EXACT SPARSE SUPER-RESOLUTION VIA MODEL AGGREGATION

Hongqing Yu and Heng Qiao

University of Michigan - Shanghai Jiao Tong University Joint Institute,
Shanghai Jiao Tong University, Shanghai, China
E-mail: YuHongQing@sjtu.edu.cn, heng.qiao@sjtu.edu.cn

ABSTRACT

This paper studies the problem of discrete super-resolution. Existing stability guarantees rely on the fact that certain separation conditions are satisfied by the true support. However, such structural conditions have not been exploited in the corresponding algorithmic designs. This paper proposes a novel Bayesian approach based on the model aggregation idea that can generate an exact sparse estimate, and maintain the required structures of the support. The proposed method is implemented within the MCMC framework and empirically provides better support recovery than available algorithms.

Index Terms **—** Exact support recovery, discrete super resolution, model aggregation, MCMC algorithm

1. INTRODUCTION

The problem of super-resolution finds important applications in a wide range of areas including imaging [1–3], astronomy [4] and radar [5], to name a few. The fundamental goal is to estimate the signal of interest from its noisy low-frequency measurements. In this paper, we adopt the basic assumption that the target signal is sparse and aim to exactly recover the support. There are two lines of work differing on whether the support lies on a pre-defined grid. We take the on grid model in [6–9] instead of the gridless one which always induces spurious spikes to the best of authors' knowledge [10–14].

The discrete super-resolution problem belongs to the classic underdetermined linear regression [15] but with a deterministic Fourier-type measurement matrix. In statistical literature, recovering the support also refers to the problem of variable selection or model selection [16]. A large number of methods have been proposed in relevant contexts from both frequentist and Bayesian viewpoints. Existing frequentist methods impose deviation constraints like basis pursuit [17], or include structural penalizations like LASSO [18], or apply the maximum likelihood principle like SPICE [19].

The Bayesian approaches, on the other hand, can be roughly classified into two categories [20]: spike-and-slab priors (two-groups model) and global-local shrinkage priors. The family of spike-and-slab priors places a discrete mixture of a point mass at zero and an absolutely continuous density on each coordinate (grid point). The global-local shrinkage priors instead place absolutely continuous shrinkage priors on the whole parameter vector to promote sparsity. The successful Bayesian methods such as sparse Bayesian learning [21] and approximate message passing [22] fall into the second category by simultaneously shrinking all the coordinates.

From either perspective, general stability analyses require certain properties (e.g. RIP, coherence) of the measurement matrix [15, 23, 24], which do not hold for the deterministic Fourier-type matrix in general [25]. Available performance analyses of super-resolution instead assume the true support satisfys certain separation conditions [7, 8, 13, 14, 30]. However, the separation conditions are not considered in the algorithmic designs. Thus, the obtained estimates generally do not satisfy the separation condition without post-processing [17].

To bridge this gap between theoretical and numerical sides of the super-resolution research, we propose to apply the idea of model aggregation, and implement a Monte-Carlo Markov Chain (MCMC) algorithm for exact sparse recovery. In particular, we adopt the methodology in [26–29] that all possible supports are assigned to proper priors and the final estimate is an aggregation of primitive estimates with respect to each support. Note that both the spike-and-slab and global-local shrinkage priors will not work here as they lack the ability to differentiate the supports and accommodate the desired separation constraint. Though the method is by nature combinatorial as it considers all individual supports, the proposed MCMC algorithm is empirically efficient and provides better support recovery compared to existing approaches.

