges where smoothness was quantified in the continuum by the availability of the first two derivatives. For such synthetic images, several transforms were analyzed. The best decay one could desire is  $1/m^2$ . The table (2) presents the decay factor of some of the transforms.

**Table 2 : Various transforms and their decay rates**

Si. No.	Transforms	Decay factor
1.	Fourier	$1/\sqrt{m}$
2.	Wavelets	$1/m$
3.	Curvelets	$\log^3 m/m^2$

A more acceptable approach is to choose flexible and tunable transform such as wavelet packets or bandlets so that the methodology could adapt them to the signals under consideration [3]. Further as these methods are behavior and content driven in these images, the accuracy still remains coarse as image acquisition devices fail to render these requirements. Hence, it necessitates the deployment of the trained dictionaries that enable to adapt this model to practically any source.

#### 3.2 Regularization and Parameters

Generally, the bound and norm of underdetermined system of equations is always ill-posed and ill-conditioned. Regularization ascertains conversion of ill-posed to well-posed via bounded approximation. Similarly, ill-conditioned representation due to large norm is replaced by well-conditioned and small norm using regularization. Thus, regularization ensures numerically stable solution for these modified set of equations [6].

##### 3.2.1 Lagrange Multiplier, $\lambda$

$\lambda$  is a positive constant, called regularization parameter, which controls the amount of penalty that can be imposed on the regularization function  $\rho(x)$  with smoothness constraints and bounds on the vector space norm [1]. The value of  $\lambda$  is found empirically, to yield the best quality of the result, and hence will be different for different applications.

##### 3.2.2 Sparsity Promoting Function, $\rho(x)$

Non-smooth surfaces typically signify lack of convexity. The bounded and lower norm approximation continues to remain smooth resulting into multiple local minima. The objective of securing global minima is supported by sparsity promoting function. This function promotes convexity. Minimizing a

convex function over a convex domain yields an optimal set of solutions which concentrate in a convex set with no local minima points. If the regularization function is strictly convex, then the best solution among the set of solutions for  $Ax=b$  will be a globally unique and globally minimum solution and hence can be implemented using available standard mathematical programming tools [2].

While the choice of  $l_0$  norm for  $\rho(x)$  is non-convex and NP-hard, the  $l_1$ -norm is used. Even though this is not strictly convex, the fact that it provides a sparse solution and the available subgradient methods for solving the same makes it advantageous. The characteristics of  $l_p$ -norm for  $p=0.1, 0.5, 1, 2$  is as shown in the figure (1) [10].

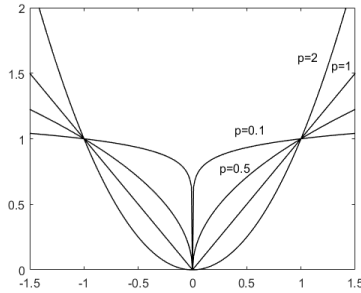


Fig 1: Behavior of  $|x|_p$  for different values of  $p$

### 3.2.3 Hessian Constraint, $c$

Convexity demands hessian to be positive definite in the constrained equation. This objective is met with hessian constraint  $c$  satisfying the condition  $cI - A^T A > 0$ ; the violation of this condition results in the unboundedness of  $(A^T A)^{-1}$  which can cause severe numerical instabilities unless additional constraints are added [6]. This is satisfied by the choice  $c > \|A^T A\|_2 = \lambda_{\max}(A^T A)$ .

## 4. ITERATIVE SHRINKAGE APPROACH AND ALGORITHMS FOR INVERSE PROBLEMS

With all the problem In order to get a mathematical expression for the rate of decay, the true The class of sparse solution is exercised for inverse problem with ill-posed linear set of equations. The solution is sought via a redundant dictionary matrix representation characterized by linear algebraic full rank property.

