Data-driven Parameter Estimation Of Contaminated Damped Exponentials

Youye Xie, Michael B. Wakin, and Gongguo Tang
Department of Electrical Engineering, Colorado School of Mines, Golden, CO, USA
{youyexie, mwakin, gtang}@mines.edu

2021 55th Asilomar Conference on Signals, Systems, and Computers | 978-1-6654-5828-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/IEEECONF53345.2021.9723189

***Abstract*—Damped exponentials appear naturally in a wide range of applications including structural health monitoring and electric machine fault detection. In this paper, given finite time-domain samples of composite, contaminated damped exponentials, we propose novel deep architectures to estimate the number of exponentials and recover the frequency and damping coefficient of each exponential. In our architecture, a damped exponential representation model maps time-domain samples to a frequency-damping spectrum representation, while a counting model then counts the number of exponentials. Combining the spectrum representation and the estimated number of exponentials, the frequencies and damping coefficients of the exponentials can be recovered automatically. Altogether, this yields an efficient learning-based method for parameter estimation of contaminated damped exponentials. Our experiments indicate that the proposed method is very effective and can robustly handle exponentials with close or even overlapping frequencies (resp. damping coefficients) as long as the damping coefficients (resp. frequencies) are sufficiently separated.**

***Index Terms*—Damped exponentials, deep learning, parameter estimation, signal decomposition**

I. Introduction

Advances in deep learning have led to a growing understanding of how to design networks for solving sparse recovery and estimation problems [1]–[4]. Recently, powerful deep networks [3], [4] have been designed for solving one of the most canonical sparse signal processing problems: estimating sinusoidal frequencies in line spectral estimation. The demonstrated performance is competitive with traditional methods such as MUSIC [5]. Meanwhile, the problem of estimating the frequencies and damping coefficients of *damped* exponentials from finite time-domain samples has wide applications, like structural health monitoring [6], [7] and fault detection [8], [9]. Unfortunately, this problem is more complicated than line spectral estimation, and the previous networks [3], [4] cannot accommodate the damped signal model. In this paper, inspired by [3], [4], we design novel deep architectures to estimate the number of exponentials and recover the frequency and damping coefficient of each exponential. Our work adds to the growing science of deep network design for signal processing.

A. Estimation Of Contaminated Damped Exponentials

Suppose a system observes a composite signal consisting of a linear combination of M (unknown) damped exponentials:

$$y(t) = \sum_{j=1}^{M} A_j e^{-\alpha_j t} e^{i2\pi f_j t} + \eta(t), \tag{1}$$

where each $A_j \in \mathbf{C}$ is a complex weight incorporating the unknown magnitude and phase of the j-th exponential, and $\eta(t)$ is additive Gaussian noise of unknown variance. By taking N samples of $y(t)$ with sampling interval T_s, we obtain a vector $\boldsymbol{y} \in \mathbf{C}^N$ whose n-th entry is

$$\begin{aligned} \boldsymbol{y}(n) &= \sum_{j=1}^{M} A_j e^{-\alpha_j T_s n} e^{i2\pi f_j T_s n} + \eta(T_s n) \\ &= \sum_{j=1}^{M} A_j e^{-\alpha_j n} e^{i2\pi f_j n} + \eta(n) \end{aligned} \tag{2}$$

where, without loss of generality (assuming non-aliased sampling), we take $T_s = 1$ and restrict the active frequencies $f_j \in [-0.5, 0.5)$. We assume $N = 50$ and each damping coefficient $\alpha_j \in [0, 0.1]$, so that a maximally damped exponential ($\alpha = 0.1$) will see its amplitude decay by approximately $100\times$ between its first and last sample.

Given the sample vector $\boldsymbol{y}$, our goals are to determine the number of exponentials M and to recover the corresponding frequency f_j and damping coefficient α_j of each exponential.