2. PROBLEM FORMULATION

To set up the super-resolution problem, we adopt the following widely used model [7, 8, 30]. In particular, we acquire a noisy low-frequency sketch of a sparse vector $\mathbf{x}^\star \in \mathbb{C}^N$ as

$$\mathbf{y} = \mathbf{A}\mathbf{x}^\star + \mathbf{w}^\star \tag{1}$$

where the support $\mathbb{S}^\star$ of $\mathbf{x}^\star$ is of size s, and $\mathbf{w}^\star \in \mathbb{C}^M$ is the additive noise. The low-pass procedure is characterized by a partial DFT matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$ without normalization:

$$[\mathbf{A}]_{m,n} = e^{j2\pi\frac{mn}{N}}, \quad -f_c \leqslant m \leqslant f_c,\ 0 \leqslant n \leqslant N-1$$

where f_c denotes the cut-off frequency. Our goal is to obtain an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}^\star$ from the noisy measurement $\mathbf{y}$.

3. STRUCTURAL SUPER-RESOLUTION VIA MODEL AGGREGATION

3.1. A model aggregation setting for super-resolution

To fulfill a possible pedagogical purpose, we first state the basic setting of model aggregation in statistical inference literature [27]. Suppose we observe $\{Y_i, X_i\}_{i=1}^{M}$ such that

$$Y_i = f(X_i) + \xi_i, i = 1, \cdots, M$$

where $f : \mathcal{X} \rightarrow \mathbb{R}$ is an unknown function to be estimated and ξ_i are independent random variables. In the context of model aggregation, we try to approximate f by linearly combining the elements in a given functional dictionary $\mathcal{F} = \{f_1, \cdots, f_L\}$ as follows

$$f_{\boldsymbol{\lambda}}(X) \triangleq \sum_{l=1}^{L} \lambda_l f_l(X), \quad \forall X \in \mathcal{X}$$

To apply the aggregation idea to the super-resolution problem, we have to properly define the dictionary $\mathcal{F}$ as well as the weights $\boldsymbol{\lambda}$ to promote desired support structure.

An intuitive construction of the functional dictionary $\mathcal{F}$ for sparse super-resolution is to associate each candidate f_l with one possible support. To be precise, a sparse pattern is a binary vector $\mathbf{p} \in \mathcal{P} \triangleq \{0,1\}^N$ [28]. The sparsity of a particular pattern $\mathbf{p}$ is denoted by $|\mathbf{p}|$, and the cardinality of $\mathcal{P}$ is 2^N. For each sparse pattern $\mathbf{p}$, we have the freedom to define the candidate model f to reflect certain structural prior information [29, 31]. In this paper, we simply exploit the ordinary least squares (OLS) to construct f. In particular, for each sparsity patter $\mathbf{p}_l$, we define f_l as

$$f_l \triangleq \hat{\mathbf{z}}_l = \arg\min_{\mathbf{z} \in \mathcal{Z}_l} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 \tag{2}$$

The feasibility set is given by

$$\mathcal{Z}_l \triangleq \{\mathbf{z} \odot \mathbf{p}_l, \mathbf{z} \in \mathbb{C}^N\} \subset \mathbb{C}^N$$

where $\odot$ denotes the element-wise (Hadamard) product to enforce the sparse pattern. If the solution of (2) is not unique, we will take the one of least norm. We can impose other structures such as sign [8] and shape [32] constraints on $\mathcal{Z}_l$.

After constructing the dictionary $\mathcal{F}$, the next step is to determine the combination coefficients $\boldsymbol{\lambda}$ such that we can simultaneously promote the sparsity of $\hat{\mathbf{x}}$ and maintain the desired separation of the support.

3.2. Sparse exponential weighting of OLS estimates

One natural way to compute $\boldsymbol{\lambda}$ is to apply the OLS as [27]

$$\hat{\boldsymbol{\lambda}}^{LS}(\Lambda) = \arg\min_{\lambda \in \Lambda} \|\mathbf{y} - \mathbf{A} f_{\boldsymbol{\lambda}}\|_2$$

where the feasibility set Λ can take many forms depending on the applications. For example, we can set $\Lambda = \mathbb{R}^L$ or

$$\Lambda \triangleq \Lambda^L = \left\{ \boldsymbol{\lambda} \in \mathbb{R}^L : \lambda_l \geqslant 0, \sum_{l=1}^{L} \lambda_l = 1 \right\} \tag{3}$$