The constraints depicted in equation (2) should ensure unique, optimal and possess fast iterative solution. The governing factors to these objectives summarize the following observations: (i) increased precision demanding increased redundancy in dictionary matrix which also entails increased sparseness in estimated  $x$  ( $\hat{x}$ ) (ii) Generally undercomplete system of linear equation can be iteratively solved using gradient descent for seeking minimum. To increase the speed, parallel co-ordinate descent can also be sought [20], (iii) To reduce deviation in descent due to non-smoothness, optimization via convexity is the solution, (iv) To have unique solution through optimization approach, a combination of  $l_2, l_1$  and/or  $l_p$ -norms are suitable, (v) To have increased sparsity, shrinkage is forced. These class of algorithms mainly focused on sparsity have been categorized as shrinkage algorithms.

In this work, out of many possible approaches, the solution to linear inverse problems in sparse domain is sought with optimal redundant dictionary. Three algorithmic classes of Surrogate Separable function (SSF), Parallel Co-ordinate

Descent (PCD) and Iterative Reweighted Least Squares (IRLS) cases are studied. As the descent is on non-smooth one, regularization is a need with a combinational norm and constrained optimization. This work will consider these cases with a case study on image deblurring application.

The unconstrained convex methods, also known as subset selection methods [23]. The key objective of these algorithms is to minimize the functional in equation (2). Minimization of this optimization problem can be addressed effectively using a family of algorithms, called iterative shrinkage methods. Roughly speaking, in these iterative methods, each iteration comprises of a multiplication by matrix  $A$  and its adjoint along with a scalar shrinkage step on the obtained result. Despite their simple structure, these algorithms are shown to be very effective in minimizing the convex optimization problems.

The three algorithms discussed are subjected to performance evaluation with image deblurring as a case study with qualitative and quantitative evaluation are presented in this paper.

Qualitative evaluation includes Peak Signal to Noise Ratio (PSNR), Improvement Signal to Noise Ratio (ISNR) and Structural Similarity (SSIM). All the three are used in this paper as individually each one has a drawback. PSNR not all time guarantees true assessment, ISNR is a relative measure with respect to quality of the image. SSIM is opted in recent years as an effective assessment metric [24]. SSIM is assessed over  $[0, 1]$  where value '1' corresponds to high quality.

## 4.1 Surrogate Functions and the Prox Method

The constrained solution to linear inverse problem can be sought by iterative substitution process. The minimizing functional to such implementation in inverse problems are called surrogate functionals. Further as dictionary matrix  $A$  is unitary and orthogonal, such functions with dictionary matrices are categorized as separable surrogate functionals that are subjected to global minimization. However, these may demand additional regularization, convex and shrinkage conditioning to obtain optimal and sparse solution.

A class of minimization functional for solving linear inverse problem that has been prevailing in sparse theory is a surrogate functional that approximates an objective function

$$\hat{f}(x) = \frac{1}{2} \|b - HAx\|_2^2 + \lambda 1^T \rho(x) + \frac{c}{2} \|x - x_0\|_2^2 - \frac{1}{2} \|Ax - Ax_0\|_2^2 \quad (3)$$

An iterative approximation facilitating minimization functional with replacement approach is termed as surrogate functional. It is computationally observed to be an efficient and less complex. It works effectively even when the objective function suffers from non-smoothness. It attempts towards finding global minima for the objective function optimally. Exploring global minima combined with speed characterizes the surrogate function based algorithm [14]. Equation (3) corresponds to modified minimization function in equation (2) with added function  $d(\cdot)$  that promotes convexity which is given as

$$d(x, x_0) = \frac{c}{2} \|x - x_0\|_2^2 - \frac{1}{2} \|Ax - Ax_0\|_2^2 \quad (4)$$

The new objective function will thus be

$$f(\hat{x}) = \frac{1}{2} \|b - HAx\|_2^2 + \lambda 1^T \rho(x) + d(x, x_0) \quad (5)$$

## 4.2 Parallel Co-ordinate Descent (PCD)

PCD uses a functional for minimization with an approach to update individual elements in the estimate of  $x$  or minimization functional is updated every time for every new element estimate in  $x$ .

$$g(x) = \frac{1}{2} \|b - Ax_0 - a_i(x - x_0[i])\|_2^2 + \lambda \rho(x) \quad (6)$$

Generally gradient descent solution heavily depends on initial value that one has to choose. This limitation is overcome by PCD where initial value of  $x$  can be zero.

