

Adaptive Algorithm in Image Reconstruction Based on Information Geometry

Meng Wang¹, Zhen Hu Ning^{2*}, Jing Yu², and Chuang Bai Xiao²

¹ School of Applied Science
Beijing Information Science and Technology University
Beijing, 100192, China
[e-mail: falnwu@163.com]

² College of Computer Science, Faculty of Information Technology
Beijing University of Technology
Beijing, 100124, China
[e-mail: ningzhenhu@126.com, boosherboo@126.com, boobooboo@126.com]
*Corresponding author: Zhen-Hu Ning

*Received July 30, 2019; revised November 27, 2019; revised June 9, 2020; accepted November 17, 2020;
published February 28, 2021*

Abstract

Compressed sensing in image reconstruction has attracted attention and many studies are proposed. As we know, adding prior knowledge about the distribution of the support on the original signal to CS can improve the quality of reconstruction. However, it is still difficult for a recovery framework adjusts its strategy for exploiting the prior knowledge efficiently according to the current estimated signals in serial iterations. With the theory of information geometry, we propose an adaptive strategy based on the current estimated signal in each iteration of the recovery. We also improve the performance of existing algorithms through the adaptive strategy for exploiting the prior knowledge according to the current estimated signal. Simulations are presented to validate the results. In the end, we also show the application of the model in the image.

Keywords: Information Geometry, Image Reconstruction, Compressed Sensing

1. Introduction

Compressed Sensing (CS) is a signal processing technique to recover sparse signals of interest from far fewer samples and has been widely used in image reconstruction due to the booming of media [1]. Chai et al. proposed an image encryption algorithm based on the memristive chaotic system, elementary cellular automata and CS [2]. Zha et al. proposed a group-based sparse representation method with non-convex regularization for image CS reconstruction [3]. Furthermore, CS was applied to reconstruction of under-sampled atomic images [4]. In 2018, the scholars proposed an image encryption method integrating CS and detour cylindrical diffraction [5].

The CS is to acquire and reconstruct a signal $x \in R^N$ by finding a solution to under-determined linear system [6-8]

$$y = \Phi x + e \quad (1)$$

where $y \in R^M$ is the compressed representation of x , $\Phi = (\phi_{ij})_{M \times N}$ is the random projection matrix, $e \in R^M$ represents the noise, and $M \ll N$. x is inherently sparse or sparse in some domain. That is to say, most of its entries are zero, or there exists an orthogonal transformational matrix Ψ satisfying most of the entries of $\Psi^{-1}x$ are zero. One may recover an estimated \hat{x} as the solution to a l_0 -minimization problem P0 [9], which is discontinuous and NP-hard. Considering that under the restricted isometry property [10], the P0 is usually relaxed to :

$$P1 \quad \hat{x} = \arg \min_{x \in R^N} \|x\|_{l_1} \quad s.t. \quad y = \Phi x \quad (2)$$

To solve the two problems (P0 and P1), many achievements have been obtained. Recovery frameworks fall into three categories: greedy algorithms [11], convex optimization algorithms [12-14], and statistical algorithms [15-17].

Because of lower computational complexity, greedy algorithms are used widely in CS. A typical greedy algorithm is Orthogonal Matching Pursuit (OMP) [18]. OMP can accommodate various scenarios. And it can exactly recover a signal with S nonzero entries in dimension N by $O(S \ln N)$ random linear measurements. So OMP has been studied by many scholars and a family of variants are supported, such as Stagewise Orthogonal Matching Pursuit (StOMP) [19], Regularized Orthogonal Matching Pursuit (ROMP) [20], Compressed Sampling Matching Pursuit (CoSaMP) [21], Iterative Hard Threshold (IHT) [22], Gradient Descent with Sparsification (GraDeS) [23] and Generalized Orthogonal Matching Pursuit (GOMP) [24], and so on.

A well-known convex optimization algorithm is the Least Absolute Shrinkage and Selection Operator (LASSO) [25]

$$\tilde{x} = \arg \min_{x \in R^N} \frac{1}{2} \|y - \Phi x\|_{l_2}^2 \quad s.t. \quad \|x\|_{l_1} \leq t \quad (3)$$

where $t \geq 0$ is a tuning parameter. A stage-wise fast LASSO algorithm for the image reconstruction from CS optimizes an insensitive Huber objective function to achieve a decision function [26]. Lian et al. [27] exploit a multilevel prior support information model and incorporate it into the LASSO using a weighted l_1 -norm penalty function. Least Angle

Regression (LARS) provides an efficient algorithm for computing the solution paths of LASSO [28]. LARS is an algorithm for fitting linear regression models to high-dimensional data [29]. In 2007, Keerthi et al. [30] applied a fast tracking algorithm for LARS to sparse kernel logistic regression.

In many applications, it may be possible to catch some prior knowledge about the vector representation of a natural image in the domain of some linear transform [31, 32]. For example, it can be known that some regions are more likely to be highlight than the others in magnetic resonance images [33]. In statistical algorithms, adding the prior knowledge to CS can improve the quality of reconstruction. There are many studies on sparse recovery with prior knowledge, which enhance the signal recovery by statistic in various degrees. Next, we introduce previous typical statistical algorithms. Babacan et al. [34] proposed an excellent greedy algorithm using Laplace priors based on Bayesian framework. Besides that, the main idea of model-CS is to introduce an extra structure to exploit the distribution of the original signal [35]. In [36], a method based on approximate message passing and Markov-tree prior was supported. Furthermore, neural networks were trained to learn the prior knowledge for recover signal [37]. However, the computational complexity is high because of the extra structure as a model. For the tree based algorithms, the reconstruction involves the travel across the whole tree in every iteration. As far as we know, the average complexity per iteration for the fastest tree-based algorithm for CS is $O(MN^2)$ [38]. In addition, the algorithm based on neural networks involves the forwarding from the bottom layer to the top layer of the networks. And the lower bound on average complexity per iteration for the algorithm is $O(2N)$ for a single forwarding computation with a simplest architecture of a hidden layer of $2[(\alpha-2)/M+1]$ hidden neurons, where α is the size of the training set and $\alpha \gg N$ [39]. Instead of the models, a modified method works by exerting punishment on the selection for a wrong support [40]. In [41], relations between entries of the original signal were built by Kalman filter based on the assumption that support changes slowly. But the Kalman filter does not always use the actual co-variances of the changes of the estimated signals. Furthermore, weighted minimization provided better upper bounds on the reconstruction error where the weights were determined by prior knowledge about a specific probability of nonzero entries [42, 43]. In 2018, the weights were chosen to minimize the expected statistical dimension of the descent cones of a weighted cross-polytope [44]. Moreover, such optimal weights have been used to improve greedy algorithms and convex optimization algorithms [45], such as the Weighted Orthogonal Matching Pursuit (WOMP) and Weighted LARS (W-LARS). The algorithms combine either greedy algorithms or convex optimization algorithms with statistic by weights. In the algorithms, a new estimated signal is generated in an iteration according to the result from its previous iterations. However, the optimal weights are fixed in all iterations in the whole procedure of the reconstruction without the consideration of the changing of the estimated signal. Thus, it is more reasonable to adjust the strategy for exploiting the prior knowledge according to the current estimated signal.