B. Related Work

When all $\alpha_j = 0$, our problem reduces to line spectral estimation [10], [11]. However, when damping is present, line spectrum estimation methods no longer apply. To take damping into account, several methods relying on the discrete Fourier transform (DFT) are proposed in [12]–[15]. These methods, however, assume either a single sinusoid or multiple sinusoids with well-separated frequencies. Some least square methods leveraging sparse techniques [16]–[22] are proposed in [23]–[25], which iteratively refine the estimated parameters and system order but are computationally cumbersome. Alternatively, two efficient time-domain approaches that allow exponentials with overlapping frequencies but separated damping coefficients are Prony's method [26], which solves a polynomial whose roots encode the parameters, and the matrix pencil method [27], [28], which constructs a matrix pencil based on the observed signal and then solves a generalized eigenvalue problem. However, the polynomial and matrix pencil methods require prior knowledge of the system order. And approaches such as Akaike information criterion (AIC) [29], minimum description length (MDL) [30], and second-order statistic of eigenvalues (SORTE) [31] are proposed to estimate the system order.

Deep learning has attracted significant interest due to its efficient end-to-end processing and competitive performance [2], [32]–[35]. Many deep architectures have been proposed for sparse recovery of different signal models [1]–[4], [36]. Of particular interest are deep neural networks such as DeepFreq [3] and PSnet [4] that achieve superior performance for line spectral estimation and [36] that extends the work of [3], [4] to multiband identification. However, those works assume no damping.

In this paper, we extend the works of [3], [4] by taking damping into account, and our work shows the potential for applying deep learning to a more complex sparse signal processing problem. Specifically, we propose a novel damped exponential representation model that maps the observed signal to a two-dimensional frequency-damping spectrum, in contrast with the one-dimensional frequency spectrum in [3], [4]. Moreover, to boost the performance in estimating exponentials with very close or overlapping frequencies but different damping coefficients, we propose a novel two-branch structure in the representation model to extract the frequency and damping information. We also propose a counting model to determine the number of exponentials, M, based on the estimated frequency-damping spectrum. Finally, the frequencies and damping coefficients can be extracted from the M peaks with the largest magnitudes in the spectrum.

The rest of the paper is organized as follows. In Section II, we propose the damped exponential representation model and counting model for parameter estimation of contaminated damped exponentials. In Section III, several experiments are conducted to evaluate the performance of the proposed networks and we conclude this paper in Section IV.

II. Proposed Methodology

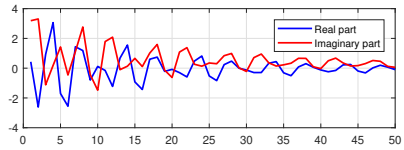
A. Damped Exponential Representation Model

In applying deep learning to line spectral estimation, [4] shows that nonparametrically predicting the frequency spectrum is more effective than parametrically predicting the frequencies directly. Inspired by that approach, our damped exponential representation model aims to map the observed signal to a nonparametric *frequency-damping spectrum* (FDS).

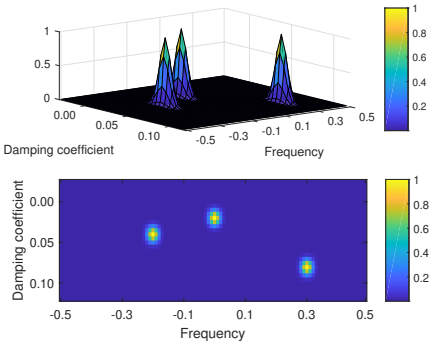
More specifically, the input of our representation model is $[\boldsymbol{y}_R^T, \boldsymbol{y}_I^T]^T \in \mathbf{R}^{2N}$ where $(.)^T$ is the transpose operator. $\boldsymbol{y}_R$ and $\boldsymbol{y}_I$ are the real and imaginary parts of the observed signal $\boldsymbol{y} = \boldsymbol{y}_R + i\boldsymbol{y}_I$. Based on the ground-truth parameters $\{(f_1, \alpha_1), (f_2, \alpha_2), \ldots, (f_M, \alpha_M)\}$ of the exponentials contained in an observed signal, the ground-truth FDS is defined to be the superposition of M generalized 2D Gaussian kernels,

$$FDS(f, \alpha) = \sum_{j=1}^{M} K(f - f_j, \alpha - \alpha_j) \tag{3}$$

where the kernel has the form $K(f, \alpha) = e^{-\left(f^2/\sigma_f^2 + \alpha^2/\sigma_\alpha^2\right)}$. Larger values of the standard deviations σ_f and σ_α allow for more informative non-zero values backpropagated during calibration but at the cost of lower resolution. In this paper, we set $N = 50$, $\sigma_f = 0.9/N$, and $\sigma_\alpha = 0.45/N$. The



(a) The real and imaginary parts of the observed signal.