This iterative shrinkage approach typically begins with several simpler co-ordinate descents and merging into joint steps. The algorithm starts by proposing a Coordinate Descent (CD) algorithm that updates one entry at a time in  $x$  while keeping the rest untouched. A sequence of such rounds of  $m$  steps (addressing each of the entries in  $x \in \mathbb{R}^m$ ) is necessarily converging. While each co-ordinate direction assumes descent, their combination needs to be properly scaled. An additional step with directional Line Search (LS) typically assumes expected descent

Compared to the SSF, PCD effectively depends on atom normalization in dictionary matrix  $A$  for accurate reweighting of back projected error. Also the descent is driven by the acceleration methods.

## 4.3 Iterative Reweighted Least Squares (IRLS)

Equation (2) is minimized for  $f(x)$  through  $l_2$ -norm weighting factor as required in place of  $1^T \rho(x)$  as the original functions are generally non- $l_2$ -norms. Accordingly the modified functional will be

$$f(x) = \frac{1}{2} \|b - HAx\|_2^2 + \frac{\lambda}{2} x^T W^{-1}(x)x \quad (7)$$

where  $W(x)$  is a diagonal matrix

$$W[k, k] = 0.5 \cdot x[k]^2 / \rho(x[k]) \quad (8)$$

Solution to IRLS is possible only by  $l_2$ -norm. This needs conversion of non- $l_2$ -norms to  $l_2$  and is achieved by iteratively weighting factor  $W$ . This feature makes IRLS more preferable. IRLS is sensitive to zero initialization and care need to be taken to avoid this initialization. Further, the algorithm fails to reveal even with isolated zeros in intermittent solutions. These isolated zeros correspond to the control exercised by local minima.

$$f(x) = \frac{1}{2} \|b - HAx\|_2^2 + \lambda 1^T \rho(x) = \frac{1}{2} \|b - HAx\|_2^2 + \frac{\lambda}{2} x^T W^{-1}(x)x \quad (9)$$

leading to the iterative equation

$$S \left( \frac{1}{c} A^T (b - Ax_k + x_k) \right) \quad (10)$$

## 4.4 Acceleration Methods

Speed-up or acceleration methods can be implemented in all the algorithms discussed in two ways. Firstly, Line Search facilitates in all the algorithms as discussed in PCD algorithm. Secondly, a technique called Sequential Subspace

Optimization (SESOP) provides additional advantage of speeding up and is preferred when given a choice.

In Line Search approach for acceleration, the updates on  $x_{k+1}$  vector, obtained from  $x_k$ , requires an error function to be added to  $x_k$  in each iteration. Error function is dependent on gradient measure and descent direction. In order to have guaranteed descent eventually resulting into optimal minimization of the objective function, it necessitates to have scaling factor with error function. Generally, scaling factor  $\mu$  is a constant or can be derived from regularization and convexity enforcing parameters along with sparse vector  $x_k$  to obtain updated  $x_{k+1}$ .

SESOP seeks update on  $x_{k+1}$  via optimization of minimization functional over an affine subspace spanned in  $q$  recent steps along with prevailing gradient.  $q+1$  dimensional optimization is performed with Newton algorithm. SESOP guarantees no additional cost with  $q+1$  dimensional computation as these can be re-stored in previous iterations (similar to recycled storage).

## 5. NUMERICAL IMPLEMENTATION AND ISSUES

Place Consider a blurred and noisy version of the ideal image  $z=Ha+v$  where the blur kernel  $H$  and the noise variance  $\sigma$  are assumed to be known as this knowledge helps us to achieve the better estimate of the solution. Sparse model assumes that an ideal image can be represented as a product of dictionary matrix 'A' with sparse vector 'x'. Hence restoration corresponds to solution of optimization problem given in (2).