Information geometry [46] is a mathematical branch that applies the techniques of differential geometry to study probability theory and statistic. It has been widely used to many applications, such as asymptotic theory of statistical inference [47], the Expectation-Maximization (EM) algorithm [48], the learning of neural networks [49-51], and many others [52-56]. Wang et al. proposed a model-CS method with neural networks based on information geometry [57]. However, the computational complexity is high because of the neural networks as an extra structure.

To address the problems of the previous algorithms mentioned above, we introduce

information geometry to improve the performance of the greedy algorithms and convex optimization algorithms with no extra structure. It is possible to reflect the image features with a specific manifold structure [58]. Based on the fact, we evaluate the candidates of the estimated signals with the Fisher distance [59] between the distribution over estimated support set and the known distribution over the true support on the original signal. Then, we choose the recovery signal which has a distribution close to the true one according to the evaluations. The strategy can be seen as adaptive to the developed support sets generated in different iterations, because we calculate the estimated distribution according to the current estimated support set in each iteration. Experiments are presented to validate the results.

2. Methods

2.1 Construction of Geometry Model

As discussed above, the true distribution of the indexes of the non-zero entries of x is able to be known by statistic. The index distribution can be described by a specific Probability Density Function (PDF) $p(k | \theta)$ where $\theta = (\theta_1, \theta_2, \dots, \theta_N)'$.

As prior knowledge, we know that the true parameters for the original signals are θ^1 . In addition, θ^2 are estimated parameters in the model. In the iterative algorithm, we estimate the distribution by computing θ^2 over the set of optional indexes in each iteration. We choose the recovery signal in the sense that its support has a distribution closer to the known distribution. The details are described as follows.

A distribution from the exponential family [60] is defined as a function in the following form

$$p(k | \theta) = \exp\{\theta \cdot k - \varphi(\theta)\} \quad (4)$$

where φ is a function of θ .

As a discrete distribution, the position distribution is from the exponential family naturally. Suppose that the frequency of the occurrence of index $i \in \{0, 1, \dots, N\}$ is $p(k = i | \theta) = p_i$, where $\theta = (\theta_1, \theta_2, \dots, \theta_N)'$. And we denote the i th element of θ as θ_i . Then, it can be rewritten as

$$p(\xi | \theta) = \exp\{\theta \cdot \xi - \varphi(\theta)\} \quad (5)$$

where

$$\xi_i = \delta_i(k) = \begin{cases} 1, & k = i \\ 0, & k \neq i \end{cases} \quad (6)$$

$$\theta_i = \log \frac{p_i}{p_0} \quad (7)$$

then,

$$p_0 = 1 - \sum_{i=1}^N p_i = 1 - p_0 \sum_{i=1}^N e^{\theta_i} \quad (8)$$

From (8), we obtain

$$\varphi(\theta) = \log(1 + \sum_{i=1}^N e^{\theta_i}) \quad (9)$$

A classic parametric space for this family of PDF's is

$$H = \{\theta \in R^N \mid \theta \geq 0\} \quad (10)$$

The Fisher Information Matrix (FIM) is defined as

$$G(\theta) = [g_{ij}(\theta)] \quad (11)$$

$$g_{ij}(\theta) = E \left[\frac{\partial \log p(k \mid \theta)}{\partial \theta_i} \cdot \frac{\partial \log p(k \mid \theta)}{\partial \theta_j} \right] \quad (12)$$

where E means the expectation. The FIM is a way of measuring the amount of information of the parameter. And a distance called Fisher distance arises from the FIM [61]. The distance between two points $\theta^1 = \theta(t_1)$ and $\theta^2 = \theta(t_2)$ in the half-plane H measures the dissimilarity between the associated distributions $p(x \mid \theta^1)$ and $p(x \mid \theta^2)$. Furthermore, the Fisher distance is given by the value which is the minimum of the lengths of all the piece-wise smooth paths $\theta(t)$ that joins θ^1 and θ^2 , $t_1 \leq t \leq t_2$. That is

$$D_F(\theta^1, \theta^2) := \min_{\{\theta(t); \theta(t_1) = \theta^1, \theta(t_2) = \theta^2\}} \int_{t_1}^{t_2} \sqrt{\left(\frac{d\theta}{dt} \right)^T G(\theta) \frac{d\theta}{dt}} dt \quad (13)$$

where t is the parameter of the curve $\theta(t)$. The above equations can be transformed to the Euler-Lagrange equations as

$$\frac{d^2 \theta_h}{dt^2} + \sum_{q=1}^n \sum_{z=1}^n \left[\frac{1}{2} \sum_{l=1}^n g^{hl} \left(\frac{\partial g_{ql}}{\partial \theta_z} + \frac{\partial g_{zl}}{\partial \theta_q} - \frac{\partial g_{qz}}{\partial \theta_l} \right) \right] \frac{d\theta_q}{dt} \frac{d\theta_z}{dt} = 0, \forall h, l, q, z \in \{1, 2, \dots, N\} \quad (14)$$

where the subscripts have the same meanings as above, while the superscripts index the elements in the inverse matrix of $G(\theta)$ in (11).

2.2 Construction of Geometry Model

How to choose support is important to the recovery algorithms. The common method is described as follows. Firstly, the correlation between each column of measurement matrix and residual is calculated. Secondly, the index (or indexes) which has (have) the highest correlation is (are) added to the support. So the correlation is the indicator of determining support.

In this paper, we improve the traditional indicator by a reward term. If the addition of an index into the support leads to a decrease of D_F between the estimated distribution and the known distribution, the reward will be set higher. Define the reward term by:

$$\text{reward}(\tilde{A}) = \frac{1}{\lambda(\theta^1, S) + D_F(\theta^1, \theta^2)} \quad (15)$$

where $\lambda(\theta^1, S)$ is a constant which offsets D_F 's influence on the indicator. It depends on the parameters θ^1 and the sparsity S . The details will be discussed in section 3.

θ^2 is determined with the support \tilde{A} generated in each iteration. In an iteration, the support set $\tilde{A} = \{k^1, k^2, \dots, k^j\}$ has j elements. We obtain θ^2 by the Maximum Likelihood Estimate (MLE) [62].

$$\theta^2 = \arg \max_{\theta} \sum_{i=1}^j p(k^i | \theta) \quad (16)$$

The reward term is used to modify the evaluation of the support in each iteration. An indicator is multiplied by its reward term. It is adaptive according to the different estimated support set in each iteration. In another word, it is an adaptive strategy based on the current estimated signal in each iteration of the recovery. The improved algorithm for greedy algorithm is shown in Algorithm 1. The improved algorithm for convex optimization algorithm is shown in Algorithm 2.