The major disadvantage of coefficients $\boldsymbol{\lambda}^{LS}$ is that we cannot guarantee the desired separation nor promote the sparsity of the aggregate estimate. We propose to use the following coefficients for exponential weighting of $\mathcal{F} = \{f_l\}_{l=1}^{L}$ [27–29]

$$\hat{\lambda}_l^{EW} = \frac{\exp(-M\hat{r}_l/\beta)\pi_l}{\sum_{k=1}^{L} \exp(-M\hat{r}_k/\beta)\pi_k} \tag{4}$$

where $\hat{r}_l = \frac{1}{M}\|\mathbf{y} - \mathbf{A}\hat{\mathbf{z}}_l\|_2^2$ is the residual corresponding to the candidate model f_l. β is a tuning parameter and $\{\pi_l\}_{l=1}^{L}$ is a customized prior on the set of primitive estimates $\mathcal{F}$.

The reason for using the above defined exponential weighting coefficients is made clear by noting that [27]

$$\hat{\boldsymbol{\lambda}}^{EW} = \arg\min_{\boldsymbol{\lambda} \in \Lambda^L} \left(\sum_{l=1}^{L} \lambda_l \hat{r}_l + \frac{\beta}{M} \sum_{k=1}^{L} \lambda_k \log \frac{\lambda_k}{\pi_k} \right) \tag{5}$$

where $\sum_{k=1}^{L} \lambda_k \log \frac{\lambda_k}{\pi_k}$ is the Kullback-Leibler divergence between $\boldsymbol{\lambda}$ and $\boldsymbol{\pi}$. Thus, the exponential weighting aims to balance the data fitting error and the fidelity to the priors.

Obviously, the prior $\boldsymbol{\pi}$ plays an important role in enforcing certain structures of the support. For instance, we can simply set $\pi_l = 0$ if the sparse pattern $\mathbf{p}_l$ does not satisfy the separation condition. However, besides the separation, we also want to promote the sparsity of the resulting estimate $\hat{\mathbf{x}}$.

One such prior $\boldsymbol{\pi}$ to promote sparsity is given by [27–29]

$$\pi_l = \begin{cases} \frac{1}{H}\left(\frac{|\mathbf{p}_l|}{2eL}\right)^{|\mathbf{p}_l|}, & \mathbf{p}_l \in \tilde{\mathcal{P}} \subseteq \mathcal{P} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $\tilde{\mathcal{P}}$ is a subset of $\mathcal{P}$ that satisfys desired structures (e.g, separation), and H is a normalizing factor to make $\boldsymbol{\pi}$ a valid prior on $\mathcal{F}$. As argued in [28], $H \leqslant 4$ with such construction.

Without losing generality, for now we assume $\tilde{\mathcal{P}} = \mathcal{P}$ and show how the construction (6) promotes sparsity. Plug it back into (5), we have

$$\begin{aligned} \hat{\boldsymbol{\lambda}}^{EW} &= \arg\min_{\boldsymbol{\lambda} \in \Lambda^L} \left(\sum_{l=1}^{L} \lambda_l \hat{r}_l + \frac{\beta}{M} \sum_{k=1}^{L} \lambda_k \log \frac{\lambda_k}{\pi_k} \right) \\ &= \arg\min_{\boldsymbol{\lambda} \in \Lambda^L} \left(\sum_{l=1}^{L} \lambda_l \hat{r}_l + \frac{\beta}{M} \sum_{k=1}^{L} \lambda_k \log \frac{\lambda_k}{\frac{1}{H}\left(\frac{|\mathbf{p}_k|}{2eL}\right)^{|\mathbf{p}_k|}} \right) \\ &= \arg\min_{\boldsymbol{\lambda} \in \Lambda^L} \left(\sum_{l=1}^{L} \lambda_l \hat{r}_l + \frac{\beta}{M} \sum_{k=1}^{L} \lambda_k \log \lambda_k \right. \\ &\quad \left. + \frac{\beta}{M} \log H + \sum_{k=1}^{L} \lambda_k |\mathbf{p}_k| \log \frac{2eL}{|\mathbf{p}_k|} \right) \end{aligned} \tag{7}$$