The description of each of the parameters is as follows:  $A$  is a full rank  $m \times n$  ( $m < n$ ) matrix under which the signal representation  $x$  (of size  $n \times 1$ ) is known to be sparse. It is built with the  $2 \times 2$  masks  $[0.5, 0.5; 0.5, 0.5]$ ,  $[0.5, 0.5; -0.5, -0.5]$ ,  $[0.5, -0.5; -0.5, 0.5]$  and  $[0.5, -0.5; 0.5, -0.5]$ . These are extended for  $4 \times 4$  masks. The number of rows ( $m$ ) should be equal to the length of the vector ( $\mathbb{R}^m$ ) containing the pixels of the available image (image with perturbations). This matrix is constructed by creating the vectors of length  $R_m$ , called atoms, one-by-one and inserting them column-wise and are normalized to have unit norm.

The noise  $v$  is assumed to be a random vector of size  $R_n$  containing white Gaussian noise as this turns the maximum likelihood criterion to least-squares criterion which is practically implementable [21].

The blur kernel  $H$  used here is a  $15 \times 15$  ( $m_H \times n_H$ ) filter with values  $1/(i^2 + j^2 + 1)$  for  $-7 \leq i; j \leq 7$ , normalized to have unit sum. The circulant version of  $H$  of size  $400 \times 400$  when multiplied with the ideal image results in the circulant blurred image. This can be avoided by appending the last ( $n_H/2$ ) columns of the ideal image ( $y_0$ ) at the beginning and the first ( $n_H/2$ ) columns of  $y_0$  at the end and the similar treatment to the rows of the resultant image yields a  $35 \times 35$  image. this is then convolved with  $H$  and only the part of the convolution computed without the zero padded edges is considered.

The sparsity promoting function  $\rho(x)$  used for all these algorithms is  $\rho(x) = |x| - \text{slog}(1+|x|/s)$  where  $s=0.01$  is a constant.

A generalized iterative shrinkage algorithm for solving the inverse problems comprise the following steps:

**Initialization:** For all the algorithms, the initial signal representation  $x$  is initialized to be zero. For IRLS,  $x$  is assumed to be a vector of ones.

**Back projection:**  $e = A^T r$

**Shrinkage:**  $e_s = S(x + e/c)$

For SSF, the values of  $e$  are weighted by  $(1/c)$ .

For PCD  $e$  is weighted by random iid (independent and identically distributed) values are used.

For IRLS these weights depend on each entry of  $x$

**Line Search or SESOP:** These are the acceleration methods to increased speed of convergence and to

**Updating the solution:** For  $k^{\text{th}}$  iteration  $x_k = x_{k-1} + \mu(e_s - x_{k-1})$

**Updating the residual:**  $r = b - HAx$

Here, the shrinkage function  $S$  maps each value of the input vector to the desired output value  $\hat{x}_{\text{opt}}$ . The proper design of this function ensures a globally minimum solution even though the input vector is non-convex. The shrinkage function used is

$$\hat{x}_{\text{opt}} = \frac{(x - s - \lambda) + \sqrt{(s + \lambda - x)^2 + 4sx}}{2}$$

More details on each of these algorithms can be found in [10].

## 5.1 Convergence Issues

The minimization objective function in equation (2) is unconstrained and open ended function that lacks unique solution in its generalized form having pure  $l_2$ -norm penalty term. Driving this objective penalty function to meet desired applications, multiple additional terms are required such as regularization terms, convexity norms etc. (2) corresponds to penalty term added with convex enforcing  $l_1$ -norm.

Optimal sparse solution for estimating sparse vector  $x$  is sought in two steps. Step 1 corresponds to promoting convergence with optimal estimation of  $x$ . Step 2 is an additional enforcement on  $x$  with hard thresholding that increases sparseness in vector  $x$ .