Algorithm 1:

Inputs:
Measurement matrix: Φ
Measurement vector: y
Max step size: $iter_{max}, res_{min}, thres_{min}$

Initialization:
 $j = 0$; **(The iteration index)**
 $r^0 = y$; **(The measurement residual)**
 $\tilde{x}, x_{A^j}^j \in R^N$; **(The estimated signal)**
 $\Phi_{A^j} = []$; **(The matrix formed by the columns ϕ_k of Φ where $k \in A^j$)**
 $A^j = \emptyset$; **(The stage index)**
 $done = 0$; **(The set of nonzero entries)**

do
{
 $j = j + 1$;

for $1 \leq k^j \leq N$ **and** $k^j \notin A^{j-1}$
 $\tilde{A}^j = \{A^{j-1} \cup k^j\}$;

 $\theta^j = \arg \max_{\theta} \sum_{i=1}^j p(k^i | \theta)$;

 $reward(\tilde{A}^j) = \frac{1}{\lambda + d_F(\theta, \theta^j)}$;

end
 $k^j = \arg \max_{1 \leq k^j \leq N, k^j \notin A^{j-1}} func(r^{j-1}, \phi_{k^j})$; **(*)**
 $A^j = A^{j-1} \cup \{k^j\}$;

 $\Phi_{A^j} = [\phi_k, k \in A^{j-1}; \phi_{k^j}]$;

 $x_{A^j}^j = \arg \min \|y - \Phi_{A^j} x\|_{l_2}$;

 $r^j = y - \Phi_{A^j} x_{A^j}^j$;

if ($j > iter_{max}$ **or** $\|r^j\|_{l_2} < res_{min}$) **{**
 $\tilde{x}_i = \begin{cases} x_i^j, & i = k^j \\ 0, & i \neq k^j \end{cases}$;

 $done = 0$;

}
while ($done == 0$)

Outputs: \tilde{x}, A ;

Algorithm 2:**Inputs:****Measurement matrix:** Φ **Measurement vector:** y **Max step size:** $iter_{max}, res_{min}, thres_{min}$ **Initialization:**

$j = 0$; **(The iteration index)**
 $r^0 = y$; **(The measurement residual)**
 $\mu^0 = 0$; **(The LARS estimate)**
 $A^0 = \emptyset$; **(The active set)**
 $\Phi_{A^0} = []$; **(The matrix formed by the columns ϕ_k of Φ where $k \in A^j$)**
 $done = 0$; **(The stage index)**

do{ $j = j + 1$; $r^j = y - \mu^{j-1}$;**for** $1 \leq k^j \leq N$ and $k^j \notin A^{j-1}$ $\tilde{A}^j = \{A^{j-1} \cup k^j\}$; $\theta^* = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^j p(k^i | \theta)$; $reward(\tilde{A}^j) = \frac{1}{\lambda + d_F(\theta, \theta^*)}$;**end** $k^j = \underset{1 \leq k' \leq N, k' \notin A^{j-1}}{\operatorname{argmax}} func(r^{j-1}, \phi_{k'})$; (*) $\hat{C} = func(r^{j-1}, \phi_{k^j})$; $A^j = A^{j-1} \cup \{k^j\}$; $\Phi_{A^j} = [sign(\langle \phi_k, r^{j-1} \rangle) \times \phi_k, k \in A^{j-1}, \phi_{k^j}]$; $G = \Phi_{A^j}^T \Phi_{A^j}$; $B = (I_{A^j}^T G^{-1} I_{A^j})^{-\frac{1}{2}}$; $w = B G^{-1} I_{A^j}$;**if** ($j = 1$) { $u = \phi_{k^j}$; } **else** { $u = \Phi_{A^j} w$; } $a = \Phi^T u$; $\gamma = \min_{k^j \in A^j} \left\{ \frac{\hat{C} - func(r^{j-1}, \phi_{k^j})}{B - a_{k^j}}, \frac{\hat{C} + func(r^{j-1}, \phi_{k^j})}{B + a_{k^j}} \right\}$; $\mu^j = \mu^{j-1} + \gamma u$; $\tilde{x}^j = \begin{cases} \tilde{x}_i^{j-1} + \gamma w_i, & i \in A^j \\ \tilde{x}_i^{j-1}, & i \notin A^j \end{cases}$;**if** ($j > iter_{max}$ or $\|r^j\|_{l_2} < res_{min}$) { $done = 0$; }**while** ($done == 0$)**Outputs:** \tilde{x} ;

$fun()$ in (*) is defined by

$$fun(r^{j-1}, \phi_i) = \begin{cases} \langle W r^{j-1}, \phi_i \rangle, & j \leq thres_{\min} \\ \langle W r^{j-1}, \phi_i \rangle \cdot reward(A^{j-1} \cup k^i), & j > thres_{\min} \end{cases} \quad (17)$$

where $\langle \rangle$ represents inner product in Euclidean space, W is a positive definite matrix and $thres_{\min}$ is a threshold. For non-weighted algorithms, W is a unit matrix and for algorithms with weights, W is a diagonal.

In the algorithms discussed above, the computing power mainly focuses on the computation of (15) to (17). In the j th iteration, the computation described above will be implemented for $N - j + 1$ candidates. However, the computation doesn't involve any travel or propagation across a model with a complex architecture. Thus, we infer that the computational complexity of the proposed algorithm is lower than that of the model-based algorithms introduced in section 1.

3. Experimental Results and Discussion

Generally, the analysis of various applications allows the extraction of prior knowledge about the specific distribution over the supports of the signal's sparse representation. In this section, we take a series of experiments with the simulated signals and imaging signals.

3.1 Simulation

We choose two PDF randomly in [Table 1](#) as the known index distributions to generate 1000 simulated signals following each of them, respectively. We suppose that the latitude of each nonzero entry follows a standard normal distribution. And N is set to 300. We set $iter_{\max} = 8 \times M$. The res_{\min} is 10^{-5} , while the $thres_{\min}$ is 30 (it is big enough for static). λ is determined based on different sparsity for each PDF respectively, as shown in [Table 1](#).

Table 1. Settings for PDF

PDF	S	λ
$f_1(x) = e^{-\frac{(x-120)^2}{1800}} / (30\sqrt{2\pi})$	45	0.5 ^a
	60	0.3
	75	0.28
$f_2(x) = 0.008e^{-0.008x}$	45	0.23
	60	0.3
	75	0.17

^aFor simplicity, λ is set to the expectation of D_F over the samples subtracted by 1.

The signal is acquired from noisy measurements with $SNR_{mes}=30\text{dB}$. Accurate recovery is declared when