(b) The corresponding frequency-damping spectrum .

Fig. 1: *A noisy observed signal consists of three damped exponentials whose parameters* (f_j, α_j) *are* $(-0.2, 0.04)$, $(0.0, 0.02)$, *and* $(0.3, 0.08)$ *respectively. The observed signal's real and imaginary parts are shown in (a) and its frequency-damping spectrum is shown in (b). We pad the spectrum with damping coefficient values smaller than* 0 *and greater than* 0.1.

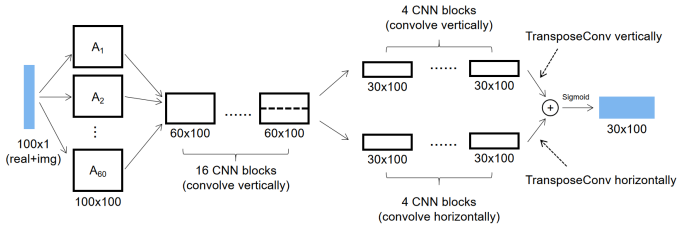


Fig. 2: *The Dank model for damped exponentials. Each convolutional neural network (CNN) block consists of a convolution layer with kernel size* 6 *and circular padding and a batch normalization layer followed by the rectified linear unit (ReLU). The sizes of linear encoders, model's input and output, and each CNN block are marked.*

Gaussian kernel applied in the FDS enforces the fact that a close estimate is more valuable than estimates far from the ground-truth. Since we use the squared error loss function to train the network, when the Gaussian kernel is applied, the loss function value for a close estimate is smaller than an estimate far from the ground-truth. If the observed signal has more than 50 samples, our network could use the first 50 samples. An example of the input and output of the representation model is shown in Fig. 1, in which the discretized FDS is of size 30×100. To account for periodicity in the frequency parameter f, we use a circular extension of the frequency axis. Since no such periodicity exists for the damping coefficient, we pad the spectrum with α values smaller than 0 and greater than 0.1. With this discretization, the resolution of f and α are 0.01 and 0.005, respectively.

The structure of our representation model, which we term

Dank (deep + damp), is shown in Fig. 2. We first linearly encode the input signal to an intermediate feature space of 60 channels. Then several convolution neural network (CNN) blocks consisting of a convolution layer with kernel size 6 and circular padding, a batch normalization layer [37], and a rectified linear unit (ReLU) layer further process the features. Despite the differences in the size of the feature channel, convolution kernel, and padding, the structure of the feature encoder and CNN block are inspired by the DeepFreq [3] and PSnet [4], respectively, which are designed for line spectrum estimation. [3] finds that the learned feature encoder for undamped exponential signals implements a Fourier-like transformation and based on that, the localized kernel in the convolution layer can accurately locate the sinusoid frequencies. However, the Fourier transform of a damped exponential is a generalized Dirichlet kernel parameterized by α [38], which suggests that a fixed size localized convolution kernel is not appropriate. Moreover, to enable the identification of exponentials with overlapping frequencies but separated damping coefficients, we introduce a novel two-branch network structure. In particular, the top branch (see Fig. 2) implements the convolution vertically as in the PSnet [4] to locate the frequencies, while the bottom branch implements the convolution horizontally to estimate the damping coefficients utilizing the information from the whole feature channel. Finally, the transposed convolution [39] decoders with kernel size 1 produce the estimated spectrum.

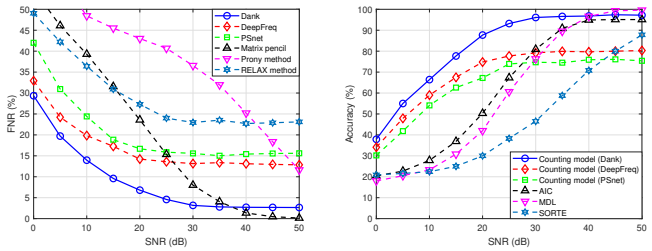
B. Damped Exponential Counting Model

Based on the predicted FDS, we train a counting model to determine the number of valid damped exponentials in the observed signal. Specifically, the input of the counting model is the estimated FDS and the output is a single value to be rounded to the nearest integer. Based on the estimated system order M, the frequency and damping coefficients are then extracted from the M peaks having the largest magnitudes in the input spectrum. Since the system order should be invariant to the translation of the peaks in the FDS, the counting model consists of 20 CNN blocks introduced in Section II-A followed by a convolution layer with output channel and kernel size 1 and a fully connected layer to predict the final value.