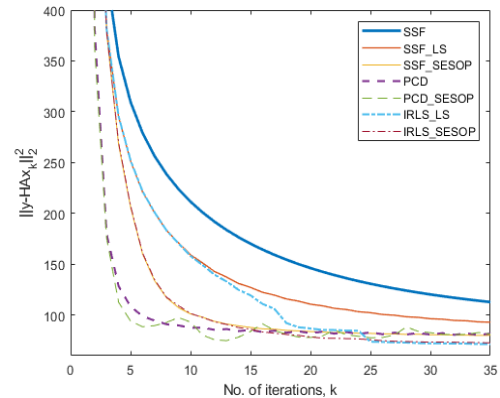
The key issue in seeking sparse solution can be attributed to imposing optimal convexity that guarantees convergence. The objective function in equation (2) and optimal sparse vector  $x$  estimation as in equation in section 5 have to be convex specific. To this end, robust formulations in choosing regularization  $\lambda$ , imposing additional norm terms along with sparsity promoting function  $\rho(x)$  and shrinkage with hard thresholding have been well established in literature [10]. Hence convergence has never been a major issue as long as convexity enforcing parameters combined with sparsity promoting functions are suitably embedded in the objective function. In addition, computation efficiency and speed-up can also be addressed in multiple ways within different choices of algorithms.

## 6. RESULTS AND INFERENCES

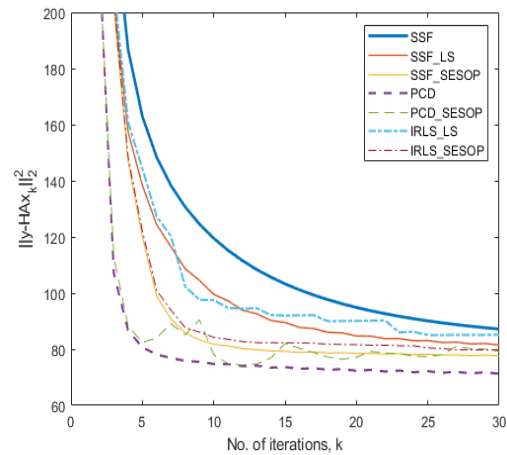
The experimented results are presented graphically, the behavior of all the discussed algorithms showing the squared  $l_2$ -norm of the error (in the objective function), as a function of iteration for the resized  $60 \times 60$  cameraman, football and lena images. These plots are very similar to the  $20 \times 20$  and  $40 \times 40$  images.

Among all the algorithms and their acceleration methods, all the three algorithms with SESOP acceleration along with PCD gives the better results. It can be seen from the figures (2, 3

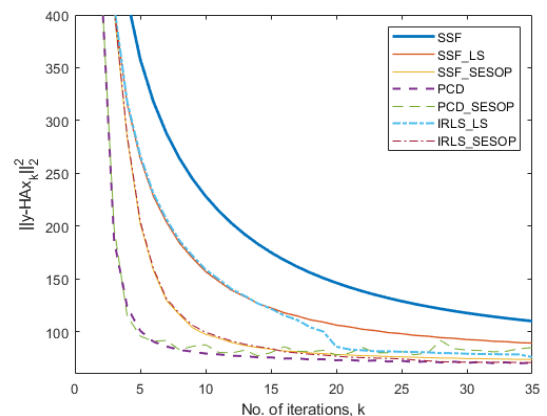
and 4) that among these algorithms, PCD converge faster for about 10 iterations with an improvement of about 7.3dB, 5.3dB and 8dB in PSNR value for cameraman, football and lena images respectively



**Fig 2: The value of the squared  $l_2$ -norm of the error (in the objective function), as a function of iteration for resized  $60 \times 60$  cameraman image**



**Fig 3: The value of the squared  $l_2$ -norm of the error (in the objective function), as a function of iteration for resized  $60 \times 60$  football image**



**Fig 4: The value of the squared  $l_2$ -norm of the error (in the objective function), as a function of iteration for resized  $60 \times 60$  lena image**

**Table 3 : Image deblurring using various Iterative Shrinkage algorithms and their quality measures for a 20 × 20, resized cameraman image with PSNR=18.088641**