$$\|x - \tilde{x}\|_{l_2} / \|x\|_{l_2} \leq 10^{-1} \quad (18)$$

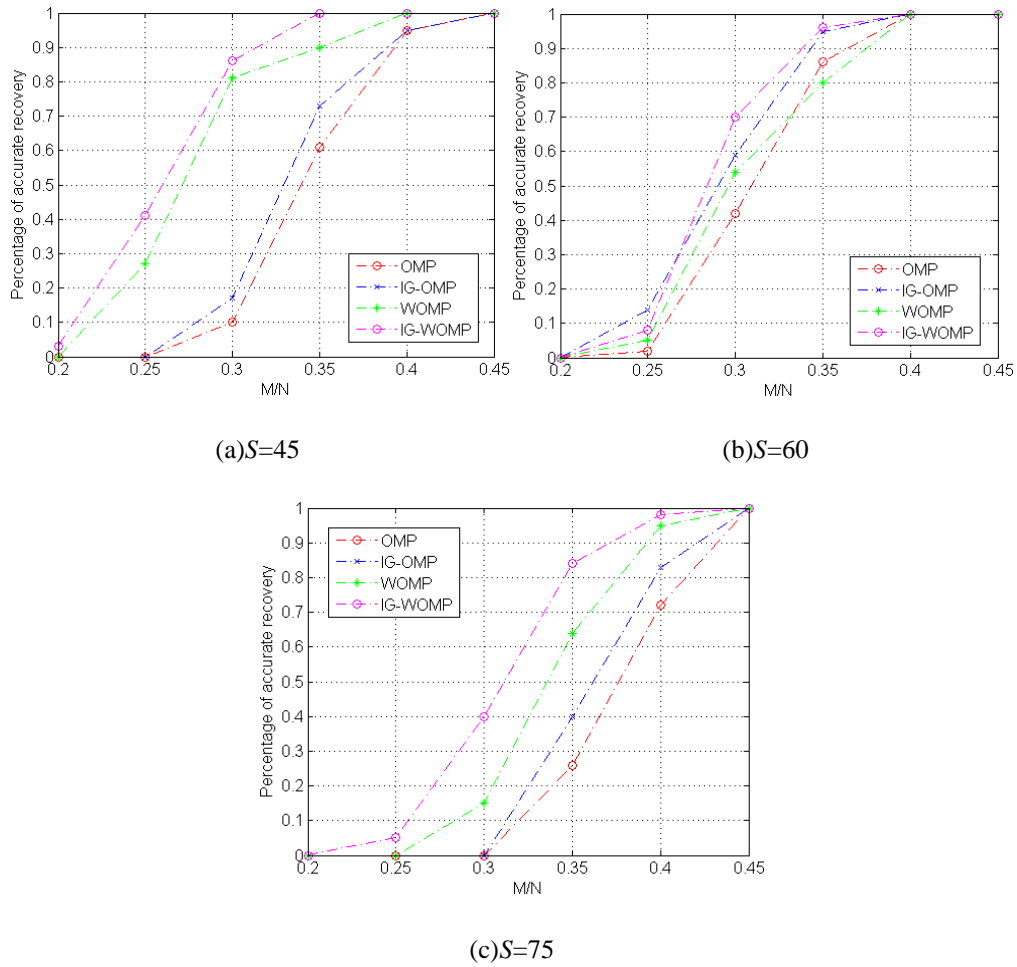


Fig. 1. Comparison among OMP, IG-OMP, WOMP and IG-WOMP with the distribution described by f_1

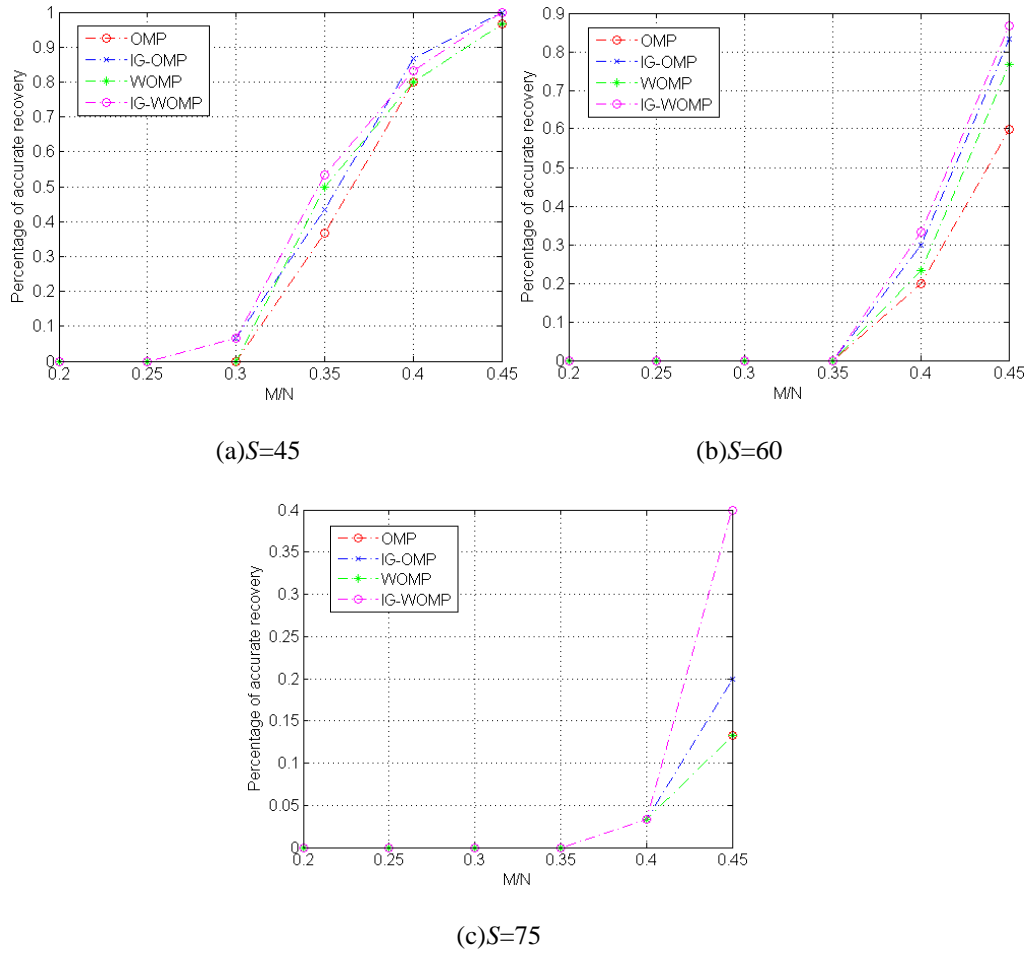
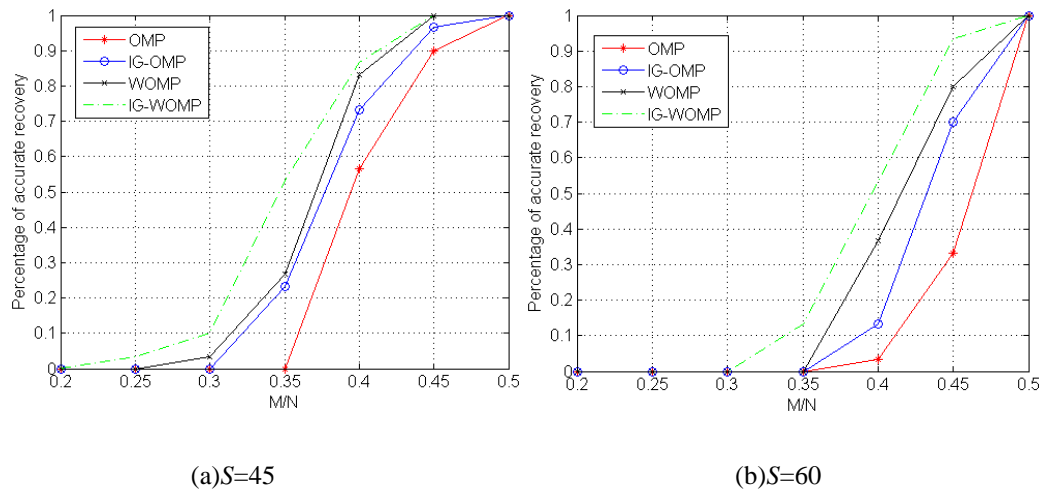
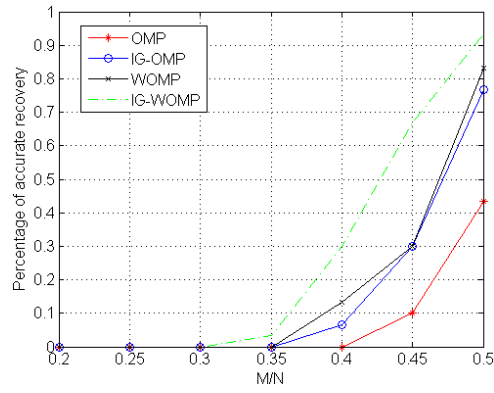