III. Numerical Experiments

A. Experiment Setup

To validate our approach, we generate simulation data based on (2). The training, validation, and testing datasets consist of 200000, 1000, and 1000 composite signals, respectively. For each signal, $N = 50$, M is uniformly selected from 1 to 5, and $A_j = (0.1 + |g|)e^{i\theta}$, where g follows the standard Gaussian distribution and θ is uniformly selected in $[0, 2\pi]$. All pairs of frequencies and damping coefficients are required to satisfy at least one minimum separation condition between the pair of frequencies $(4/N)$ **or** damping coefficients (0.04). First, the frequency f_1 and damping coefficient α_1 are selected uniformly in the range of $[-0.5, 0.5)$ and $[0.0, 0.1]$, respectively. When $M \geq 2$, the second exponential is generated to have its frequency close to the frequency of the first exponential. Specifically, $f_2 = f_1 + u$ where u is uniformly selected in $[-1/N, 1/N]$. Thus, the first and second exponentials have very close or overlapping frequencies but well-separated damping coefficients (due to the separation condition). We apply the same procedure for generating the third and fourth exponentials if $M \geq 4$. Moreover, exponentials with well-separated frequencies may have very close or overlapping damping coefficients. We train the proposed Dank model by applying the Adam algorithm [40] for 200 epochs, minimizing the squared ℓ_2 error between the network estimated FDS and the ground-truth FDS, $||FDS_{est} - FDS_{gt}||_2^2$. The initial learning rate is 0.0003 which reduces by half when the loss function does not decrease for 3 consecutive epochs on the validation set. The batch size is 256. During each epoch, we add scaled Gaussian noise to the signals so that the SNR for each example is chosen uniformly at random between 0 and 50 dB. The counting model is trained using a fixed-weight representation model to generate the FDS and following the same training setup with 100 epochs. Similarly, the counting model's loss function is the squared ℓ_2 norm error between the network output and the ground-truth system order.



(a) False negative rate. (b) Counting accuracy.

Fig. 3: *Performance of Dank and counting models.*

B. Performance Of Dank And Counting Models

We first validate the performance of the Dank representation model in terms of the false negative rate (FNR) given the ground-truth number of exponentials for each observed signal. A successful recovery is counted for a damped exponential when the frequency and damping coefficient errors are both smaller than $1/N$. In addition, we compare the proposed Dank model to several representative methods: DeepFreq [3], PSnet [4], total-least-squares (TLS) matrix pencil [27], TLS Prony's method [41], and RELAX [23] which minimizes a nonlinear least squares problem. Note that because the original DeepFreq [3] and PSnet [4] only concern the frequency spectrum, we modify their final layers by changing the output dimension so that they too can predict the FDS. We train these networks following the same training setup as our model. The results are recorded in Fig. 3 (a). The proposed model outperforms other methods in the low to medium SNR regimes and achieves less than 3% FNR when SNR ≥ 35 dB.

We also examine the system order estimation performance of the proposed counting model, which determines the number

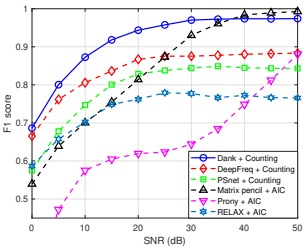
of exponentials based on the FDS estimated by the pre-trained Dank model. We compare with the AIC [29], MDL [30], and SORTE [31] methods which estimate the system order based on the signal covariance matrix. Moreover, we train separate counting models of the same structure using the FDS estimated by DeepFreq [3] and PSnet [4] respectively. The results are shown in Fig. 3 (b). We observe that our counting model outperforms the covariance matrix based methods except at high SNR, and the quality of the predicted spectrum has a great impact on the counting accuracy.