Note that $-\log L \leqslant \sum_{k=1}^{L} \lambda_k \log \lambda_k \leqslant 0$ on simplex Λ^L, the constructed coefficients $\hat{\boldsymbol{\lambda}}^{EW}$ approximately minimize the following regularized problem with prior in (6)

$$\min_{\boldsymbol{\lambda} \in \Lambda^L} \sum_{l=1}^{L} \lambda_l \hat{r}_l + \lambda_l |\mathbf{p}_l| \log \frac{2eL}{|\mathbf{p}_l|} \tag{8}$$

Compared (8) with the ideal ℓ_0 norm regularized regression

$$\min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_0$$

We can clearly see that the model aggregation based sparse estimation implicitly balances the data fitting error and the sparsity. Additionally, the proposed method has a finer view of the problem in the sense that we treat each pair of residual $\hat{r}_l$ and sparsity $|\mathbf{p}_l|$ individually. Again, more structural constraints on the support can be imposed by selecting a proper subset of candidates $\tilde{\mathcal{P}}$ in the aggregation.

4. NUMERICAL IMPLEMENTATION VIA MCMC

Using the exponential weighting coefficients $\hat{\boldsymbol{\lambda}}^{EW}$ defined in (4), we have the aggregate estimate given by

$$f_{\hat{\boldsymbol{\lambda}}^{EW}} = \sum_{\mathbf{p}_l \in \tilde{\mathcal{P}}} \hat{\lambda}_l^{EW} f_l = \frac{\sum_{\mathbf{p}_l \in \tilde{\mathcal{P}}} \hat{\mathbf{z}}_l \exp(-M\hat{r}_l/\beta)\pi_l}{\sum_{\mathbf{p}_l \in \tilde{\mathcal{P}}} \exp(-M\hat{r}_l/\beta)\pi_l}$$

Note that the exact computation of $\hat{\mathbf{x}}$ requires computing OLS for all the sparse patterns in $\tilde{\mathcal{P}}$. This will suffer great computational complexity if we have a fine grid that $N \gg 1$. Thus, an implementable algorithm is necessary to apply the model aggregation idea to the super-resolution applications.

As in [28, 29], we propose to find an approximate aggregate solution by implementing a proper Monte-Carlo Markov Chain. Different from prior work, we will have to adaptively design the transition and proposal distribution to enforce the structures of $\tilde{\mathcal{P}}$.

The use of MCMC is motivated by the fact that the desired aggregate solution $f_{\hat{\boldsymbol{\lambda}}^{EW}}$ is the expectation of a random variable $\hat{\mathbf{z}}_l$ with probability mass

$$\nu_l \propto \exp(-M\hat{r}_l/\beta)\pi_l, \qquad \mathbf{p}_l \in \tilde{\mathcal{P}} \tag{9}$$

where we ignore the normalizing factor for convenience. This distribution can be designated as the stationary distribution of a MCMC generated by the classic Metropolis-Hastings (MH) algorithm [33]. Under the setting of super-resolution, the proposed MH algorithm is defined on the L-hypercube graph of which each vertex corresponds to a sparse pattern.

The proposed algorithm follows the common MH framework and one of our contributions lies in the proposal distribution. For completeness, the high-level numerical procedure of MH is summarized in Algorithm 1.

As in [28, 29], we approximate the desired aggregate solution $f_{\hat{\boldsymbol{\lambda}}^{EW}}$ by

$$\hat{\mathbf{x}} = \frac{1}{T} \sum_{t=T_0+1}^{T_0+T} \hat{\mathbf{z}}_t \tag{10}$$

where T_0, T are user-defined parameters to approximate the exact expectation under certain time complexity.