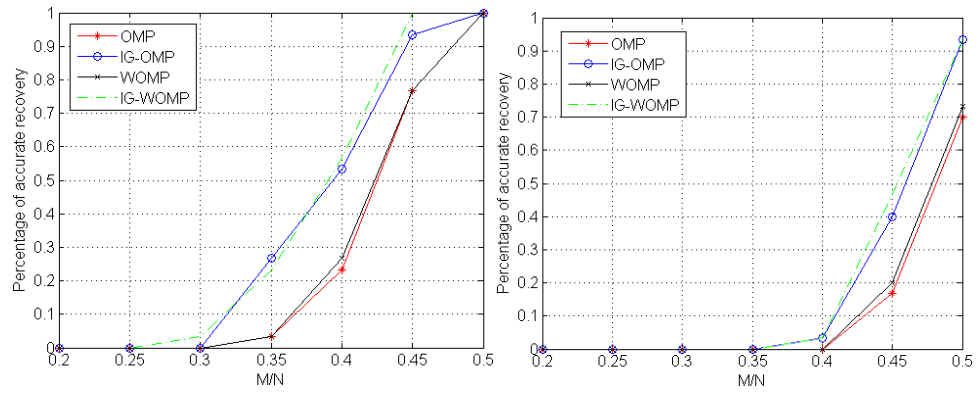
Fig. 2. Comparison among OMP, IG-OMP, WOMP and IG-WOMP with the distribution described by f_2





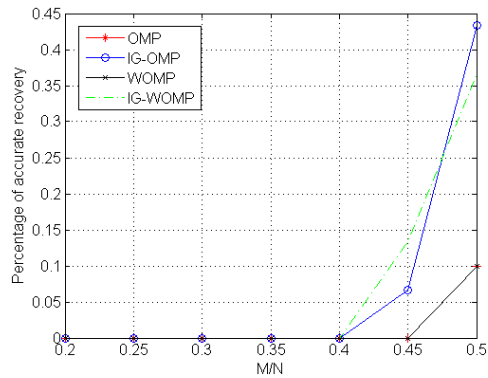
(c)S=75

Fig. 3. Comparison among LARS, IG-LARS, WLARS and IG-WLARS with the distribution described by f_1



(a)S=45

(b)S=60



(c)S=75

Fig. 4. Comparison among LARS, IG-LARS, WLARS and IG-WLARS with the distribution described by f_2

During the experiment, M/N is increasing gradually. Meanwhile, we compute the accurate recovery rate. Firstly, the proposed method is used to improve OMP and WOMP, respectively. The improved algorithms are called IG-OMP and IG-WOMP, respectively. We illustrate the results in Fig. 1 and Fig. 2. Secondly, proposed method is used to improve LARS and WLARS, respectively. The improved algorithms are called IG-LARS and IG-WLARS, respectively. We illustrate the results in Fig. 3 and Fig. 4.

To validate the superiority of the proposed algorithm in the computational complexity, we compare the performance of an algorithm based on information geometry with a typical model-based algorithm for CS (i.e., neural networks for CS introduced in section 1) in both of the average running time per iteration and the accurate recovery rate. The experimental settings are as the same as above. For each algorithm, we take the average over all conditions. Naturally, the improved algorithms are more complex than the original algorithms. In addition, OMP has a lower time complexity than LARS. Similarly, IG-WOMP has a lower time complexity than IG-WLARS. Hence, IG-WLARS is the most complex one among the algorithms (OMP, IG-OMP, WOMP, IG-WOMP, LARS, IG-LARS, WLARS and IG-WLARS). So the running time of the lower complex algorithms are meaningless in the validation. Then, we choose the result of the most complex one (i.e., IG-WLARS) for the comparison in Table 2. For simplicity, Neural Networks for CS is abbreviated as NNCS.

Table 2. Results

Algorithm	Percentage(%)	Time(second)
IG-WLARS	36.24	1.07
NNCS	39.05	5.51

3.2 Application Examples

Imaging is a typical signal whose support follows a specific distribution. Here, x is an imaging signal in DCT domain. The image is blocked, so that N is 32×32 . And the M is set to 200.

In Table 3 and Table 4, the results are shown numerically. Moreover, the results are shown intuitively in Fig. 5 and Fig. 6.

Table 3. Results

Image	Algorithm	PSNR(dB) ^a
fruits	OMP	25.0959
	IG-OMP	27.1202
	WOMP	25.6369
	IG-WOMP	29.4092
boat	OMP	24.2169
	IG-OMP	26.8963

camera man	WOMP	25.8620
	IG-WOMP	28.6659
	OMP	22.9748
	IG-OMP	25.4498
	WOMP	26.5065
	IG-WOMP	28.4797

^a The definition of the Peak-Signal-to-Noise Ratio (PSNR) is given in [63].

Table 4. Results

Image	Algorithm	PSNR(dB)
fruits	LARS	28.8631
	IG-LARS	29.7615
	WLARS	28.9837
	IG-WLARS	30.0083
boat	LARS	27.4469
	IG-LARS	29.3894
	WLARS	29.0135
	IG-WLARS	30.0485
camera man	LARS	25.8389
	IG-LARS	27.3939
	WLARS	29.5036
	IG-WLARS	30.3571



(a)



(b)



(c)



(d)



(e)



(a)



(b)



(c)



(d)



(e)



(a)



(b)



(c)



(d)



(e)

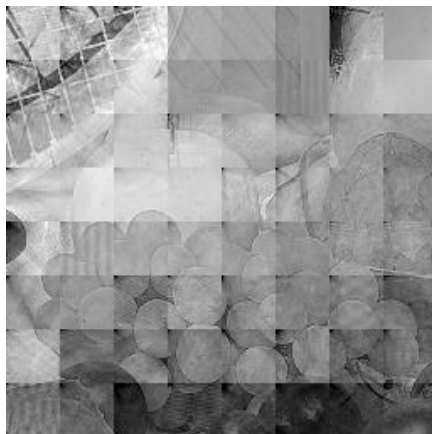
Fig. 5. (a) Original image (b) IG-WOMP (c) WOMP (d) IG-OMP (e) OMP



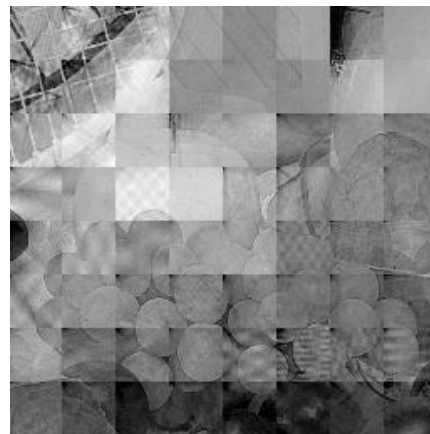
(a)



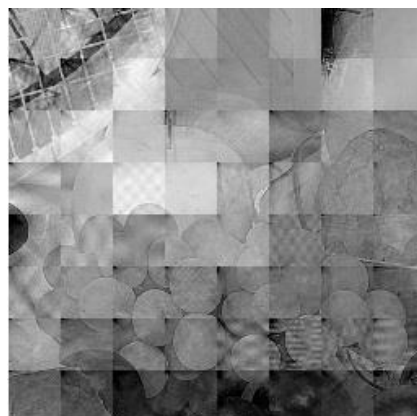
(b)