C. The Overall Performance Of The Combined Models

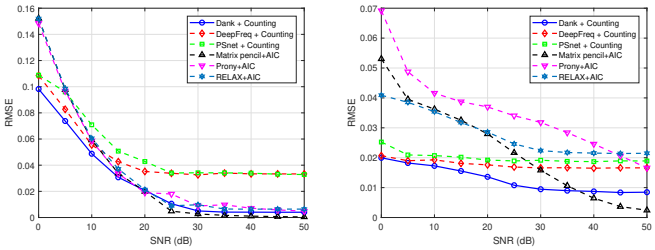
We assess the overall performance of the combined models in terms of the F1 score and root mean square error (RMSE). Based on the success recovery criteria in Section III-B, we calculate the $PRECISION$ and $RECALL$ and define the F1 score to be $2/(PRECISION^{-1} + RECALL^{-1})$. Moreover, for each ground-truth exponential, we calculate the recovery errors of the frequency and damping coefficient. Then the RMSEs of the frequency and damping coefficient are calculated based on the recovery errors of all ground-truth exponentials in the testing dataset. We combine the system order estimation method AIC with the matrix pencil, Prony's method, and RELAX method and the DeepFreq and PSnet with their trained counting models for comparison. The results are recorded in Fig. 4. We see that, in terms of F1 score, the proposed method outperforms other deep learning methods over the whole range of SNRs and the traditional methods by a large margin in low to medium SNRs. And in terms of RMSE, although the traditional methods and proposed method achieve comparable performance for frequency estimation, our method is much more effective in estimating the damping coefficients in low to medium SNRs. Results show that our method can robustly handle damped exponentials with close or overlapping frequencies and well-separated damping coefficients.

D. Without Overlapping Frequencies And Real Data

In this section, we generate a synthetic signal without overlapping frequencies based on a measured nuclear magnetic resonance (NMR) signal in [25]. We set $M = 3$, $(f_1, f_2, f_3) = (0.078, 0.196, 0.287)$, $(\alpha_1, \alpha_2, \alpha_3) = (0.015, 0.023, 0.017)$, $(|A_1|, |A_2|, |A_3|) = (7.09, 2.31, 5.98) \times 10^4$, and $(\angle A_1, \angle A_2, \angle A_3) = (-0.102, -0.283, -0.173) \times 2\pi$. We add complex Gaussian noise with zero mean and different variances, σ_η^2, and we define the magnitude-to-noise ratio $\Psi = 10 \log_{10} \left(\sum_{j=1}^{M} |A_j|^2 / \sigma_\eta^2 \right)$. The ground-truth system order, 3, is provided for all methods and we record their accumulated RMSE in Fig. 5. The accumulated RMSEs of frequency and damping coefficient are defined as the sum of the RMSE of (f_1, f_2, f_3) and $(\alpha_1, \alpha_2, \alpha_3)$ respectively across 1000 trials. The Cramér-Rao Lower Bound (CRLB) is calculated based on [23]. The 2-D cubic interpolation is applied to the estimated FDS by Dank for better accuracy. We observe that when there are no exponentials with overlapping frequencies, our method has a comparable performance to traditional methods. In addition, measured on a system with

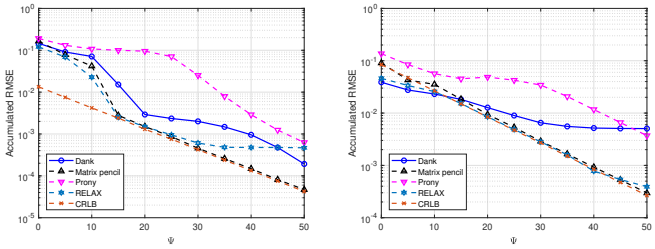


(a) F1 score.



(b) RMSE of frequency. (c) RMSE of damp.

Fig. 4: *The overall performance of the combined models.*



(a) Accumulated RMSE of frequency. (b) Accumulated RMSE of damp.

Fig. 5: *Damped exponentials without overlapping frequencies.*

i7-6700 CPU and GTX 1080 GPU, Dank takes around 2.05 seconds to process 1000 signals of length 50. In contrast, matrix pencil, the Prony's method, and RELAX method take 0.87, 0.22, and 11.02 seconds respectively. Although Dank is efficient in testing, it takes around eight hours to train.