4.1. Structure-aware proposal distribution choice

In [28,29], the authors do not impose any additional structural constraints on $\mathcal{P}$ that it contains all the possible support of sizes ranging from 0 to rank$(\mathbf{A})$. In this paper, our goal is to maintain the separation among the spikes for a more accurate and robust exact sparse recovery.

The key idea to design the proposal distribution $q(\cdot|\mathbf{p}_t)$ is to prune the vertices on the hypercube that are not in $\tilde{\mathcal{P}}$. At

Algorithm 1: Metropolis-Hastings algorithm

Select initial sparse pattern $\mathbf{p} = 0 \in \mathbb{R}^N$;
for *t=1,2,...* **do**
- 1. Generate a tentative move on $\tilde{\mathcal{P}}$ via sampling proposal distribution: $\tilde{\mathbf{p}}_t \sim q(\cdot|\mathbf{p}_t)$
- 2. Compute the acceptance probability: $r(\mathbf{p}_t, \tilde{\mathbf{p}}_t) = \min(\frac{\nu_{\tilde{p}}}{\nu_p}, 1)$, where $\nu_p, \nu_{\tilde{p}}$ are computed as (9) for $\mathbf{p}_t, \tilde{\mathbf{p}}_t$ respectively
- 3. Generate a random variable $u \sim \text{Uniform}(u; 0, 1)$
- 4. **if** $u < r$ **then**
  - Accept the proposal: $\mathbf{p}_{t+1} \leftarrow \tilde{\mathbf{p}}_t$;
- **else**
  - Reject the proposal: $\mathbf{p}_{t+1} \leftarrow \mathbf{p}_t$;
- **end**
- 5. Compute the $\hat{\mathbf{z}}_{t+1}$ with respect to $\mathbf{p}_{t+1}$

end

any iteration with sparse pattern $\mathbf{p}_t \in \tilde{\mathcal{P}}$, instead of randomly selecting an adjacent neighbor as in [28, 29], we first exclude the invalid candidates in $\mathcal{P}\backslash\tilde{\mathcal{P}}$ by temporarily freezing the coordinates within the neighborhood of $\mathbf{p}_t$ of which the size is determined by the desired separation. Next, we randomly select a neighboring vertex in $\tilde{\mathcal{P}}$. In particular, we take uniform distribution to search for the next valid move.

5. SIMULATION

In this section, we conduct extensive numerical experiments to demonstrate the superior performance of the proposed method compared to existing algorithms. Following the measurement model (1), the entries of signal $\mathbf{x}^\star$ and noise $\mathbf{w}^\star$ are generated as complex Gaussian random variables. We require that the true support satisfies the separation condition and the normalized mutual distances are no smaller than a predetermined value $\Delta \in (0, 1)$. The tuning parameter β in (4) is set to be four times the entry-wise noise power σ_w^2. The number of iterations of the proposed method is chosen to be 3000 with $T_0 = 2000$ and $T = 1000$ in (10). We consider three competing frequentist/Bayesian algorithms: the LASSO [18], the Square-root Lasso (equivalent to SPICE with single snapshot) [19, 34] and the SBL [21]. Given an estimate $\hat{\mathbf{x}}$, we claim a success if the support is exactly recovered without any post-processing. We will also compare the algorithms with respect to the normalized estimation error $\epsilon \triangleq \frac{\|\hat{\mathbf{x}}-\mathbf{x}^\star\|_1}{\|\mathbf{x}^\star\|_1}$. In sample runs, the largest and smallest amplitudes of $\mathbf{x}^\star$ are denoted by $x_{\max}, x_{\min}$ respectively.

We first show two sample runs in Fig. 1 and Fig. 2 under the settings that the sources are close or have sharp amplitude contrast. In both cases, the proposed method can provide exact sparse estimates while the other algorithms fail.