(c)



(d)



(e)



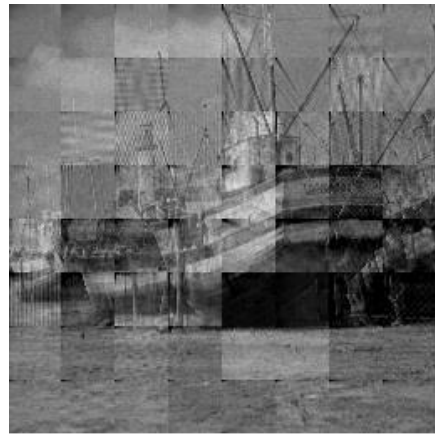
(a)



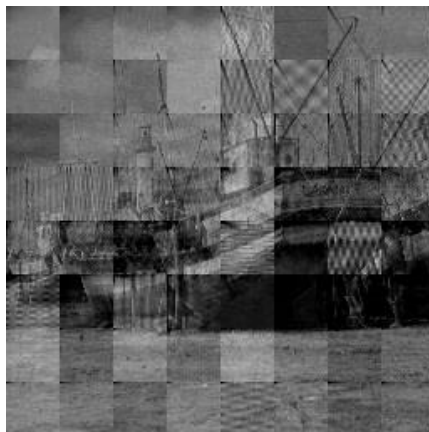
(b)



(c)



(d)



(e)

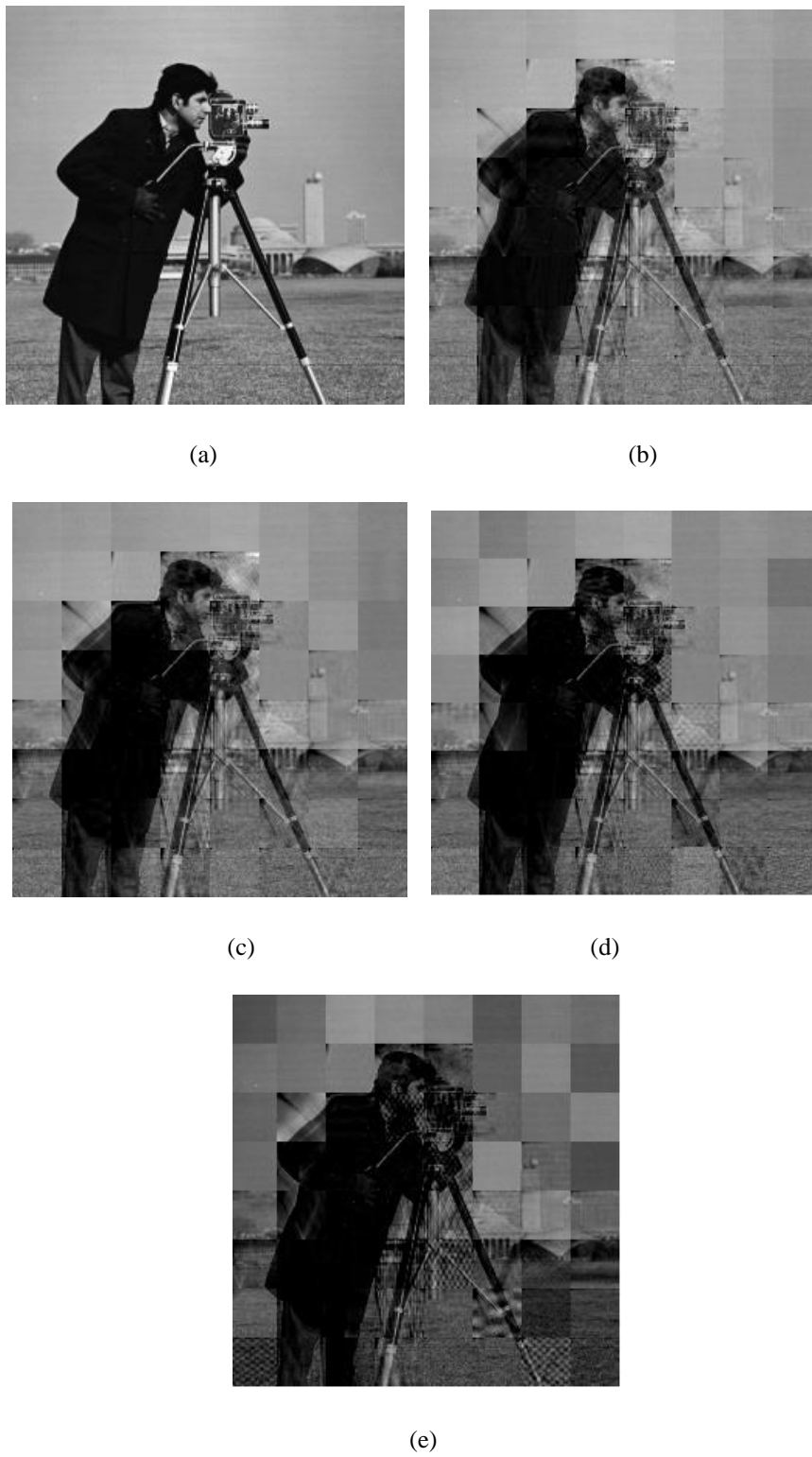


Fig. 6. (a) Original image (b) IG-WLARS (c) WLARS (d) IG-LARS (e) LARS

To validate the superiority of the proposed algorithm in the computational complexity, we compare the performance of the most complex algorithm based on information geometry in the paper (i.e., IG-WLARS) with a typical model based algorithm for CS (i.e., NNCS) in both of the average running time and the recovery qualities of the images (i.e., the average PSNR over all images). And we list the results in [Table 5](#).

Table 5. Results

Algorithm	PSNR(dB)	Time(second)
IG-WLARS	30.138	27.13
NNCS	29.8146	70.19

3.3 Discussion

Firstly, we discuss [Fig. 1](#), [Fig. 2](#), [Fig. 3](#), and [Fig. 4](#). S -sparse means x is with S nonzero entries. Under different sparsities, algorithms improved by information geometry outperform original algorithms constantly. The results of two types of decoders are shown: original decoders (OMP, WOMP, LARS, and WLARS) and improved decoders (IG-OMP, IG-WOMP, IG-LARS, and IG-WLARS). Original decoders have worse performance.

Secondly, we discuss the results in [Fig. 5](#), [Fig. 6](#), [Table 3](#), and [Table 4](#). As it is shown in [Fig. 5](#), [Fig. 6](#), [Table 3](#), and [Table 4](#), the recovery is applied to different standard test images (fruits, boat, camera man). In [Table 3](#) and [Table 4](#), a higher PSNR indicates that the reconstruction is of higher quality. In [Fig. 5](#) and [Fig. 6](#), the reconstructions of images show that algorithms improved by the proposed method outperform the original algorithms in real data.

Thirdly, we discuss the results in [Table 2](#) and [Table 5](#). In [Table 2](#), the running time of IG-WLARS is lower than NNCS, while the values of accurate recovery rate are close. In [Table 5](#), the running time of IG-WLARS is lower than NNCS, while the recovery quality of the IG-WLARS is better. The results of running time of these algorithms contrast sharply with each other, while the values of the recovery quality are very close. The results support the conclusion that the proposed algorithm has a lower computational complexity, while keeping a commensurate recovery accuracy.