Sl. No.	Algorithms	$\lambda$	Output PSNR (dB)	ISNR (dB)	SSIM	No. of iterations	Time elapsed (seconds)
1.	SSF	0.031	25.435446	7.846805	0.789356	230	2.448662
2.	SSF-LS	0.02	25.417816	7.329175	0.777811	130	2.837949
3.	SSF-SESOP-5	0.04	25.583603	7.494961	0.789401	30	1.966703
4.	PCD	0.05	25.805269	7.716627	0.788631	28	4.452650
5.	PCD-SESEOP-5	0.07	25.142301	7.053660	0.780596	25	4.063416
6.	IRLS	0.01	25.644224	7.555582	0.793015	60	2.622169
7.	IRLS-SESEOP-5	0.01	25.598283	7.509641	0.791428	40	3.128366

**Table 4.: Image deblurring using various Iterative Shrinkage algorithms and their quality measures for a 40 × 40, resized cameraman image with PSNR=19.52821**

Sl. No.	Algorithms	$\lambda$	Output PSNR (dB)	ISNR (dB)	SSIM	No. of iterations	Time elapsed (seconds)
1.	SSF	0.05	26.556837	7.028621	0.742670	250	8.953577
2.	SSF-LS	0.05	26.844592	7.316376	0.741316	200	12.582175
3.	SSF-SESOP-5	0.04	26.782034	7.253818	0.741122	30	3.925382
4.	PCD	0.07	26.854199	7.325983	0.747544	35	17.904596
5.	PCD-SESEOP-5	0.09	26.586498	7.058282	0.740936	25	18.748126
6.	IRLS	0.01	27.059598	7.531382	0.734499	60	5.011519
7.	IRLS-SESEOP-5	0.01	27.032880	7.504664	0.738624	40	5.000347

**Table 5 : Image deblurring using various Iterative Shrinkage algorithms and their quality measures for a 60 × 60, resized cameraman image with PSNR=20.112524**

Sl. No.	Algorithms	$\lambda$	Output PSNR (dB)	ISNR (dB)	SSIM	No. of iterations	Time elapsed (seconds)
1.	SSF	0.05	27.346605	7.234081	0.683433	250	45.167878
2.	SSF-LS	0.05	27.13882	7.026358	0.685160	100	24.316454
3.	SSF-SESOP-5	0.06	27.405975	7.293451	0.681953	50	15.345012
4.	PCD	0.09	27.425877	7.313353	0.700659	35	84.304598
5.	PCD-SESEOP-5	0.09	27.571170	7.458646	0.689928	25	82.771071
6.	IRLS	0.01	27.721745	7.609221	0.666722	50	14.591339
7.	IRLS-SESEOP-5	0.01	27.660806	7.548282	0.664549	35	12.383515

For a quantitative analysis of the obtained results, the algorithm uses PSNR, ISNR and SSIM and these values for cameraman, football and lena images each with dimensions 20 × 20, 40 × 40 and 60 × 60 are tabulated in the Tables (3-5). Due to space constraint, only tabulated results for Cameraman are presented. This work gave an improvement of 6-7dB in PSNR for all the three images of dimensions 20 × 20, 40 × 40 and 60 × 60 respectively. (The experiments here are restricted to the smaller dimensions of the image because of the constraints on the cache).

These algorithms are tuned with  $\lambda$  to get the better results and are stopped at the iteration that gives a comparatively better values of PSNR and SSIM. However, if proceeded or stopped at the earliest, the value of these measures deteriorates.

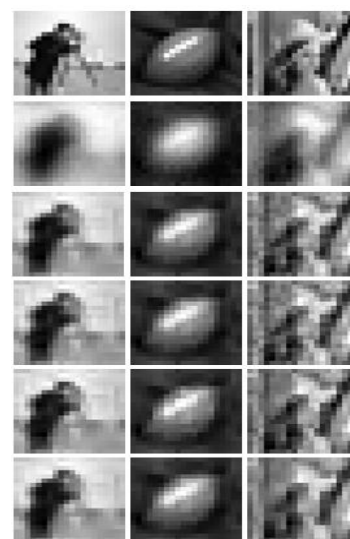
Finally, the results of all the algorithms for all the three images of dimensions 20 × 20, 40 × 40 and 60 × 60 are presented. As stated above, all the algorithms with SESOP acceleration gives better results.