Finally, we examine the reconstruction error of different methods in a real data experiment. Specifically, we hang an iPhone from a 1 centimeter line, release the phone from a 45 degree initial position, and record its accelerometer data. The recorded data is real-valued; we pass 0s as the imaginary part to the input of Dank. Based on the rank of the Hankel matrix formed by the recorded data [9], we set the system order to be 5 for all methods. We then attempt to reconstruct the data based on the parameters estimated by each method. And the average relative reconstruction errors among 10 trials for Dank, matrix pencil, Prony's method, and RELAX method are 12.8%, 10.5%, 13.6%, and 8.3% respectively. This demonstrates the effectiveness of Dank on real data.

IV. Conclusion

We apply deep learning to the parameter estimation problem for contaminated damped exponentials. We propose two

novel networks, the Dank representation model and the corresponding counting model, to map the observed signal to an FDS and subsequently determine the number of exponentials based on the estimated spectrum. Experiments show that the proposed approach can handle composite signals of damped exponentials with (i) very close or overlapping frequencies (resp. damping coefficients) but different damping coefficients (resp. frequencies) and (ii) contamination by Gaussian noise at varying noise levels.

REFERENCES

[1] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the 27th international conference on machine learning*, pp. 399–406, 2010.

[2] Y. Xie, Z. Wang, W. Pei, and G. Tang, "Fast approximation of non-negative sparse recovery via deep learning," in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 2921–2925, IEEE, 2019.

[3] G. Izacard, S. Mohan, and C. Fernandez-Granda, "Data-driven estimation of sinusoid frequencies," in *Advances in Neural Information Processing Systems*, pp. 5127–5137, 2019.

[4] G. Izacard, B. Bernstein, and C. Fernandez-Granda, "A learning-based framework for line-spectra super-resolution," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3632–3636, IEEE, 2019.

[5] W. Liao and A. Fannjiang, "Music for single-snapshot spectral estimation: Stability and super-resolution," *Applied and Computational Harmonic Analysis*, vol. 40, no. 1, pp. 33–67, 2016.

[6] J. Ndambi, B. Peeters, J. Maeck, J. De Visscher, M. Wahab, J. Vantomme, G. De Roeck, and W. De Wilde, "Comparison of techniques for modal analysis of concrete structures," *Engineering Structures*, vol. 22, no. 9, pp. 1159–1166, 2000.

[7] H. Qarib and H. Adeli, "A new adaptive algorithm for automated feature extraction in exponentially damped signals for health monitoring of smart structures," *Smart Materials and Structures*, vol. 24, no. 12, p. 125040, 2015.

[8] M. Kanemaru, M. Tsukima, T. Miyauchi, and K. Hayashi, "Bearing fault detection in induction machine based on stator current spectrum monitoring," *IEEJ Journal of Industry Applications*, vol. 7, no. 3, pp. 282–288, 2018.

[9] Y. Xie, D. Liu, H. Mansour, and P. T. Boufounos, "Robust parameter estimation of contaminated damped exponentials," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5500–5504, IEEE, 2020.

[10] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, 2013.

[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, 2015.

[12] E. Aboutanios, "Estimation of the frequency and decay factor of a decaying exponential in noise," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 501–509, 2009.

[13] R.-C. Wu and C.-T. Chiang, "Analysis of the exponential signal by the interpolated dft algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 12, pp. 3306–3317, 2010.

[14] E. Aboutanios and S. Ye, "Efficient iterative estimation of the parameters of a damped complex exponential in noise," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 975–979, 2014.

[15] K. Wang, H. Wen, W. Tai, and G. Li, "Estimation of damping factor and signal frequency for damped sinusoidal signal by three points interpolated dft," *IEEE Signal Processing Letters*, vol. 26, no. 12, pp. 1927–1930, 2019.

[16] Y. Xie, S. Li, G. Tang, and M. B. Wakin, "Radar signal demixing via convex optimization," in *2017 22nd International Conference on Digital Signal Processing (DSP)*, pp. 1–5, IEEE, 2017.