As only the proposed method can give sparse resolution, we show in Fig. 3 the empirical probability of success in terms of support recovery. We control the separation of two equal spikes and the results show the fact that the larger the separation, the higher the chance we can exactly locate the spikes. In Fig. 4, we compare the averaged normalized error of the four algorithms. It can be seen that the larger the con-

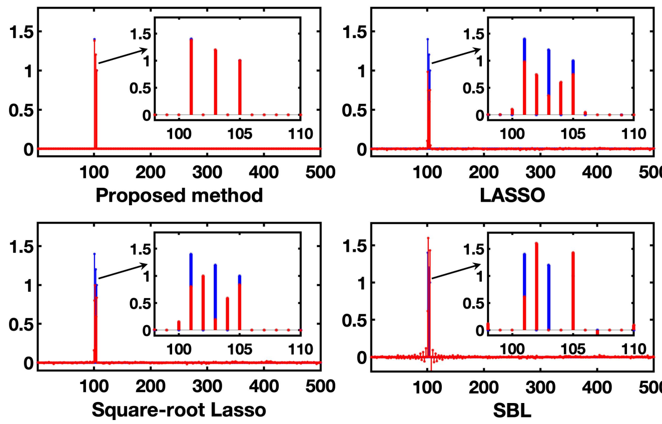

Fig. 1. Sample run of close sources. Blue spikes are true signals and red spikes are the estimates. $N = 500, M = 200, s = 3, \Delta = 0.004, x_{min} = 1, x_{max} = 1.4, \sigma_w = 0.1$.

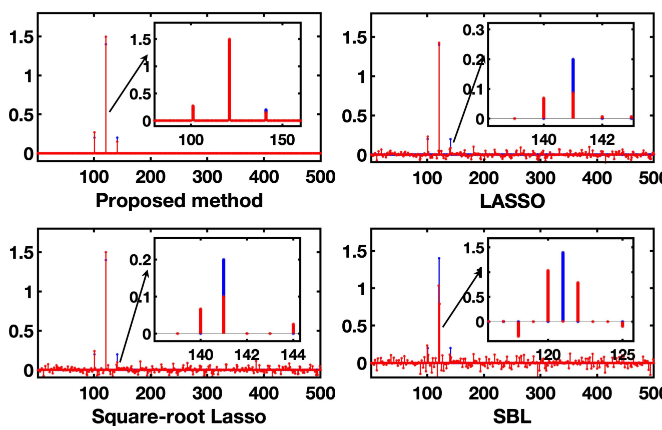

Fig. 2. Sample run of sharp amplitude contrast. $N = 500, M = 200, \Delta = 0.04, s = 3, x_{\min} = 0.2, x_{\max} = 1.4, \sigma_w = 1$.

trast in amplitudes is, the larger the estimation errors are. The proposed method is by nature combinatorial and we study its temporal complexity in Fig. 5. All four algorithms are run on a Dell desktop with Core i7-9700. The computational complexity of the proposed algorithm is comparable to others but it can provide better estimates as shown in earlier simulations.

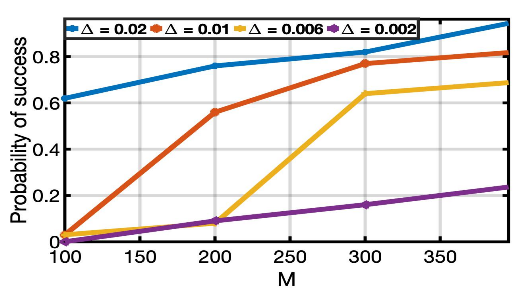

Fig. 3. Probability of exact support recovery. $N = 500, s = 3, x_{\min} = 1, x_{\max} = 1, \sigma_w = 0.1$. Averaged over 100 runs.