4. Conclusion

With the theory of information geometry, we propose an adaptive strategy based on the current estimated signal in each iteration of the recovery. And we improve the performance of recovery algorithms through the adaptive strategy for exploiting the prior knowledge about the index distribution. Simulations are presented to validate the results. In the end, we also show the application of the model in the image.

Acknowledgement

This research was funded by the Beijing Science and Technology Planning Program of China (Z171100004717001), Beijing Natural Science Foundation (4172002), Natural Science

Foundation of China (61701009), Conventional Project of Promoting of the Connotation Developing of Colleges and Universities (5112011030), and Natural Science Foundation of China (11772063).

References

- [1] J. Zhou, J. Ai, Z. Wang, S. Chen, and Q. Wei, "Discovering attractive segments in the user-generated video streams," *Lecture Notes in Computer Science*, vol. 11642, pp. 236-250, 2019. [Article \(CrossRef Link\)](#)
- [2] X. Chai, Z. Xiaoyu, G. Zhihua, H. Daojun, and Y. Chen, "An image encryption algorithm based on chaotic system and compressive sensing," *Signal Processing*, vol. 148, pp. 124-144, July 2018. [Article \(CrossRef Link\)](#)
- [3] Z. Zha, X. Zhang, Q. Wang, L. Tang, and X. Liu, "Group-based sparse representation for image compressive sensing reconstruction with non-convex regularization," *Neurocomputing*, vol. 296, pp. 55-63, June 2018. [Article \(CrossRef Link\)](#)
- [4] G. Han and B. Lin, "Optimal sampling and reconstruction of undersampled atomic force microscope images using compressive sensing," *Ultramicroscopy*, vol. 189, pp. 85-94, June 2018. [Article \(CrossRef Link\)](#)
- [5] J. Wang, Q. Wang, and Yu. Hu, "Image Encryption Using Compressive Sensing and Detour Cylindrical Diffraction," *IEEE Photonics Journal*, vol. 10, no. 3, pp.1-14, June 2018. [Article \(CrossRef Link\)](#)
- [6] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Trans signal processing magazine*, vol. 25, no. 2, pp. 21-30, Mar. 2008. [Article \(CrossRef Link\)](#)
- [7] D. L. Donoho, "Compressed sensing," *IEEE Trans Inform Theory*, vol. 52, no. 4, pp. 1289-1306, May 2006. [Article \(CrossRef Link\)](#)
- [8] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions Inform. Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006. [Article \(CrossRef Link\)](#)
- [9] S. Muthukrishnan, "Data Streams: Algorithms and Applications," *Foundations & Trends in Theoretical Computer Science*, p. 135, 2005. [Article \(CrossRef Link\)](#)
- [10] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans on information theory*, vol. 51, no. 12, pp. 4203-4215, Dec. 2005. [Article \(CrossRef Link\)](#)
- [11] S. G. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397-3415, Dec. 1993. [Article \(CrossRef Link\)](#)
- [12] M. Figueiredo, D. Robert, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586-597, Dec. 2007. [Article \(CrossRef Link\)](#)
- [13] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *Society for Industrial and Applied Mathematics*, vol. 43, no. 1, pp. 129-159, 2001. [Article \(CrossRef Link\)](#)
- [14] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413-1457, Dec. 2004. [Article \(CrossRef Link\)](#)
- [15] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, "Bayesian pursuit algorithm for sparse representation," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009. [Article \(CrossRef Link\)](#)
- [16] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian Compressive Sensing Via Belief Propagation," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 269-280, Jan. 2010. [Article \(CrossRef Link\)](#)
- [17] M. W. Seeger, "Bayesian inference and optimal design for the sparse linear model," *Journal of Machine Learning Research*, vol. 9, pp. 759-813, May 2008. [Article \(CrossRef Link\)](#)

- [18] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans on information theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007. [Article \(CrossRef Link\)](#)
- [19] D. L. Donoho, Y. Tsaig, and I. Drori, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *Statistics Technical Report Aerospace Corp.*, vol. 58, no. 2, pp. 1094-1211, Feb. 2012. [Article \(CrossRef Link\)](#)
- [20] D. Needell and R. Vershynin, "Uniform Uncertainty Principle and Signal Recovery via Regularized Orthogonal Matching Pursuit," *Foundation of Computational Mathematics*, vol. 9, pp. 317-334, Jan. 2009. [Article \(CrossRef Link\)](#)
- [21] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301-321, May 2009. [Article \(CrossRef Link\)](#)
- [22] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, pp. 265-274, Feb. 2009. [Article \(CrossRef Link\)](#)
- [23] R. Garg and R. Khandeka, "Gradient descent with sparsification: an iterative algorithm for sparse recovery with restricted isometry property," in *Proc. of the 26th International Conference on Machine Learning*, pp. 337-344, 2009. [Article \(CrossRef Link\)](#)
- [24] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 60, pp. 6202-6216, Dec. 2012. [Article \(CrossRef Link\)](#)
- [25] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, vol. 73, no. 3, pp. 273-282, 2011. [Article \(CrossRef Link\)](#)
- [26] W. Jiao, L. Fang, and J. Licheng, "Reconstruction of images from compressive sensing based on the stagewise fast LASSO," in *Proc. of SPIE*, vol. 7498, 2009. [Article \(CrossRef Link\)](#)
- [27] L. Lian, A. Liu, and V. K. N. Lau, "Weighted LASSO for Sparse Recovery with Statistical Prior Support Information," *IEEE Transactions on Signal Processing*, vol. 66, no. 6, pp. 1607-1618, 2018. [Article \(CrossRef Link\)](#)
- [28] C. Y. Yau and T. S. Hui, "LARS-type algorithm for group lasso," *Statistics and Computing*, vol. 27, pp. 1041-1048, 2017. [Article \(CrossRef Link\)](#)
- [29] B. Radley EFron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least Angle Regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407-499, 2004. [Article \(CrossRef Link\)](#)
- [30] S. S. Keerthi and S. Shevade, "A Fast Tracking Algorithm for Generalized LARS/LASSO," *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1826-1830, 2007. [Article \(CrossRef Link\)](#)
- [31] M. A. Khajehnejad, W. Xu, A. Avestimehr, and B. Hassibi, "Analyzing Weighted l1 Minimization for Sparse Recovery with Nonuniform Sparse Models," *IEEE Transactions on Signal Processing*, vol. 59, pp.1985-2001, Jan. 2011. [Article \(CrossRef Link\)](#)
- [32] Z. Wang, K. Chen, M. Zhang, P. He, Y. Wang, P. Zhu, and Y. Yang. "Multi-scale aggregation network for temporal action proposals," *Pattern Recognition Letters*, vol. 122, no. 1, pp. 60-65, 2019. [Article \(CrossRef Link\)](#)
- [33] W. Lu and N. Vaswani, "Modified compressive sensing for real-time dynamic MR imaging," *Proc. IEEE Int. Conf. Image Process*, 2009. [Article \(CrossRef Link\)](#)
- [34] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Transactions Image Process.*, vol. 19, no. 1, pp. 53-63, Jan. 2010. [Article \(CrossRef Link\)](#)
- [35] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions Inf. Theory*, vol. 56, no. 4, pp. 1982-2001, 2010. [Article \(CrossRef Link\)](#)
- [36] S. Som and P. Schniter, "Compressive imaging using approximate message passing and a Markov-tree prior," *IEEE Transactions Signal Process*, vol. 60, no. 7, pp. 3439-3448, 2012. [Article \(CrossRef Link\)](#)
- [37] D. Merhej, C. Diab, M. Khalil, and R. Prost, "Embedding prior knowledge within compressed sensing by neural networks," *IEEE Transactions Neural Netw.*, vol. 22, no. 10, pp. 1638-1649, Oct. 2011. [Article \(CrossRef Link\)](#)
- [38] H. Q. Bui, C. N. H. La, and M. N. Da, "A fast tree-based algorithm for Compressed Sensing with sparse-tree prior," *Signal Processing*, vol. 108, pp. 628-641, 2015. [Article \(CrossRef Link\)](#)