## 7. NOTE ON SPARSITY AND IT'S RELEVANCE IN RESEARCH

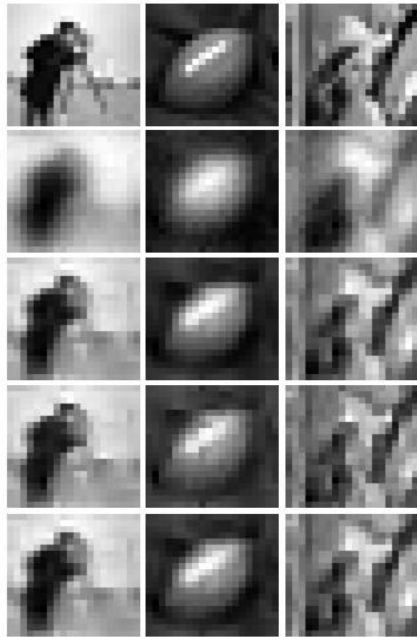
With the availability of vast redundant atomized dictionary matrix, several possible approximations of sparse vector for the target image are possible. Additionally, with multiple avenues in choosing the optimal approach the objective minimization functional can be suitably defined within sparsity framework for optimal approximation. Further, the choice of kernels in dictionary atoms along with optimal framework can be suitably steered to cover large applications.

Thus, there is a lot of freedom in how to relate Sparseland

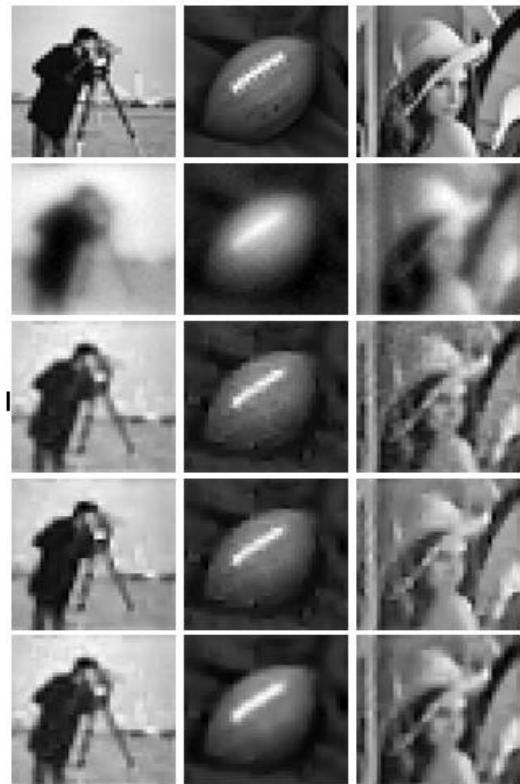
model to the real world problems. The path from theory to practice is not direct and hence fitting various applications to this model is the art. Several applications have emerged from past decades based on this model. With this vast potential offered by sparse theory combined with multitude of applications, especially to linear inverse problems, can be seen as an independent area of study.



**Fig 5: Original resized 20 × 20 cameraman, football and Lena images (First row), their blurred versions (second row), Restored images using: SSF (third row), SSF-LS (fourth) PCD-LS (fifth) and IRLS-LS (sixth)**



**Fig 6:** Original resized  $20 \times 20$  cameraman, football and Lena images (First row), their blurred versions (second row), Restored images using: SSF-SESOP (third row), PCD-SESOP (fourth) and IRLS-SESOP (fifth)



**Fig 8:** Original resized  $40 \times 40$  cameraman, football and Lena images (First row), their blurred versions (second row), Restored images using: SSF-SESOP (third row), PCD-SESOP (fourth) and IRLS-SESOP (fifth)



**Fig 7:** Original resized  $40 \times 40$  cameraman, football and Lena images (First row), their blurred versions (second row), Restored images using: SSF (third row), SSF-LS (fourth) PCD-LS (fifth) and IRLS-LS (sixth)