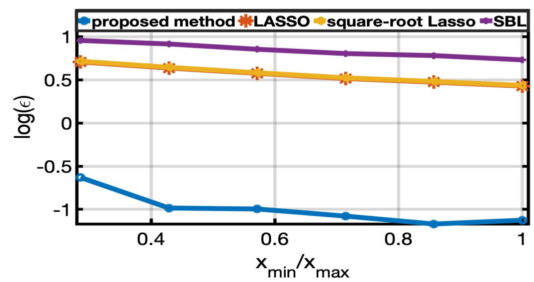

Fig. 4. Averaged normalized error ϵ as a function of $x_{\min}/x_{\max}$. $N = 500, M = 200, \Delta = 0.03, s = 3, \sigma_w = 1$. Averaged over 50 runs.

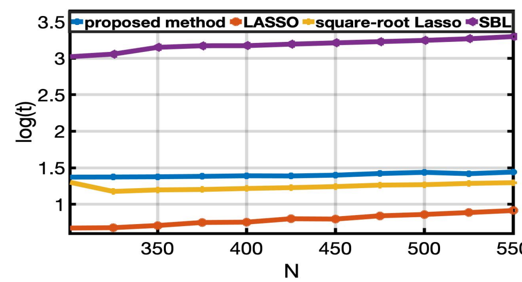

Fig. 5. Time complexities of four algorithms as functions of N. $M = 200, \Delta = 0.03, s = 3, x_{\min} = 1, x_{\max} = 1, \sigma_w = 1$. Averaged over 50 runs.

6. CONCLUSION

We propose a novel Bayesian discrete super-resolution algorithm based on the idea of model aggregation. By selecting priors over the possible supports, the proposed method resembles the frequentist sparse recover method with ℓ_0 norm based regularization. To efficiently implement the Bayesian approach, we exploit the MCMC framework with a customized proposal distribution to enforce the desired structural constraint like the separation condition in prior work. The numerical results demonstrate the advantages of the proposed method over other widely used methods in literature.

7. REFERENCES

[1] S. C. Park, M. K. Park and M. G. Kang, "Super-resolution image reconstruction: a technical overview", *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21-36, May 2003.

[2] J. Lindberg, "Mathematical concepts of optical superresolution", *J. Opt*, vol. 14, no. 8, July 2012.

[3] H. Greenspan, "Super-Resolution in Medical Imaging", *The Computer Journal*, vol. 52, no. 1, Jan. 2009.

[4] K. G. Puschmann and F. Kneer, "On super-resolution in astronomical imaging", *Astronomy and Astrophysics*, vol. 436, no. 1, pp. 373-378, June 2005.

[5] R. Heckel, V. I. Morgenshtern and M. Soltanolkotabi, "Super-resolution radar", *Information and Inference: A Journal of the IMA*, vol. 5, no. 1, pp. 22-75, March. 2016.

[6] D. L. Donoho, “Superresolution via sparsity constraints”, *SIAM J. Math. Anal.*, vol. 23, no. 5, pp. 1309-1331, Sep. 1992.

[7] E. J. Candès and C. Fernandez-Granda, “Towards a mathematical theory of super-resolution”, *Commun. Pure Appl. Math.*, vol. 67, no. 6, pp. 906-956, June 2014.

[8] V. I. Morgenshtern and E. J. Candès, “Super-Resolution of Positive Sources: the Discrete Setup”, *SIAM J. Imaging Sciences*, vol. 9, no. 1, pp. 412-444, 2016.

[9] T. Bendory, “Robust Recovery of Positive Stream of Pulses”, *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 2114-2122, April 2017.

[10] G. Tang, B. N. Bhaskar and B. Recht, “Near minimax line spectral estimation”, *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 499 -512, Jan. 2015.

[11] Z. Yang and L. Xie, “On gridless sparse methods for line spectral estimation from complete and incomplete data”, *IEEE Transactions on Signal Processing*, vol. 63, no. 12, pp. 3139-3153, June 2015.

[12] Z. Yang, L. Xie and C. Zhang, “A discretization-free sparse and parametric approach for linear array signal processing”, *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4959 - 4973, Oct. 2014.

[13] W. Li, W. Liao and A. Fannjiang, “Super-resolution limit of the ESPRIT algorithm”, *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4290 -4308, July. 2013.