- [39] Z. Zhang, X. Ma, and Y. Yang, "Bounds on the number of hidden neurons in three-layer binary neural networks," *Neural Networks*, vol. 16, no. 7, pp. 995-1002, 2003. [Article \(CrossRef Link\)](#)
- [40] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Transactions Signal Process.*, vol. 58, no. 9, pp. 4595-4607, Sep. 2010. [Article \(CrossRef Link\)](#)
- [41] N. Vaswani, "Kalman filtered compressed sensing," in *Proc. of IEEE International Conference Image Process*, 2008. [Article \(CrossRef Link\)](#)
- [42] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi, "Analyzing Weighted Minimization for Sparse Recovery With Nonuniform Sparse Models," *IEEE Transactions Signal Process*, vol. 59, no. 5, pp. 1985-2001, May 2011. [Article \(CrossRef Link\)](#)
- [43] C. La and M. Do, "Tree-based Orthogonal Matching Pursuit algorithm for signal reconstruction," in *Proc. of IEEE International Conference on Image Processing*, pp. 1277-1280, May 2006. [Article \(CrossRef Link\)](#)
- [44] D. Mateo, J. Mauricio, R. Felipe, and V. Mauricio, "Compressed sensing of data with a known distribution," *Applied and Computational Harmonic Analysis*, vol. 45, no. 3, pp. 486-504, Nov. 2018. [Article \(CrossRef Link\)](#)
- [45] D. Escoda, L. Granai, and P. Vandergheynst, "On the use of a priori information for sparse signal approximations," *IEEE Transactions Signal Processing*, vol. 54, no. 9, pp. 3468-3482, Sep. 2006. [Article \(CrossRef Link\)](#)
- [46] N. Ay, J. Jost, H. V. Lê, and L. Schwachhöfer. *Information Geometry*, Springer, 2017.
- [47] R.E. Kass and P.W. Vos, *Geometrical Foundations of Asymptotic Inference*, New York, USA: Wiley-Interscience, 1997.
- [48] S. Amari, "Information geometry of the EM and em algorithms for neural networks," *Neural Networks*, vol. 8, no. 9, pp. 1379-1408, Dec. 1995. [Article \(CrossRef Link\)](#)
- [49] D. Meng and R. Liu, "Information geometry-Geometric methods for Computational Neuroscience," *Acta Biophysica Sinica*, vol. 15, pp. 243-248, May 1999.
- [50] M. Hu, Y. Yang, F. Shen, N. Xie, R. Hong, and H. T. Shen, "Collective Reconstructive Embeddings for Cross-Modal Hashing," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2770-2784, 2019. [Article \(CrossRef Link\)](#)
- [51] Z. Yang and J. Laaksonen, "Principal whitened gradient for information geometry," *Neural Networks*, vol. 21, pp. 232-240, Nov. 2008. [Article \(CrossRef Link\)](#)
- [52] S. Amari, K. Kurata, and H. Nagaoka, "Information geometry of Boltzmann machines," *IEEE Transactions on Neural Networks*, vol. 3, no. 2, pp. 260-271, Jan. 1992. [Article \(CrossRef Link\)](#)
- [53] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, pp. 251-276, Sep. 1998. [Article \(CrossRef Link\)](#)
- [54] S. Amari, "Fisher information under restriction of Shannon information in multi-terminal situations," *Annals of the Institute of Statistical Mathematics*, vol. 41, pp. 623-648, Jan. 1989. [Article \(CrossRef Link\)](#)
- [55] L. Campbell, "The relation between information theory and the differential geometry approach to statistics," *Information Sciences*, vol. 35, no. 3, pp. 199-210, Feb. 1985. [Article \(CrossRef Link\)](#)
- [56] S. Amari, "Information geometry on hierarchy of probability distributions," *IEEE Transactions on Information Theory*, vol. 47, no. 5, pp. 1701-1711, Mar. 2001. [Article \(CrossRef Link\)](#)
- [57] M. Wang, C. B. Xiao, and Z. H. Ning, "Neural Networks for Compressed Sensing Based on Information Geometry," *Circuits Systems and Signal Process*, vol. 38, no. 2, pp. 569-589, Feb. 2019. [Article \(CrossRef Link\)](#)
- [58] M. Hu, Y. Yang, F. Shen, N. Xie, R. Hong, and H. T. Shen, "Collective Reconstructive Embeddings for Cross-Modal Hashing," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2770-2784, 2019. [Article \(CrossRef Link\)](#)
- [59] S. He, B. Wang, Z. Wang, Y. Yang, F. Shen, Z. Huang, and H. T. Shen, "Bidirectional Discrete Matrix Factorization Hashing for Image Search," *IEEE Transactions on Cybernetics*, vol. 5, no. 9, pp. 4157-4168, 2020. [Article \(CrossRef Link\)](#)
- [60] B. R. Frieden, *Science from Fisher Information: A Unification*, Cambridge Univ. Press, 2004.

- [61] F. Nielsen and V. Garcia, "Statistical exponential families: A digest with flash cards," arXiv.org:0911.4863, 2009. [Article \(CrossRef Link\)](#)
- [62] M. Menendez, D. Morales, L. Pardo, and M. Salicr j, "Statistical tests based on geodesic distances," *Applied Mathematics Letters*, vol. 8, no. 1, pp. 65-69, Jun. 1995. [Article \(CrossRef Link\)](#)
- [63] H. Quan and G. Mohammed, "The accuracy of PSNR in predicting video quality for different video scenes and frame rates," *Telecommunication Systems*, vol. 49, pp. 35-48, Feb. 2012. [Article \(CrossRef Link\)](#)



Meng Wang holds a lecturer position in School of Applied Science at Beijing Information Science and Technology University. She received a Ph.D. degree in Computer Science from the Beijing University of Technology in 2020. Her research interests are in the areas of signal processing and artificial intelligence.



Zhen Hu Ning holds a lecturer position in the Faculty of Information Technology at the Beijing University of Technology. He received a Ph.D. degree in Computer Science from the Beijing University of Technology in 2016. His research interests are in the areas of information safety and artificial intelligence.



Jing Yu holds a lecturer position in the Faculty of Information Technology at the Beijing University of Technology. She received a Ph.D. degree in Information and Communication Engineering from the Tsinghua University in 2011. From 2011 to 2014, she took a job of postdoctoral scientific research pursues in workstation at the Tsinghua University. Her research interests are in the areas of image processing and artificial intelligence.



Chuangbai Xiao received a Ph.D. degree from the Tsinghua University in 1995. Since 2001, he has been teaching and researching in the Faculty of Information Technology at the Beijing University of Technology, where he is currently a Professor. He has been an author or co-author of more than 100 papers in peer-reviewed journals, conferences, or workshops.