[17] S. Foucart and H. Rauhut, "A mathematical introduction to compressive sensing," *Bull. Am. Math*, vol. 54, no. 2017, pp. 151–165, 2017.

[18] Y. Xie, M. B. Wakin, and G. Tang, "Support recovery for sparse recovery and non-stationary blind demodulation," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pp. 235–239, IEEE, 2019.

[19] Y. Xie, M. B. Wakin, and G. Tang, "Sparse recovery and non-stationary blind demodulation," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5566–5570, IEEE, 2019.

[20] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[21] Y. Xie, M. B. Wakin, and G. Tang, "Simultaneous sparse recovery and blind demodulation," *IEEE Transactions on Signal Processing*, vol. 67, no. 19, pp. 5184–5199, 2019.

[22] Y. Xie, M. B. Wakin, and G. Tang, "Support recovery for sparse signals with unknown non-stationary modulation," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1884–1896, 2020.

[23] Z.-S. Liu, J. Li, and P. Stoica, "Relax-based estimation of damped sinusoidal signal parameters," *Signal processing*, vol. 62, no. 3, pp. 311–321, 1997.

[24] J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, "High resolution sparse estimation of exponentially decaying n-dimensional signals," *Signal Processing*, vol. 128, pp. 309–317, 2016.

[25] M. Juhlin, F. Elvander, J. Swärd, and A. Jakobsson, "Fast gridless estimation of damped modes," in *2018 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pp. 346–351, IEEE, 2018.

[26] T. Sauer, "Prony's method in several variables," *Numerische Mathematik*, vol. 136, no. 2, pp. 411–438, 2017.

[27] T. K. Sarkar and O. Pereira, "Using the matrix pencil method to estimate the parameters of a sum of complex exponentials," *IEEE Antennas and Propagation Magazine*, vol. 37, no. 1, pp. 48–55, 1995.

[28] P. Guillaume, J. Schoukens, and R. Pintelon, "Sensitivity of roots to errors in the coefficient of polynomials obtained by frequency-domain estimation methods," *IEEE Transactions on Instrumentation and Measurement*, vol. 38, no. 6, pp. 1050–1056, 1989.

[29] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, 1985.

[30] M. Wax and I. Ziskind, "Detection of the number of coherent signals by the mdl principle," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 8, pp. 1190–1196, 1989.

[31] Z. He, A. Cichocki, S. Xie, and K. Choi, "Detecting the number of clusters in n-way probabilistic clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2006–2021, 2010.

[32] W. Pei, Y. Xie, and G. Tang, "Spammer detection via combined neural network," in *International Conference on Machine Learning and Data Mining in Pattern Recognition*, pp. 350–364, Springer, 2018.

[33] A. Mousavi and R. G. Baraniuk, "Learning to invert: Signal recovery via deep convolutional networks," in *2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2272–2276, IEEE, 2017.

[34] Y. Xie, Y. Tang, G. Tang, and W. Hoff, "Learning to find good correspondences of multiple objects," in *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 2779–2786, IEEE, 2021.

[35] Y. Tang, K. Kojima, T. Koike-Akino, Y. Wang, P. Wu, Y. Xie, M. H. Tahersima, D. K. Jha, K. Parsons, and M. Qi, "Generative deep learning model for inverse design of integrated nanophotonic devices," *Laser & Photonics Reviews*, vol. 14, no. 12, p. 2000287, 2020.

[36] Y. Xie, M. B. Wakin, and G. Tang, "Contaminated multiband signal identification via deep learning," in *2021 IEEE Statistical Signal Processing Workshop (SSP)*, IEEE, 2021.

[37] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, pp. 448–456, PMLR, 2015.

[38] M. Bertocco, C. Offelli, and D. Petri, "Analysis of damped sinusoidal signals via a frequency-domain interpolation algorithm," *IEEE Transactions on Instrumentation and measurement*, vol. 43, no. 2, pp. 245–250, 1994.

[39] H. Gao, H. Yuan, Z. Wang, and S. Ji, "Pixel transposed convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 5, pp. 1218–1227, 2019.

[40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[41] W. M. Steedly, C.-H. Ying, and R. L. Moses, "A modified TLS-Prony method using data decimation," *IEEE transactions on signal processing*, vol. 42, no. 9, pp. 2292–2303, 1994.