[14] D. Batenkov, G. Goldman and Y. Yomdin, “Super-resolution of near-colliding point sources”, *arXiv: 1904.09186V1*, April 2019.

[15] S. Foucart and H. Rauhut, “A Mathematical Introduction to Compressive Sensing”, Birkhäuser/Springer, New York, 2013.

[16] C. R. Rao and Y. Wu, “On model selection”, IMS Lecture Notes-Monograph Series, vol. 38, 2001.

[17] H. Qiao and P. Pal, “Guaranteed Localization of More Sources than Sensors with Finite Snapshots in Multiple Measurement Vector Models Using Difference Co-Arrays”, *IEEE Transactions on Signal Processing*, vol. 67, no. 22, pp. 5715-5729, Nov. 2019.

[18] P. Pal and P. P. Vaidyanathan, “Pushing the limits of sparse support recovery using correlation information”, *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 711-726, Feb. 2015.

[19] P. Stoica, P. Babu and J. Li, “SPICE: A sparse covariance-based estimation method for array processing”, *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629-638, Feb. 2011.

[20] A. Bhadra, J. Datta, N. G. Polson and B. Willard, “Lasso meets horseshoe: a survey”, *Statistical Science*, vol. 34, no. 3, pp. 405-427, 2019.

[21] D. P. Wipf and B. D. Rao, “An empirical Bayesian strategy for solving the simultaneous sparse approximation problem”, *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704-3716, July 2007.

[22] D. L. Donoho, A. Maleki and A. Montanari, “Message passing algorithms for compressed sensing: I. motivation and construction”, *Proceedings of IEEE Information Theory Workshop on Information Theory* , Cairo, 2010.

[23] S. van de Geer, “Estimation and Testing Under Sparsity”, *Lecture Notes in Mathematics 2159*, Springer, 2016.

[24] I. Castillo, J. Schmidt Hieber and A. Van Der Vaart, “Bayesian linear regression with sparse priors”, *The Annals of Statistics*, vol. 43, no. 5, pp. 1986-2018, 2015.

[25] H. Qiao, “A universal technique for analysis discrete super-resolution algorithms”, *IEEE Signal Processing Letters*, vol. 27, pp. 1829-1833, 2020.

[26] D. Dai, P. Rigollet and T. Zhang, “Deviation optimal learning using greedy Q-aggregation”, *The Annals of Statistics*, vol. 40, no. 3, pp. 1878-1905, 2012.

[27] A. B. Tsybakov, “Aggregation and minimax optimality in high-dimensional estimation”, *Proceedings of International Congress of Mathematicians (Seoul, 2014)*, 3:225-246, 2014.

[28] P. Rigollet and A. B. Tsybakov, “Exponential screening and optimal rates of sparse estimation”, *The Annals of Statistics*, vol. 39, no. 2, pp. 731-771, 2011.

[29] P. Rigollet and A. B. Tsybakov, “Sparse estimation by exponential weighting”, *Statistical Science*, vol. 27, no. 4, pp. 558-575, 2012.

[30] E. J. Candès and C. Fernandez-Granda,“Super-resolution from noisy data”, *J. Fourier Anal. Appl.*, vol. 19, no. 6, pp. 1229-1254, 2013.

[31] P. C. Bellec, “Optimal bounds for aggregation of affine estimators”, *The Annals of Statistics*, vol. 46, no. 1, pp. 30-59, 2018.

[32] S. van de Geer and P. Bühlmann, “On the conditions used to prove oracle results for the Lasso”, *Electronic Journal of Statistics*, vol. 3, pp. 1360-1392, 2009.

[33] C. Robert and G. Casella, “Monte Carlo Statistical Methods”, Springer, New York, 2004.

[34] H. Qiao, “A universal technique for analysis discrete super-resolution algorithms”, *IEEE Signal Processing Letters*, vol 28, pp. 543-547, 2021.