**Fig 9:** Original resized  $60 \times 60$  cameraman, football and Lena images (First row), their blurred versions (second row), Restored images using: SSF (third row), SSF-LS (fourth) PCD-LS (fifth) and IRLS-LS (sixth)



**Fig 10: Original resized  $60 \times 60$  cameraman, football and Lena images (First row), their blurred versions (second row), Restored images using: SSF-SESOP (third row), PCD-SESOP (fourth) and IRLS-SESOP (fifth)**

## 8. REFERENCES

- [1] Christopher M Bishop. Pattern recognition and machine learning. Springer, 2006.
- [2] Stephen Boyd and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.
- [3] Emmanuel J Candès and David L Donoho. New tight frames of curvelets and optimal representations of objects with piecewise  $c_2$  singularities. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(2):219–266, 2004.
- [4] Chunhong Cao, Jie Yu, Chengyao Zhou, Kai Hu, Fen Xiao, and Xieping Gao. Hyperspectral image denoising via subspace-based nonlocal low-rank and sparse factorization. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019.
- [5] Cuthbert Daniel and Fred S Wood. Fitting equations to data: computer analysis of multifactor data. John Wiley & Sons, Inc., 1980.
- [6] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.
- [7] Ronald A DeVore. Nonlinear approximation. *Acta numerica*, 7:51–150, 1998.
- [8] Minh N Do and Martin Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on image processing*, 14(12):2091–2106, 2005.
- [9] Weisheng Dong, Lei Zhang, Guangming Shi, and Xin Li. Nonlocally centralized sparse representation for image restoration. *IEEE transactions on Image Processing*, 22(4):1620–1630, 2012.
- [10] Michael Elad. Sparse and redundant representations: from theory to applications in signal and image processing. Springer Science & Business Media, 2010.
- [11] Ramin Eslami and Hayder Radha. The contourlet transform for image denoising using cycle spinning. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, volume 2, pages 1982–1986. IEEE, 2003.
- [12] Ramin Eslami and Hayder Radha. Translation-invariant contourlet transform and its application to image denoising. *IEEE Transactions on Image Processing*, 15(11):3362–3374, 2006.
- [13] Rafael C Gonzalez, Richard E Woods, et al. Digital image processing [m]. Publishing house of electronics industry, 141(7), 2002.
- [14] H-M Gutmann. A radial basis function method for global optimization. *Journal of global optimization*, 19(3):201–227, 2001.
- [15] Roger A Horn and Charles R Johnson. Matrix analysis. Cambridge university press, 2012.
- [16] Marvin Marcus and Henryk Minc. A survey of matrix theory and matrix inequalities, volume 14. Courier Corporation, 1992.
- [17] Jesus Pineda, Jhacson Meza, Erik M Barrios, Lenny A Romero, and Andres G Marrugo. Noise-robust processing of phase dislocations using combined unwrapping and sparse inpainting with dictionary learning. In *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, pages 1–5. IEEE, 2019.
- [18] Jean-Luc Starck, Emmanuel J Candès, and David L Donoho. The curvelet transform for image denoising. *IEEE Transactions on image processing*, 11(6):670–684, 2002.
- [19] Jean-Luc Starck, Fionn Murtagh, and Jalal M Fadili. Sparse image and signal processing: wavelets, curvelets, morphological diversity. Cambridge university press, 2010.
- [20] Rachael Tappenden, Martin Takáč, and Peter Richtárik. On the complexity of parallel coordinate descent. *Optimization Methods and Software*, 33(2):372–395, 2018.
- [21] Albert Tarantola. Inverse problem theory and methods for model parameter estimation, volume 89. siam, 2005.
- [22] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.
- [23] Aaron Tropp. Topics in sparse approximation. Joel PhD thesis, 2004.
- [24] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [25] Joaquin Zepeda, Christine Guillemot, and Ewa Kijak. Image compression using sparse representations and the iteration tuned and aligned dictionary. *IEEE Journal of Selected Topics in Signal Processing*, 5(5):1061–1073, 2011.